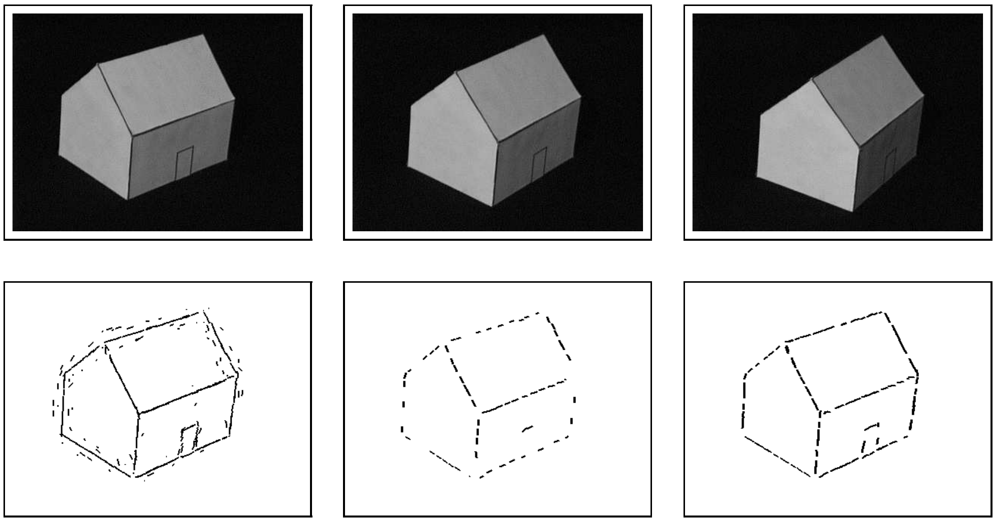
Interaktion verschiedener visueller Modalitäten zur stabilen Extraktion von Objektrepräsentationen



Diplomarbeit
am Lehrstuhl für kognitive Systeme
des Instituts für Informatik und praktische Mathematik
der Christian-Albrechts-Universität zu Kiel

vorgelegt von

Thomas Jäger

Kiel, im März 2002
Prof. Dr. Gerald Sommer
Betreuer: Dr. Norbert Krüger, Dr. Christian Perwass

Zusammenfassung

Die Arbeit beschreibt ein Verfahren, das aus kalibrierten Stereobildsequenzen dreidimensionalen Strukturinformationen gewinnt. Diese erhaltene Strukturinformation ist aufgrund der Unsicherheit lokaler visueller Verfahren fehlerbehaftet. Das vorgestellte Verfahren nutzt den *Bewegungszusammenhang* in den Bildsequenzen aus, um die dreidimensionale Information zu stabilisieren und Fehler zu erkennen. Dies geschieht durch eine Integration der visuellen Modalitäten Stereorekonstruktion, optischer Fluß, Poseschätzung und Akkumulation.

Zusätzlich wird ein heuristisches Filterungsverfahren entwickelt, das Fehler in dreidimensionalen Strukturen aufgrund von *statistischen Regularitäten* dieser Strukturen erkennen und entfernen kann. Es wird gezeigt, daß dieses Filterungsverfahren notwendig ist, um das beschriebene Verfahren erfolgreich anzuwenden.

Es werden Experimente präsentiert, anhand derer der Erfolg und die Qualität des vorgestellten Verfahrens aufgezeigt werden. Sowohl eine visuelle als auch eine statistische Auswertung der Ergebnisse findet statt.

Schlüsselworte: Stereorekonstruktion, Poseschätzung, optischer Fluß, Akkumulation

Inhaltsverzeichnis

Kapitel 1

Einleitung

1.1 Motivation

Die vorliegende Arbeit beschäftigt sich mit der Extraktion von dreidimensionalen Informationen aus Stereobildsequenzen. Diese Sequenzen werden mit kalibrierten Kameras aufgenommen.

Aufgrund der generellen Ungenauigkeit lokaler visueller Verfahren sind die auf diesem Weg gewonnenen Informationen über die dreidimensionale Struktur der betrachteten Objekte mit Fehlern behaftet. Die Ausnutzung von Bewegungszusammenhängen hilft, Fehler zu erkennen und damit zu entfernen. Die extrahierte Information wird parametrisch stabilisiert, d.h. durch eine Zuordnung eines Konfidenzmaßes zu den gewonnenen Informationen werden vorhandene Fehler in diesen Informationen beschrieben.

Krüger, Ackermann et al. [28, 1] geben ein solches Verfahren an. Durch Bewegung eines Objektes in verschiedene Ansichten werden unterschiedliche 3D-Repräsentationen von diesem Objekt gewonnen. Gemeinsamkeiten der in den einzelnen Ansichten extrahierten Repräsentationen werden genutzt, um zu einer stabileren Objektrepräsentation zu gelangen. Das Wissen um die Bewegung des Objektes in die einzelnen Ansichten ist dazu nötig. Das heißt, Ansichten werden generiert, indem das Objekt aktiv und kontrolliert mit einem Roboterarm bewegt wird.

Wissen um Bewegung kann auch durch die Analyse von Bildfolgen auf visuellem Weg ermittelt werden. Mit Hilfe des optischen Flusses (siehe z.B. Mallot [32]) bzw. einem Poseschätzverfahren (Rosenhahn et al. [41, 44]) ist zwei- bzw. dreidimensionale Bewegung extrahierbar. Ein Einsatz dieser beiden Techniken soll es möglich machen, die Bewegungsparameter anstatt durch aktive Kontrolle auch durch visuelle Beobachtung zu erhalten. Ziel dieser Arbeit ist es, die Interaktion der verschiedenen visuellen Modalitäten

Stereorekonstruktion, Akkumulation, optischer Fluß und Poseschätzung für die Erlangung einer stabilen 3D-Repräsentation von Objekten zu benutzen.

Die Poseschätzung reagiert sensitiv, wenn die zur Schätzung verwendete Repräsentation große Fehler enthält. Neben der Ausnutzung des Bewegungszusammenhanges zur Gewinnung einer verläßlichen Repräsentation werden deshalb speziell für die Poseschätzung zusätzlich statistische Regelmäßigkeiten der betrachteten Objekte herangezogen, um eine stabile Schätzung der Bewegung zu ermöglichen. Zu diesem Zweck wird ein Verfahren entwickelt, welches auf der Basis von Kollinearität in dreidimensionalen Objektstrukturen Fehler in der extrahierten Information entfernen kann.

1.2 Literatur

In der Literatur finden sich Arbeiten, die ähnlich wie in dieser Arbeit eine Integration unterschiedlicher Modalitäten verwenden, um 3D-Strukturinformation zu extrahieren. So geben Brodský et al. [8] ein monokulares Verfahren an, welches den Zusammenhang zwischen Bewegung und Tiefenstruktur ausnutzt, um eine Schätzung der Bewegung und der Tiefenstruktur einer Szene vorzunehmen. Eine Ausnutzung des Bewegungszusammenhanges über mehrere Ansichten der Szene, um eine Stabilisierung der gewonnenen Informationen zu erlangen, findet nicht statt.

Shi et al. [42] verwenden den optischen Fluß und ein Stereokamerasystem, um die Tiefe der Szene und die 3D-Bewegung betrachteter Objekte zu berechnen. Die Informationen werden nicht an charakteristischen Punkten extrahiert, sondern es werden Tiefen- und Bewegungskarten für die gesamte Bildfläche errechnet. Eine Stabilisierung und Fehlerbeschreibung dieser Karten wird nicht durchgeführt. Die Ergebnisse überzeugen nur für translatorische Bewegungen

Von Wang und Duncan [46] ist ein Verfahren angegeben, welches in einem Stereosystem mit parallelen Kameras an charakteristischen Punkten im linken und rechten Bild den optischen Fluß berechnet, und die Unterschiede zwischen den linken und rechten Flußwerten nutzt, um die 3D-Bewegung zu rekonstruieren. Das Verfahren kann mit mehreren, sich unabhängig voneinander bewegenden Objekten in einer Szene umgehen. Eine Fehlerbeschreibung findet auch hier nicht statt.

Ho und Pong [21] benutzen die Eigenschaften der Stereorekonstruktion und die des optischen Flusses, um für eine Menge von charakteristischen Punkten einer Szene die Tiefeninformation und die 3D-Bewegung zu erhalten. Eine Verfahren zur Stabilisierung der geschätzten Tiefeninformationen wird bei dieser Arbeit nicht angewendet.

Bregler und Malik [7] benutzen Flußverfahren und Poseschätzung, um anhand von Videosequenzen, die Menschen bei der Ausführung verschiedener Bewegungen zeigen, das vorgegebene 3D-Modell eines Menschen zu animieren. Im Unterschied zur vorliegenden Arbeit ist das 3D-Modell des Menschen gegeben und wird nicht aus der betrachteten Szene extrahiert.

1.3 Aufbau

Die Arbeit ist wie folgt aufgebaut: Im Kapitel 2 wird eine Übersicht über die Problematik der 3D-Stereorekonstruktion gegeben, und wie diese durch Ausnutzung von Bewegungszusammenhängen gelöst werden kann. Der Unterschied zur Arbeit von Krüger, Ackermann et al. [28, 1] wird dargestellt.

In Kapitel 3 werden mathematische Grundlagen und die verwendeten visuellen Modalitäten Stereorekonstruktion, optischer Fluß, Poseschätzung und das Akkumulationsverfahren eingeführt und erläutert. Anschließend wird in Kapitel 4 die Interaktion der einzelnen Modalitäten dargestellt und der Algorithmus formalisiert. Die durchgeführten Experimente und deren Ergebnisse sind in Kapitel 5 erklärt. Es wird sowohl eine visuelle als auch eine statistische Auswertung vorgenommen. Abschließend faßt das Kapitel 6 die Arbeit zusammen und gibt einen Ausblick.

Kapitel 2

Problemstellung

2.1 Unsicherheit der Informationsgewinnung

Ein Grundproblem bei der Gewinnung von Informationen aus visuellen Eingabedaten ist, daß die extrahierten Informationen generell unsicher sind (siehe z.B. Aloimonos und Shulman [2], Ayache [5]). Die lokalen Operationen wie z.B. die Kantendetektion, die Berechnung des optischen Flusses oder die Stereorekonstruktion können aufgrund ihrer Eigenschaften und der Eigenschaften der zugrunde liegenden Bilder nur fehlerhafte Ergebnisse erzielen.

So wird bereits beim Vorgang der Bildaquirierung die zur Verfügung stehende Information eingeschränkt. Bedingt durch das Auflösungsvermögen des CCD-Chips der verwendeten Kamera wird die eigentlich kontinuierlich vorliegende Helligkeitsverteilung der aufgenommemen Szene in einzelne Bildpunkte diskretisiert. Die Helligkeitsinformation für jeden einzelnen Bildpunkt wird quantisiert. Zusätzlich zu diesen Verlusten durch Diskretisierung und Quantisierung ist das aufgenommene Bild systembedingt durch einen gewissen Rauschanteil gestört. Bei zeitlich aufeinanderfolgenden Bildaufnahmen (Videosequenzen) bzw. Stereosystemen tritt zusätzlich das Problem von Beleuchtungsschwankungen bzw. Beleuchtungsunterschieden zwischen den verschiedenen Bildern auf. Diese verursachen eine Veränderung der Reflexionseigenschaften von Objekten und sind nur schwer analytisch zu modellieren (siehe z.B. Ikeuchi et al. [23]).

Bei der Informationsextraktion aus mehreren Bildern, wie es bei der Stereorekonstruktion oder der Berechnung des optischen Flusses nötig ist, treten zusätzliche Schwierigkeiten in Form des Korrespondenz- und des Aperturproblems auf (siehe z.B. Mallot [32], Klette [26]). Das Aperturproblem und das Korrespondenzproblem treten bei der Bestimmung des optischen Flusses auf, und für die Stereorekonstruktion ist das Korrespondenzproblem von

Relevanz. Beide Probleme werden im Folgenden detaillierter beschrieben.

2.1.1 Korrespondenzproblem

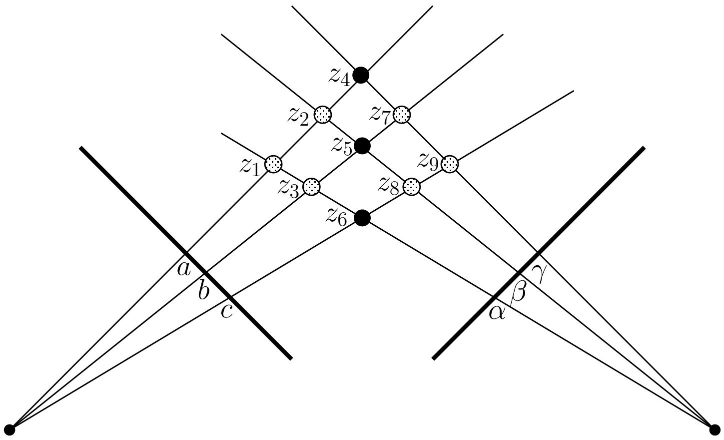


Abbildung 2.1: Korrespondenzproblem

Möchte man in Stereobildern eine Tiefenschätzung vornehmen, so ist es erforderlich, zu den im linken Bild detektierten Objektpunkten die korrespondierenden Punkte im rechten Bild zu finden. Das heißt, es müssen die beiden Orte in den Stereobildern gefunden werden, an denen sich jeweils die Projektion eines Punktes in der dreidimensionalen Szene befindet.

In Abb. 2.1 sind im linken Bild an den Orten a, b und c Merkmale vorhanden, und im rechten an den Orten α, β und γ. Lassen sich die extrahierten Merkmale in den Bilder nicht unterscheiden, so ergeben sich in dieser Konstellation 9 mögliche Zuordnungen $z_1, \ldots, z_9$. Wird eine der falschen Zuordnungen für die Tiefenschätzung verwendet, so ergibt sich eine geschätzte Position im Raum, die nicht mit dem realen Ort übereinstimmt. Es handelt sich bei dem dabei auftretenden Fehler nicht nur um eine relativ geringe ungenaue Positionsbestimmung. Der Fehler kann nicht durch eine Gaußsche Verteilung modelliert werden. Durch falsche Korrespondenzen auftretende Fehler können sehr groß werden, die bestimmte Position ist völlig verkehrt.

2.1.2 Aperturproblem

Das Apertur- oder Blendenproblem tritt auf, wenn die Bewegung einer Linie oder Kante über mehrere, zeitlich aufeinanderfolgende Bilder, bestimmt werden soll. Viele gängige Verfahren zum Berechnen von optischen Flußvektoren beruhen auf lokalen Operatoren ([6]). Die Szene wird daher gewissermaßen nur durch eine kleine Blende (Apertur) beobachtet. Die Abb. 2.2 macht das

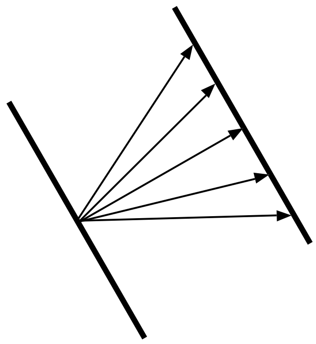


Abbildung 2.2: Aperturproblem, wohin wurde die Linie tatsächlich verschoben?

Problem deutlich. Es ist aufgrund des kleinen Betrachtungsfenstern nicht möglich, verläßlich zu sagen, in welche Richtung die Linie sich tatsächlich bewegt hat.

Es ist nur möglich, den Anteil der Bewegung zu bestimmen, der senkrecht zur Orientierung der Linie besteht. Abb. 2.3 zeigt die wirkliche Bewegung der Linie (links) und den Ausschnitt, der betrachtet werden kann, mit der daraus resultierenden Bewegungsdetektion senkrecht zur Linie (rechts). Es wird die sich eigentlich im zweidimensionalen abspielende Bewegung um eine Dimension reduziert, eine Aussage über die eigenliche Richtung der Bewegung ist nicht mehr möglich.

Dieses Phänomen der Dimensionsreduktion findet man auch in anderen Verfahren. Im Allgemeinen läßt sich sagen, daß die meisten informationstragenden Bereiche in Bilder intrinsisch eindimensional sind (siehe Zetsche, Barth [48]).

2.2 Rekonstruktion von 3D-Repräsentationen

Es wird dargestellt, welche Schwierigkeiten bei der Rekonstruktion von 3D-Repräsentationen auftreten, und angegeben, wie man die auftretenden Probleme behandeln kann.

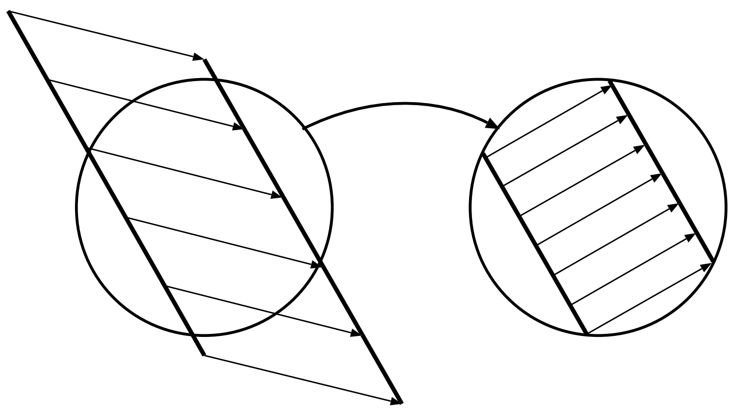


Abbildung 2.3: Aperturproblem, reale Bewegung (links) und detektierte Bewegung (rechts)

2.2.1 Schwierigkeiten der 3D-Rekonstruktion

Im oben beschriebenen Kontext der Unzuverlässigkeit von visuell gewonnenen Informationen wird nun die dreidimensionale Rekonstruktion eines Objektes betrachtet. Verwendet werden dazu Bilder, die mit Hilfe eines kalibrierten Stereokamerasystems aufgenommen werden. In diesen Bildern werden Kanten der abgebildeten Objekte detektiert. Die Eigenschaften Phase und Orientierung, gewonnen durch sphärische Quadraturfilter (siehe Felsberg, Sommer [12]) dieser detektierten Kanten werden dazu herangezogen, um im linken und rechten Bild korrespondierende Kantensegmente zu finden. Aus diesen beiden 2D-Liniensegmenten[1] in den Bildebenen wird ein 3D-Liniensegment[1] im Raum rekonstruiert. Da mit kalibrierten Kameras gearbeitet wird, kann die Tiefe des neuen 3D-Segments im Raum berechnet werden. Somit erhält man einen 3D-Punkt, an dem das 3D-Liniensegment liegt (Poerksen [37], Krüger und Felsberg [29]). Zusätzlich zu dieser Punkt-Information wird aus den 2D-Orientierungen der zugrunde liegenden 2D-Liniensegmente eine 3D-Orientierung berechnet. Diese wird dem 3D-Punkt zugeordnet. Das 3D-Liniensegment besteht also aus der Ortsangabe im Raum und der Orientierungsinformation.

Aufgrund des Korrespondenzproblems kommt es dabei zu einem Anteil von Liniensegmenten, welche falsch berechnet werden. Die Abb. 2.4 zeigt die Aufsicht auf ein rekonstriertes 3D-Modell eines Hauses. Zu erkennen ist, daß es sich um die Gibelansicht handelt. Doch darüber hinaus sind falsche

[1] Wir benutzen den Begriff Liniensegmente hier als Punkte mit assoziierter Orientierung, siehe Kap. 3.1.3.

Liniensegmente extrahiert worden, welche nicht zu dem Modell des Hauses gehören. Auch finden sich Liniensegmente, deren Lage im Raum offenbar richtig rekonstruiert wurde, aber deren Orientierung falsch ist.

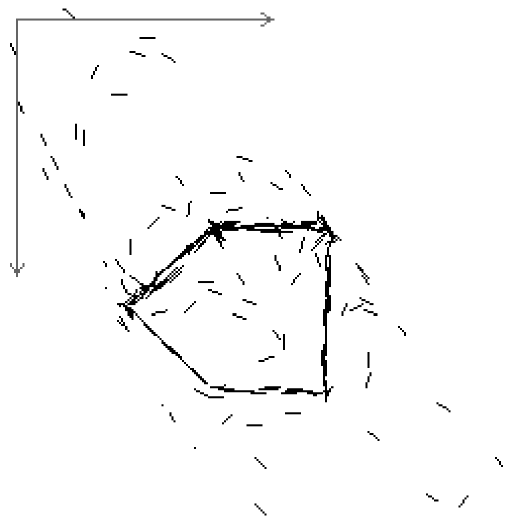


Abbildung 2.4: 3D-Stereorekonstruktion mit fehlerhaft extrahierten Liniensegmenten

2.2.2 Stabilisierung durch Ausnutzung des Bewegungszusammenhanges

Ein Verfahren zur Verbesserung der 3D-Stereorekonstruktion wird von Krüger, Ackermann in [28, 1] angegeben. Das Verfahren kann aus einer Sequenz von verschiedenen Stereoansichten eines Objektes durch Akkumulation eine stabilere Objektrepräsentation gewinnen.

Dabei werden dem System neben der Ansicht des Objektes auch jeweils die zugehörigen Bewegungsparameter übergeben. Das Objekt wird dabei unter kontrollierten Bedingungen aktiv durch einen Roboterarm bewegt. Dabei wird vom System eine vom Benutzer vorgegebene Liste mit den gewünschten Bewegungsparametern abgearbeitet. Daher sind die Bewegungsparameter explizites Vorwissen, welches dem System von außen zugeführt wird. Sie werden nicht auf dem visuellem Weg gefunden.

In der neuen Ansicht des Objektes wird eine 3D-Stereorepräsentation berechnet. Mit Hilfe der Bewegungsparameter läßt sich diese in die Ausgangsposition überführen. Dort überlagert sich die neue Repräsentation mit den bisher errechneten. Mittels definierter Abstandsmaße für 3D-Liniensegmente

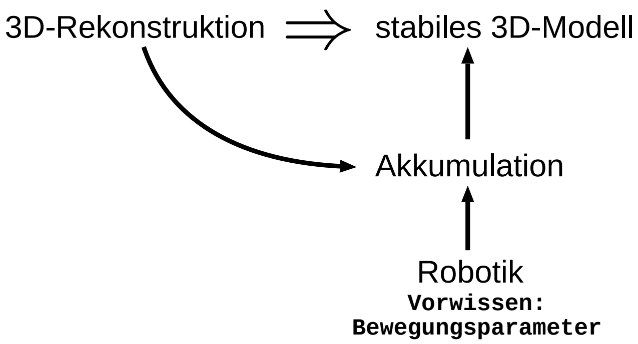


Abbildung 2.5: Akkumulationsverfahren mit explizitem Vorwissen der Bewegungsparameter

können die in den verschiedenen Ansichten ermittelten 3D-Repräsentationen akkumuliert werden. Dies geschieht durch eine Konfidenzbewertung der 3D-Liniensegmente. Korrekt rekonstruierte Segmente werden mit hoher Wahrscheinlichkeit in mehr Ansichten vorkommen als falsch rekonstruierte. Darauf basierend werden die Segmente mit höherer Konfidenz bewertet, je öfter sie in den unterschiedlichen Ansichten erscheinen. Segmente, die seltener auftreten, erhalten eine niedrige Konfidenz. Die Bewegung stellt damit den Zusammenhang zwischen den einzelnen Ansichten her.

Segmente, welche während der Akkumulation mit einer hoher Bewertung versehen werden, sind also mit hoher Wahrscheinlichkeit als korrekt rekonstruiert anzusehen. Sie sind für die modellhafte Beschreibung des Objektes brauchbar. Hingegen sind die Segmente mit niedriger Konfidenzbewertung fehlerhaft und gehören damit nicht zu dem untersuchten Objekt.

Je mehr Ansichten dem System präsentiert werden, umso stabiler wird das akkumulierte Modell werden. Die Abb. 2.5 zeigt schematisch dieses Akkumulationsverfahren.

In Krüger, Ackermann ([28, 1]) wurde dieses Verfahren auf 3D-Punkte mit assoziierten Orientierungen angewendet. In Poerksen ([37]) wurde die Phase als Strukturmerkmal der Kanten hinzugefügt. Generell ist das Verfahren erweiterbar auf alle Beschreibungselemete, deren Veränderung unter Bewegung beschreibbar ist und welche im metrischen Raum anwendbar sind (siehe [28]).

2.3 Interaktion verschiedener Modalitäten

Das beschriebene Akkumulationsverfahren erzeugt sich selbst, durch die Verwendung der Robotik, die benötigten unterschiedlichen Ansichten der Modelle. Daher sind die Bewegungsparamter exakt bekannt.

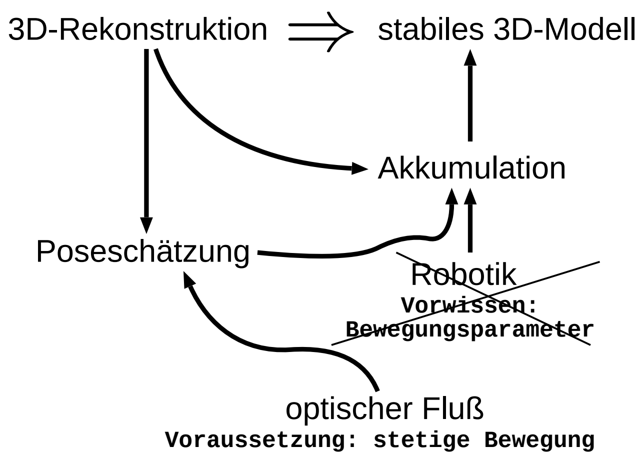


Abbildung 2.6: Interaktion der verschiedenen visuellen Modalitäten

Im Rahmen dieser Arbeit soll gezeigt werden, daß es möglich ist, basierend auf dem Akkumulationsverfahren, durch Hinzunahme weiterer visueller Modalitäten die Bewegungsparameter nicht explizit vorgeben zu müssen, sondern diese ebenfalls aus den Bildern zu schätzen. Das Wissen um die Bewegung wird nicht mehr durch die Robotik vorgegeben, sondern soll nun durch Beobachtung der Welt gewonnen werden. Die folgenden Modalitäten werden neu hinzugenommen:

- Ein Verfahren zur Poseschätzung (siehe [44]). Dieses ermittelt aus einer dreidimensionalen Repräsentation eines Objektes in einer bestimmten Pose und einer zweidimensionalen Projektion des Objektes in einer anderen Pose die Bewegungsparameter, welche die Lageveränderung zwischen den beiden Posen beschreiben.

  Nötig für die Poseschätzung ist das Wissen um die Korrespondenz zwischen den 3D-Liniensegmenten der 3D-Repräsentation und den 2D-Liniensegmenten der 2D-Projektion.

- Ein Verfahren zur Berechnung des optischen Flusses (siehe [6]). Dieser gibt für eine Sequenz von Bildern, welche ein bewegtes Objekt zeigen, jeweils an, wie sich die Bewegung des Objektes in der Verschiebung der einzelnen Bildpunkte auswirkt.

Dieses Wissen über die Verschiebung von Bildpunkten wird angewendet, um die Korrespondenz zwischen den 3D-Liniensegmenten und den 2D-Liniensegmenten der Projektion zu erhalten, welche für die Poseschätzung benötigt wird.

Damit ist es nicht mehr notwendig, die Bewegungsparameter als explizites Vorwissen anzugeben, sondern diese werden jetzt durch Anwendung und geeignete Interaktion der genannten visuellen Modalitäten ebenfalls aus den Bilddaten berechnet. In Abb. 2.6 ist schematisch die Interaktion von 3D-Stereorekonstruktion, Poseschätzung, optischem Fluß und Akkumulation dargestellt. Das Vorwissen um die Bewegungsparameter wird nicht mehr benötigt. Für die Berechnung des optischen Flusses ist es allerdings notwendig, eine stetige Bewegung des Objektes vorauszusetzen (siehe Kapitel 3.3). Dies ist in vielen Situationen gegeben, und damit gerechtfertigt.

Kapitel 3

Vorstellung der Module

3.1 Mathematische Grundlagen

Die für das Verständnis der Arbeit notwendigen mathematischen Grundlagen werden eingeführt. Zunächst wird das Modell zur Beschreibung von geometrischen Objekten und deren Transformationen, also Rotation und Translation, erläutert. Dann werden die mathematischen Strukturen vorgestellt, welche in dieser Arbeit für die Repräsentation von zwei- und dreidimensionalen Objekten verwendet werden.

3.1.1 Rotationen und Translationen

Die Bewegung eines Objektes im Raum bezeichnet man als Transformation. Sie wird in zwei Einzelbewegungen zerlegt, die Translation (Verschiebung) der Objektpunkte, und der anschließenden Rotation dieser Punkte.

3.1.1.1 Spezielle orthogonale Gruppe

Rotationen können mathematisch durch orthogonale Matrizen mit Determinante gleich eins beschrieben werden. Ausgangspunkt für die exaktere Beschreibung dieser Matrizen ist die *allgemeine lineare Gruppe*. Es wird vorrausgesetzt, daß der Leser mit dem mathematischen Begriff der Gruppe vertraut ist. Die allgemeine lineare Gruppe umfaßt die Menge der invertierbaren Matrizen:

$$GL(n) = \{M \in \mathbb{R}^{n \times n} | \det M \neq 0\}. \tag{3.1}$$

Darauf aufbauend kann jetzt die *orthogonale Gruppe* der orthogonalen Matrizen mit der Eigenschaft $M^{-1} = M^{\mathrm{T}}$ definiert werden:

$$O(n) = \{M \in GL(n) | M^{-1} = M^{\mathrm{T}}\}. \tag{3.2}$$

Die Eigenschaft $M^{-1} = M^{\mathrm{T}}$ der orthogonalen Matrizen bedingt, daß die Determinante dieser Matrizen nur den Wert ± 1 annehmen kann. Die orthogonale Gruppe $O(n)$ kann anhand dessen in zwei Untergruppen zerlegt werden. Diese enthalten nur die Matrizen mit Determinante 1 bzw. jene mit Determinante -1. Bei der ersteren der beiden handelt es sich um die *spezielle orthogonale Gruppe*:

$$SO(n) = \{M \in GL(n) | M^{-1} = M^{\mathrm{T}} \wedge \det M = 1\}. \tag{3.3}$$

Die Matrizen aus dieser Gruppe beschreiben Rotationen um den Ursprung des verwendeten Koordinatensystems. Dies bedeutet z.B. für den $\mathbb{R}^3$, daß Matrizen aus der Gruppe $SO(3)$ Rotationen im Raum um den Nullpunkt beschreiben.

3.1.1.2 Spezielle Euklidische Gruppe

Die spezielle orthogonale Gruppe $SO(n)$ beschreibt nur Rotationen. Eine Drehung um einen beliebigen Punkt im Raum erfordert eine Erweiterung dieses Modelles.

Die *spezielle Euklidische Gruppe* $SE(n)$ ist die Vereinigung von Rotation und Translation:

$$SE(n) = \{(t, M) | t \in \mathbb{R}^n, M \in SO(n)\} \tag{3.4}$$

Rotationen um einen beliebigen Punkt können durchgeführt werden, indem zunächst eine Translation durchgeführt wird, so daß der Drehpunkt im Koordinatenursprung zu liegen kommt. Nun kann die Rotation um den Ursprung durchgeführt werden. Anschließend wird die Translation rückgängig gemacht, so daß der Drehpunkt wieder seine ursprüngliche Position erhält.

3.1.2 Geometrische Algebra

Zur Beschreibung von geometrischen Objekten wie Punkten und Linien, und zur Transformation dieser Objekte durch Rotationen und Translationen wird in dieser Arbeit die Clifford-Algebra, auch geometrische Algebra (GA) genannt, verwendet ([18, 19, 36]). Eine kurze Einführung soll das Verständnis erleichtern.

Die Clifford-Algebra $\mathcal{C}l_n(\mathcal{V}^n)$ über einem Vektorraum $\mathcal{V}^n$ ist eine Erweiterung der bekannten Vektoralgebra. In dieser ist das Skalarprodukt $s = a \cdot b$ für Vektoren definiert, welches einen skalaren Wert s als Ergebnis liefert. Der Wert s gibt den Anteil, den der Vektor a entlang des Vektors b hat. Stehen die Vektoren senkrecht aufeinander, so ist der Wert folglich Null. Das Kreuzprodukt $c = a \times b$ liefert einen Vektor c, der senkrecht zu a und b steht.

3.1.2.1 Das Geometrische Produkt

In der geometrischen Algebra finden diese beiden Produkte eine Entsprechung im inneren und äußeren Produkt. Das äußere Produkt $a \wedge b$ hat den selben Betrag wie $a \times b$, aber es beschreibt weder ein Skalar, noch einen Vektor, sondern eine gerichtete Ebene, auch Bivektor oder Blade vom Grad Zwei genannt. Es handelt sich dabei um einen zweidimensionalen Unterraum von $\mathcal{V}^n$.

Da das innere Produkt $a \cdot b$ einen Skalar und das äußere Produkt $a \wedge b$ einen Bivektor ergibt, werden durch diese Operationen der Grad eines Vektors erniedrigt respektive erhöht:

$$a \cdot \underbrace{(b \wedge c)}_{\text{Bivektor}} = \underbrace{(a \cdot b)c - (a \cdot c)b}_{\text{Vektor}}. \tag{3.5}$$

Desweiteren haben sie, angewendet auf Vektoren, gegensätzliche kommutative Eigenschaften:

$$\begin{aligned} a \cdot b &= b \cdot a \qquad \text{(kommutativ)}, && (3.6)\\ a \wedge b &= -b \wedge a \qquad \text{(anti-kommutativ)}. && (3.7) \end{aligned}$$

Zusammengenommen bilden sie den symmetrischen und den antisymmetrischen Teil eines neuen Produktes (definiert von Grassmann [17] und Clifford [10]), dem sogenannten geometrischen Produkt ab:

$$ab = a \cdot b + a \wedge b. \tag{3.8}$$

Das geometrische Produkt paralleler Vektoren ist demnach ein Skalar, und das Produkt orthogonaler Vektoren ist ein Bivektor. Das Produkt von Vektoren, die weder parallel noch orthogonal zueinander sind, enthält sowohl einen skalaren als auch einen Bivektoranteil und wird als Multivektor bezeichnet.

Der Multivektor, welchen man erhält, wenn man alle Basisvektoren eines Vektorraumes mit dem geometrischen Produkt verknüpft, also

$$I = e_1 e_2 \dots e_n, \tag{3.9}$$

hat in der entsprechenden geometrischen Algebra maximalen Grad, und spannt den gesamten Vektorraum auf. I wird Pseudoskalar genannt.

3.1.2.2 Dualität

Das Dual A^* eines Blades $A \in \mathcal{C}l_n(\mathcal{V}^n)$ ist dasjenige Blade, welches an A fehlt, um den gesamten Raum aufzuspannen. Es ist folgendermaßen definiert:

$$A^* = AI^{-1}, \tag{3.10}$$

wobei I^{-1} das Inverse des Pseudoskalars ist, für das gilt $II^{-1} = 1$. Das Produkt eines Blades mit seinem Dual ergibt

$$AA^* = AAI^{-1} = \lambda I^{-1} \simeq I, \tag{3.11}$$

d.h. bis auf einen skalaren Faktor λ den Pseudoskalar.

3.1.2.3 Rotoren

Rotationen werden in der geometrischen Algebra mit Hilfe von Rotoren beschrieben. Ein Rotor R ist das geometrische Produkt zweier normierter Vektoren $\hat{a}$ und $\hat{b}$, welche die (normierte) Rotationsebene $\hat{U}$ aufspannen. Eine Rotation um den Winkel θ stellt sich wie folgt dar:

$$\begin{aligned} R &= \hat{a}\hat{b} = \hat{a} \cdot \hat{b} + \hat{a} \wedge \hat{b} \\ &= \cos\frac{\theta}{2} + \sin\frac{\theta}{2}\hat{U}, \text{ wobei } \angle(\hat{a}, \hat{b}) = \frac{\theta}{2} && (3.12) \\ &= e^{\frac{\theta}{2}\hat{U}}, \text{ wobei } \hat{U} = (\hat{a} \wedge \mathrm{rej}(\hat{b}, \hat{a}))\widehat{\ }. && (3.13) \end{aligned}$$

Die Rejektion $\mathrm{rej}(x, y) = x_\perp$ ist der Vektor $x_\perp$, der den Anteil von x bezeichnet, welcher senkrecht zu y steht.

Die Rotation eines Punktes x geschieht durch die links- und rechtsseitige Multiplikation mit dem Rotor R:

$$Rx\tilde{R} = x_\perp + x_\| e^{-\theta\hat{U}}. \tag{3.14}$$

Dabei bezeichnet $x_\|$ den Anteil von x, welcher parallel zur Rotationsebene $\hat{U}$ liegt. Dies bedeutet, daß der parallel zur Rotationsebene liegende Anteil von x in der Ebene rotiert wird, und der senkrecht zur Ebene stehende Anteil von x unverändert erhalten bleibt.

3.1.2.4 Projektive Geometrie

Die projektive Geometrie $\mathbb{P}^3$ ist eine Einbettung des dreidimensionalen Euklidischen Raumes $\mathbb{E}^3$ mit der Basis $\{e_1, e_2, e_3\}$ in den vierdimensionalen projektiven Raum $\mathbb{P}^3$ mit der Basis $\{e_1, e_2, e_3, e_4\}$ durch Hinzunahme der vierten projektiven (oder homogenen) Komponente e_4. Ein euklidischer Punkt a wird zu dem homogenen Punkt A durch

$$a = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} \xrightarrow{\mathbb{P}^3} A = a + e_4 = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ 1 \end{pmatrix}. \tag{3.15}$$

Der Übergang von $X \in \mathbb{P}^3$ zur $x \in \mathbb{E}^3$ wird durch normalisieren von X, d.h. Teilen von X durch seine vierte Komponete und dem anschließenden Weglassen dieser vierten, zu Eins gewordenen Komponente:

$$a = \sum_i \frac{A \cdot e^i}{A \cdot e^4} e_i, \qquad \text{wobei } i \in \{1,2,3\}, \qquad e_i \cdot e^i = \delta^i_i. \tag{3.16}$$

3.1.2.5 Darstellung Geometrischer Objekte

Im projektiven Raum kann eine Linie, gegeben durch zwei Punkte A und B, durch einen Bivektor L dargestellt werden:

$$L = A \wedge B. \tag{3.17}$$

Analog wird eine Ebene durch drei Punkte A, B und C durch einen Trivektor P dargestellt:

$$P = A \wedge B \wedge C. \tag{3.18}$$

Ein Punkt X, der auf der Linie L oder einer Ebene P liegt, gehorcht

$$X \wedge L = 0 \text{ bzw. } X \wedge P = 0. \tag{3.19}$$

3.1.3 Darstellung von Liniensegmenten

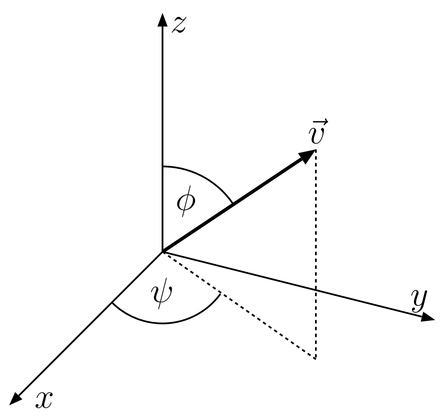


Abbildung 3.1: Darstellung der Orientierung im Raum

Zur Repräsentation von Objekten werden 2D- und 3D-Liniensegmente verwendet. Dargestellt werden diese hier nicht durch ihre Anfangs- und Endpunkte im zwei- oder dreidimensionalen Raum, sondern nur durch einen

Punkt, assoziiert mit einer Orientierung. Solch einem Liniensegment kann für den Zweck seiner Visualisierung in einem Bild eine willkürliche Länge zugeordnet werden.

Die Orientierung eines 2D-Liniensegments wird angegeben im Winkel zur x-Achse des Koordinatensystems. Die 3D-Orientierung wird angegeben in sphärischen Koordinaten (ϕ, ψ), wobei ϕ den Winkel zwischen der x-Achse und der Projektion des Orientierungsvektors $\vec{v}$ auf die xy-Ebene des Koordinatensystems angibt, und ψ den Winkel von $\vec{v}$ zur z-Achse (siehe Abb. 3.1). Der Orientierungsvektor $\vec{v}$ berechnet sich durch (Zavodnik [47]):

$$\vec{v} = \begin{pmatrix} \sin\phi\cos\psi \\ \sin\phi\sin\psi \\ \cos\phi \end{pmatrix} \tag{3.20}$$

Für die Winkel gilt $0 \leq \phi < \pi, 0 \leq \psi < 2\pi$. Veranschaulicht wird die

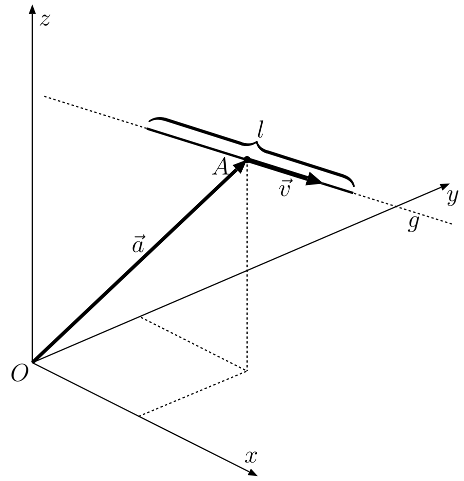


Abbildung 3.2: Darstellung eines 3D-Liniensegmentes mit Aufpunkt A und Orientierung $\vec{v}$

Darstellung eines 3D-Liniensegmentes, bestehend aus Aufpunkt A und Orientierung $\vec{v}$ in Abbildung 3.2. Der Punkt A, dargestellt durch den Vektor $\vec{a}$, beschreibt die Lage des Liniensegmentes im Raum. Durch den Vektor $\vec{v}$ wird die zu A assoziierte Orientierung angegeben. Zur Visualisierung von Lage und Orientierung kann auf der durch A und $\vec{v}$ beschriebenen Geraden g ein Liniensegment mit der Länge l eingezeichnet werden.

3.2 3D-Rekonstruktion mittels Stereo

Durch die Verwendung von zwei Kameras, die eine Szene von unterschiedlichen Blickpunkten aus betrachten, lassen sich 3D-Szeneninformationen aus den beiden aufgenommenen Bildern gewinnen (siehe z.B. Faugeras [11]). Dazu ist es nötig, daß die Parameter der Kamerageometrie bekannt sind. Diese erhält man durch eine Kalibration der Kameras (siehe z.B. Faugeras [11], Tsai [45]). Die Aufgabe, 3D-Informationen mittels Stereo und mit Hilfe von kalibrierten Kameras zu extrahieren, kann in zwei Teilschritte zerlegt werden, die Korrespondenzermittlung, und die anschließende 3D-Rekonstruktion.

3.2.1 Korrespondenzermittlung

Zunächst muß die Korrespondenz zwischen Punkten im linken und rechten Bild bestimmt werden. Dies bedeutet, es muß festgestellt werden, welche Bildpunkte des linken und rechten Bildes Projektionen desselben Punktes in der dreidimensionalen Szene sind.

Für die Ermittlung korrespondierender Bildpunkte kann man korrelationsbasierte Verfahren anwenden, welche mittels einer Ähnlichkeitsfunktion zu einem Patch im linken Bild ein entsprechendes Patch im rechten Bild suchen. Merkmalsbasierte Methoden verwenden eine Menge von Merkmalen wie z.B. Grauwertkante oder -kontur, um entsprechende Punkte in beiden Bildern zu finden. Hier wird eine Korrespondenzfindung auf der Grundlage von lokaler Orientierung d und Phase ϕ verwendet (siehe Poerksen [37], Felsberg und Sommer [12]). Diese beiden Eigenschaften werden durch sphärische Quadraturfilter gewonnen und sind in dem Strukturmultivektor zusammengefaßt. Damit ist es möglich, auf sehr kompakte Art und Weise geometrische und strukturelle Information (siehe Felsberg, Sommer [12]) zu benutzen und diese in ihrer unterschiedlichen Funktion separat zu behandeln (siehe [29]). Erweiterungsmöglichkeiten wie z.B. durch Farbe (Gebken [14]) sind gegeben.

Die Ähnlichkeitsfunktion d, welche zur Korrespondenzfindung benutzt wird, ist die gewichtete quadrierte Summe der Differenzen von Orientierung (Δd) und Phase ($\Delta\phi$):

$$\mathrm{d} = \alpha(\Delta d)^2 + (1-\alpha)(\Delta\phi)^2. \tag{3.21}$$

Um den Suchbereich einzuschränken, können analytische oder geometrische Zwänge verwendet werden. Analytische Zwänge sind z.B. *uniqeness* oder *continuity* (siehe [11]). Hier verwendet wird der *epipolare Zwang*, der zu den geometrischen Zwängen gehört.

3.2.1.1 Der epipolare Zwang

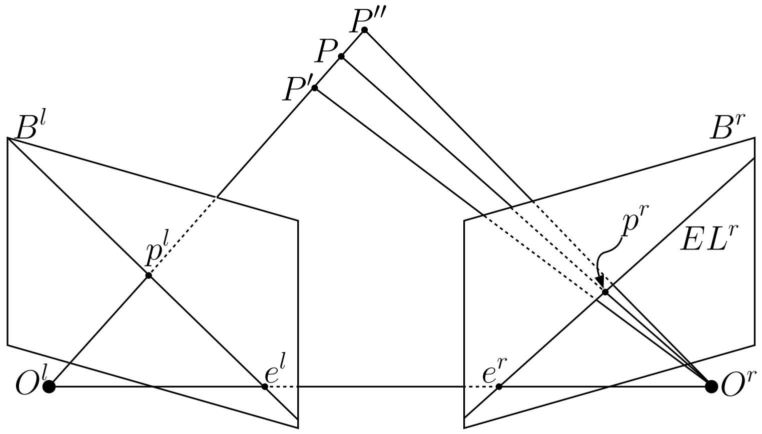


Abbildung 3.3: Epipolargeometrie

Dieser Zwang ergibt sich aus der Geometrie des Stereo-Sehens. Durch die Punkte O^l, O^r und P in Abb. 3.3 wird die *Epipolarebene* definiert. Diese Ebene schneidet die Bildebenen B^l und B^r in den *Epipolargeraden* EL^l bzw. EL^r. Der zu Punkt p^l korrespondierende Punkt p^r muß auf der Epipolargeraden EL^r liegen. Der Suchbereich kann damit auf eine Dimension reduziert werden. Da gerade der Punkt P rekonstruiert werden soll, liegt das Problem vor, die Epipolarebene ohne Kenntnis über diesen Punkt zu bestimmen. Dieses Problem läßt sich lösen, da als Grundannahme gilt, daß mit kalibrierten Kameras gearbeitet wird, denn hieraus ist die Epipolargeometrie berechenbar (siehe Faugeras [11]).

3.2.2 3D-Rekonstruktion

Sind korrespondierende Bildpunkte gefunden, so kann die *Rekonstruktion* der 3D-Information vorgenommen werden. Anhand der *Disparität*, der Verschiebung der zusammengehörigen Elemente zwischen dem linken und rechten Bild, und der bekannten Geometrie des Stereokamera-Systems, kann die Tiefeninformation des Szeneelements berechnet werden. Sind die korrekten Korrespondenzen (p^l, p^r) bzw. (q^l, q^r) bestimmt worden, so erhält man die Punkte P bzw. Q durch den Schnitt der Strahlen $\overrightarrow{O^l p^l}$ und $\overrightarrow{O^r p^r}$ bzw. $\overrightarrow{O^l q^l}$ und $\overrightarrow{O^r q^r}$ (siehe Abb. 3.4). Schneiden sich die Strahlen nicht, so wird der Punkt auf dem linken Strahl gewählt, welcher dem rechten Strahl am näch-

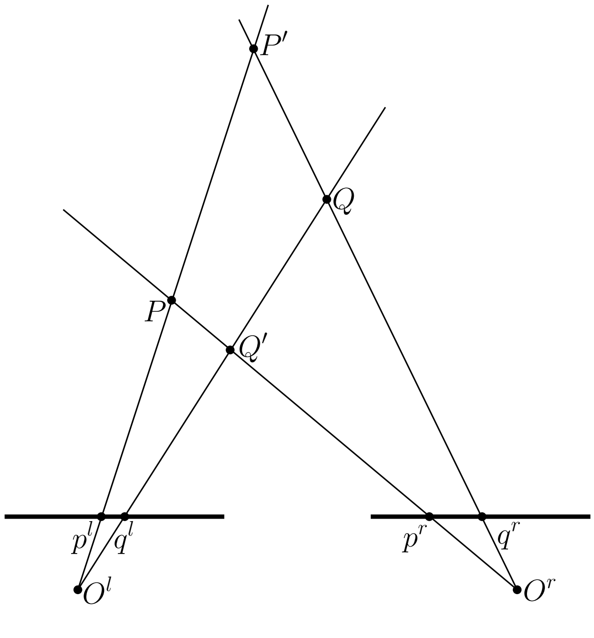


Abbildung 3.4: 3D - Rekonstruktion

sten liegt.. Für falsche Korrespondenzen (p^l, q^r) bzw. (q^l, p^r) erhielte man die Punkte P' bzw. Q' (siehe auch Kapitel 2.1.1).

An den Stellen, wo Korrespondenzen gefunden werden, erhält man auf diese Weise Tiefeninformationen. Zusammen mit dem zugehörigen Ort im projizierten Bild kann ein dreidimensionaler Szenenpunkt rekonstruiert werden. Anhand der an diesen Stellen aus der Bildstruktur extrahierten Orientierungsinformation wird dieser Szenepunkt zusätzlich mit einer 3D-Orientierung versehen. Diese erhält man, indem man die dem linken und rechten Bildpunkt zugeordnete Orientierungen zu 3D-Ebenen erweitert, und aus dem Schnitt dieser Ebenen im Raum die dreidimensionale Orientierung gewinnt.

Man erhält als beschreibende Elemente für die Szene 3D-Punkte. Diese werden mit der extrahierten 3D-Orientierungsinformation assoziiert. Diese sind darstellbar durch eine Menge von 3D-Liniensegmenten, wobei die Länge der Segmente willkürlich gewählt ist.

3.2.3 Probleme der 3D-Rekonstruktion

Aufgrund von den in Kapitel 2.1 beschriebenen Mängeln der Informationsextraktion werden wegen ungenauer oder falscher Korrespondenzermittlung eine unbekannte Anzahl von 3D-Liniensegmenten falsch rekonstruiert. Dies betrifft sowohl die absolute Lage im Raum, als auch die Orientierung.

Dabei können zwei Arten von Fehlern unterschieden werden. Aufgrund falsch bestimmter Korrespondenzen zwischen dem linken und dem rechten Bild ergeben sich 3D-Liniensegmente, deren Lage im Raum mit einem sehr großen Fehler behaftet ist. Für diese Liniensegmente liegt eine Gleichverteilung über den 3D-Raum vor.

Der andere auftretende Fehler ist in der leichten Ungenauigkeit der Informationsextraktion begründet. Die zur Rekonstruktion eines Liniensegmentes gefundene Korrespondenz zwischen linkem und rechtem Bildpunkt ist korrekt. Aber wegen der beschriebenen Schwächen wie z.B. Diskretisierung der Bildinformation oder Ungenauigkeit in der Bestimmung der Orientierung von Bildkanten ergibt sich für das rekonstruierte 3D-Segment ein leichter Fehler. Dieser betrifft sowohl die Position, als auch die Orientierung. Der Fehler dieser auf korrekten Korrespondenzen beruhenden 3D-Liniensegmente läßt sich durch eine Gaußsche Normalverteilung modellieren.

Fehlsegmente der ersten Art sollen in dieser Arbeit durch die Konfidenzbewertung herausgefunden werden. Zusätzlich wird ein Filterungsverfahren angegeben, welches die gleichverteilten Segmente mit großem Fehler aus der 3D-Repräsentation entfernt.

Dem durch eine Gaußverteilung darstellbaren kleinen Fehler in Form leichter Ungenauigkeiten in Lage und Orientierung von 3D-Liniensegmente wird durch die Mittelung über eng benachbarte Liniensegmente begegnet.

3.3 Optischer Fluß

Betrachtet man eine Folge von Bildern, die von einer nicht statischen Szene in kurzen zeitlichen Abständen aufgenommen wurden, so lassen sich von Bild zu Bild Verschiebungen detektieren. Man erhält systematische Bewegungsmuster, die von der Art der Kamerabewegung abhängen, oder von der Bewegung einzelner Objekte, die sich in der Szene bewegt haben.

Die Bewegungsmuster werden durch ein Feld von Bewegungsvektoren dargestellt. Diese geben für die einzelnen Bildpunkte (x, y) in der Bildebene an, wohin diese verschoben werden. Dieses Vektorfeld der Verschiebungen wird als optischer Fluß bezeichnet:

$$\mathbf{v} = (u, v) = (\frac{\partial x}{\partial t}, \frac{\partial y}{\partial t})^{\mathrm{T}} \in \mathbb{R}^2. \tag{3.22}$$

Für die Berechnung des optischen Flusses nach differentiell basierten Verfahren ([6]) werden die folgenden Annahmen gemacht:

- Der Grauwertverlauf ist stetig und differenzierbar in der Zeit und im Bild.

- Die Grauwerte der Projektion bewegter Objekte sind annähernd konstant, zumindest für eine kurze Zeitspanne.

Aufgrund dieser Annahmen gilt die folgende Differentialgleichung:

$$\frac{\partial I(x(t), y(t), t)}{\partial t} = 0, \tag{3.23}$$

wobei die Bildfunktion $I(x(t), y(t), t)$ den Intensitätsverlauf beschreibt. Anwendung der Kettenregel auf Gleichung (3.23) führt zu

$$\frac{\partial I}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial I}{\partial y}\frac{\partial y}{\partial t} + \frac{\partial I}{\partial t} = 0. \tag{3.24}$$

Mit folgender Notation für den Gradienten der Bildfunktion

$$\nabla I = \begin{pmatrix} \frac{\partial I}{\partial x} \\ \frac{\partial I}{\partial y} \end{pmatrix}, \tag{3.25}$$

der abgekürzten Schreibweise

$$I_t = \frac{\partial I}{\partial t} \tag{3.26}$$

und der Verwendung des Vektorfeldes $\mathbf{v}$ aus Gleichung (3.22) läßt sich die Gleichung (3.24) auch formulieren als

$$\nabla I \cdot \mathbf{v} + I_t = 0. \tag{3.27}$$

Die Gleichung (3.27) ist bekannt als *Image Brightness Constancy Equation (IBCE)*. Sie definiert eine Zwangsbedingung (Constraint) zwischen der Grauwertvariation im Raum und dem zeitlichen Bewegungsfeld.

Allerdings ist diese Zwangsbedingung für die Berechnung beider Komponenten von $\mathbf{v}$ nicht ausreichend, da die IBCE unterbesetzt ist[1]. Deshalb läßt sich nur die Normale des Flußvektors bestimmen. Es handelt sich hierbei um das Aperturproblem, welches in Kapitel 2.1.2 erläutert wurde.

Für die Berechnung des optischen Flusses sind eine Anzahl von unterschiedlichen Verfahren bekannt. In [6] werden eine Reihe der bekanntesten Methoden zur Berechnung des optischen Flusses analysiert. Die untersuchten Verfahren lassen sich in vier Gruppen unterteilen:

1. Differentielle Methoden
2. Korrelationsbasierte Methoden

[1] Bei der IBCE handelt es sich um eine lineare Gleichung mit den Unbekannten $\mathbf{v} = (u, v)$.

3. Energiebasierte Methoden
4. Phasenbasierte Methoden

Die dort präsentierten Ergebnisse für die einzelnen Verfahren wurden unter dem Gesichtspunkt untersucht, ein Verfahren zu wählen, welches für die Zwecke dieser Arbeit am besten geeignet ist. Das Augenmerk lag dabei darauf, welches Verfahren an den Objektkanten gute Flußwerte liefert[2]. Um zu erläutern, warum welches Verfahren als für diese Arbeit geeignet gewählt wurde, wird für jede dieser Gruppen exemplarisch ein Anwendungsbeispiel vorgestellt.

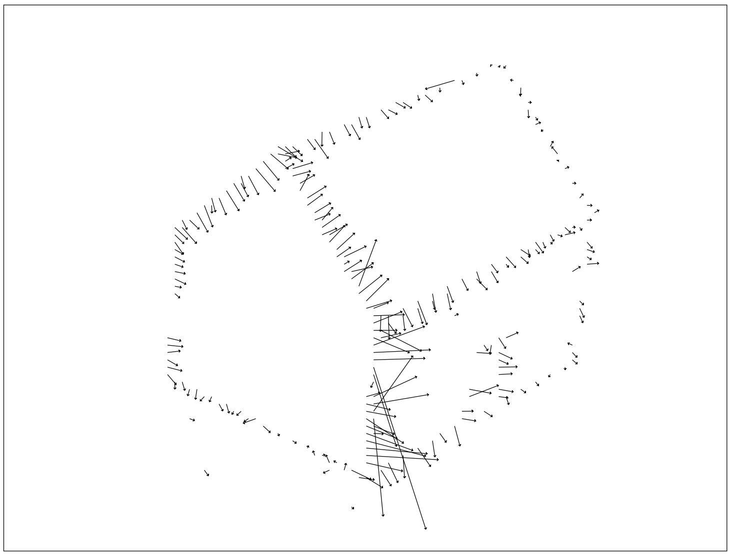


Abbildung 3.5: Verfahren von Horn und Schunck

3.3.1 Differentielle Methoden

Die differentiellen Methoden berechnen den optischen Fluß aus raum-zeitlichen Ableitungen der Grauwertintensitäten. Der Bildraum wird dafür als stetig und differenzierbar angenommen, sowohl im Raum als auch in der Zeit. Zusätzlich wird die Annahme gemacht, daß innerhalb kleiner Bildregionen ein konstantes Vektorfeld vorliegt (*global smoothness constraint*). Die Abbildung 3.5 zeigt das Flußfeld, welches nach dem Verfahren von Horn und Schunck

[2]Benutzt werden soll der optische Fluß ja für die Verfolgung von 2D-Liniensegmenten, siehe Kapitel 3.4.

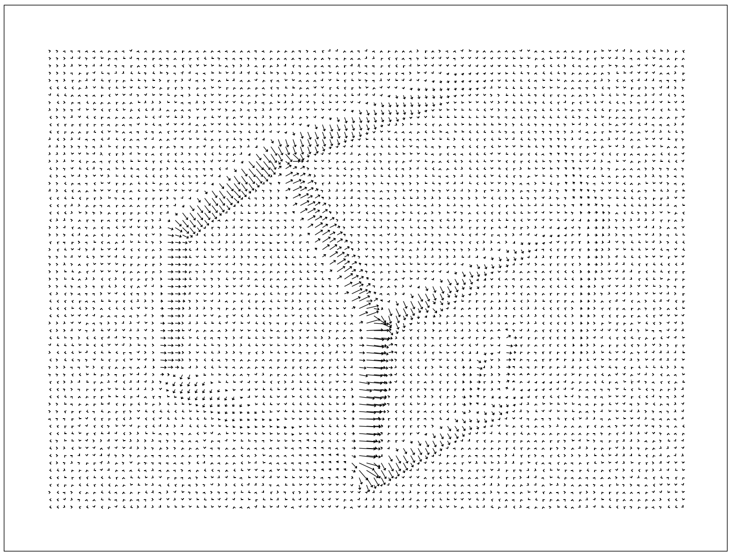


Abbildung 3.6: Verfahren von Nagel

[22] berechnet wurde. Es ist ein Verfahren, welches die IBCE (3.27) mit einem *global smoothness constraint* kombiniert.

Ein weiteres differentielles Verfahren ist von Nagel [33, 34] angegeben worden. Es handelt sich um eine Methode, welche Ableitungen zweiter Ordnung verwendet, in Kombination mit einer Alternative zum *global smoothness constraint*, dem *oriented smoothness constraint*. Hier wird an Gradienten mit hoher Intensität (z.B. Kanten) kein konstantes Vektorfeld gefordert. Dahinter steckt die Überlegung, daß an solchen Gradienten hoher Intensität eine Verdeckung eines Objektes durch ein anderes vorliegt. Da nicht davon auszugehen ist, daß sich beide Objekte in derselben Art und Weise bewegen, würde die Annahme eines konstanten Flußvektorfeldes an solchen Orten im Bild zu einer verfälschten Rekonstruktion des optischen Flusses führen. Nagel formuliert ein zu minimierendes Funktional, welches die Variation des Flusses senkrecht zu einem Gradienten abschwächt. Die Abb. 3.6 zeigt das berechnete Flußfeld. Im Vergleich mit Abb. 3.5 ist zu erkennen, daß die Flußvektoren bei Nagel senkrecht zu den Objektkanten eine geringere Variation aufweisen.

3.3.2 Korrelationsbasierte Methoden

Im Fall von starkem Rauschen, oder nur einer geringen Anzahl von Bildern, können differentielle Methoden eventuell nicht anwendbar sein. Bessere Ergebnisse können dann mit korrelationsbasierten Verfahren erzielt werden.

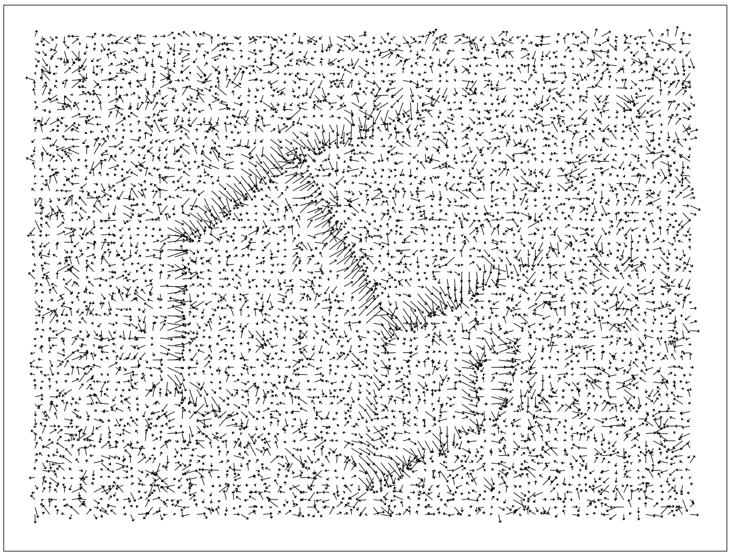


Abbildung 3.7: Verfahren von Anandan

Dabei wird versucht, die Position eines Punktes im nächsten Bild anhand von *region based matching* zu ermitteln. Es kann bei Annahme einer stetiger Bewegung der Suchraum für den Bildpunkt aufgrund vorhergehender Trajektion eingeschränkt werden. Zum Einsatz kommen u.a. SSD-Verfahren (sum of squared differences), um ein Abstandsmaß zu minimieren, um dadurch die neue Position des Bildpunktes zu finden. Aber auch der Kalman-Filter ([25]) kann eingesetzt werden.

In der Abb. 3.7 ist das Flußfeld nach dem Berechnungsverfahren von Anandan [3, 4] dargestellt. Das Verfahren verwendet eine Laplace-Pyramide und ein SSD-basiertes Matching.

3.3.3 Energiebasierte Methoden

Energiebasierte Methoden, auch frequenzbasierte Methoden genannt, verwenden Filter im Fourier-Raum für die Berechnung des optischen Flusses. Die von Barron, Fleet und Beauchemin [6] durchgeführten Untersuchungen des energiebasierten Verfahrens mit Gaborfiltern von Heeger [20] führten zu dem Ergebnis, daß diese Technik als nicht so zuverlässig anzusehen ist, wie die anderen untersuchten Verfahren. Das von Heeger vorgeschlagene Verfahren war demnach sehr empfindlich in Bezug auf die Ausgangsbedingungen, und konnte keine zuverlässigen Resultate erzielen.

3.3.4 Phasenbasierte Methoden

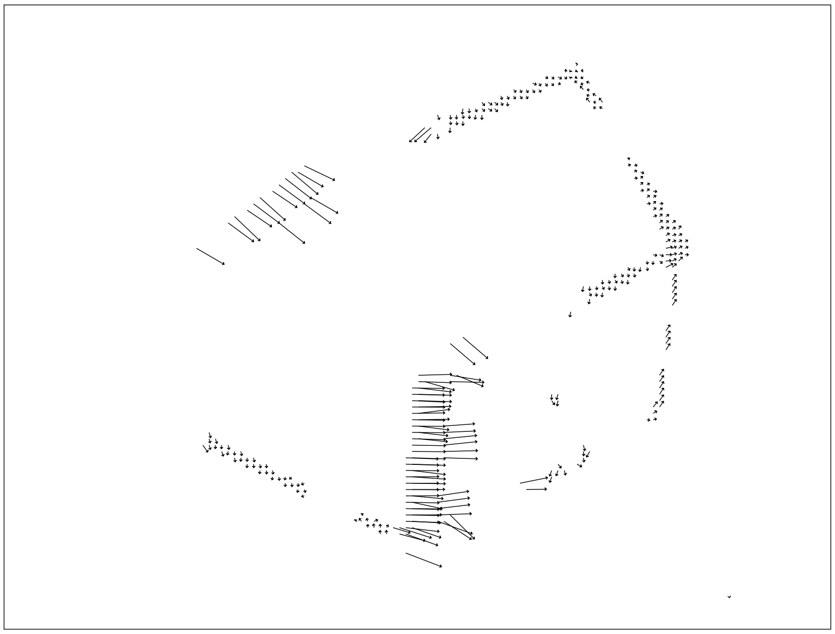


Abbildung 3.8: Verfahren von Fleet und Jepson

Phasenbasierte Methoden nutzen die Phaseninformationen der Ergebnisse von Bandpassfiltern, um die Verschiebungen in Bildsequenzen zu berechnen. Das Verfahren von Fleet und Jepson [13] basiert auf Gaborfiltern. In Abb. 3.8 ist das Ergebnis der Flußberechnung dargestellt. Barron, Fleet und Beauchemin [6] kommen zu dem Ergebnis, daß dieses Verfahren die besten Resultate aller untersuchten Methoden liefert. Die Ergebnisse dieses Verfahrens hier (siehe Abb. 3.8) weisen nicht eine so gute Qualität auf.

3.3.5 Fazit

Zusätzlich zu den Erkenntnissen über die Ergebnisqualität der unterschiedlichen Verfahren zur Berechnung des optischen Flusses wurden noch die Laufzeiten auf einer CPU mit 100 MHz gemessen:

Verfahren	Laufzeit	Abbildung
horn	2.66 sec	3.5
nagel	7.05 sec	3.6
anandan	48.33 sec	3.7
fleet	416.54 sec	3.8

Die differentiellen Verfahren und das korrelationsbasierte Verfahren von Anandan haben im Vergleich zu dem phasenbasierten Ansatz von Fleet und Jepson eine deutlich geringere Laufzeit. Obwohl dieses Verfahren in [6] als das Beste herausgestellt wurde, fand es wegen seiner langen Laufzeit und der für unsere Zwecke erzielten weniger guten Ergebnisse in dieser Arbeit keine Verwendung.

Gute Ergebnisse an Kanten erzielen die differentiellen Verfahren, während die Laufzeit gering ist. Gewählt für die Implementierung wurde das Verfahren von Nagel.

3.4 Verfolgung von 2D-Liniensegmenten

Für die Schätzung der Pose ([44]) ist ein 3D-Modell des Objektes zum Zeitpunkt t_1 und ein 2D-Modell zum Zeitpunkt t_2 nötig. Darüber hinaus muß die Korrespondenz zwischen den 3D-Liniensegmenten des 3D-Modells und den 2D-Liniensegmenten des 2D-Modells bekannt sein.

3.4.1 Gewinnung des 2D-Modells

Das 3D-Modell M^{3D} wird mit Hilfe der Stereo-Rekonstruktion berechnet. Um das 2D-Modell M^{2D} zu erhalten, wird das zum Zeitpunkt t_1 erstellte 3D-Modell auf die Bildebene projiziert. Dieses 2D-Modell ist allerdings ebenfalls für den Zeitpunkt t_1 gültig, und nicht, wie erforderlich, für t_2. Mit Hilfe des optischen Flusses werden die 2D-Liniensegmente des Modells M^{2D} in der Bildebene bewegt. Dadurch erhält man das für die Schätzung der Pose notwendige 2D-Modell M'^{2D} zum Zeitpunkt t_2.

Beispielhaft dargestellt ist dies in Abb. 3.9. Das 3D-Liniensegment L_{t_1} des 3D-Modells zum Zeitpunkt t_1 wird in die Bildebene B^l projiziert. Man erhält das 2D-Liniensegment l_{t_1}. Die an diesem beginnenden Vektoren $\vec{v}_{1...i}$ des optischen Flusses werden nun verwendet, um das Liniensegment l_{t_1} zu bewegen und damit das neue 2D-Liniensegment l_{t_2} zum Zeitpunkt t_2 zu erhalten. Dabei wird ein vereinfachtes Verfahren angewendet. Alle Flußvektoren, die auf dem Liniensegment beginnen, werden aufsummiert, und anschließend durch ihre Anzahl geteilt. Es findet also eine Mittelung über die Flußvektoren statt. Mit dem gemittelten Fluß wird der Aufpunkt des 2D-Liniensegmentes nun im Bild verschoben. Durch die Mittelung wird eine Robustheit bei der dem Liniensegment zugeordnete Verschiebung erreicht. Die Orientierung wird dabei beibehalten.

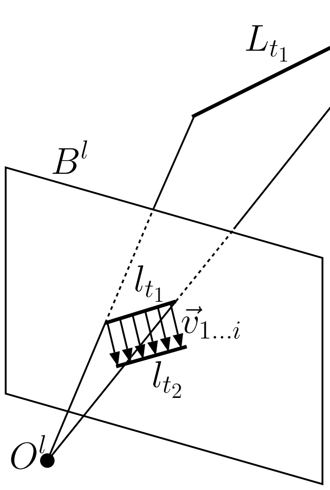


Abbildung 3.9: Verfolgung eines projizierten 3D-Liniensegmentes mit dem optischen Fluß

3.4.2 Probleme bei der Verfolgung von Linien

Aufgrund des Aperturproblems stehen die Flußvektoren unabhängig von der wirklichen Bewegungsrichtung einer Kante oder Linie im Bild immer senkrecht zu dieser. Daher wird das 2D-Liniensegment bei der Bewegung mittels der Flußvektoren nicht an den richtigen Ort verschoben. Es liegt lediglich auf der Geraden, die entsteht, wenn man das eigentlich korrekte 2D-Liniensegment verlängert. Dies tritt selbst bei optimal berechnetem Fluß auf. Da nur kleine Bewegungen auftreten, ist dieses Problem von geringer Relevanz.

3.5 Poseschätzung

Die Poseschätzung liefert die für die Akkumulation der Objektrepräsentation notwendigen Bewegungsparameter. Verwendet wird hier ein in [44, 41, 39, 38, 40, 16] beschriebenes Verfahren. Die Bewegung eines starren Objektes aus einer Position A in eine Position B wird aufgrund der folgenden Informationen geschätzt:

- 3D-Modell des Objektes in der Position A.

- Projizierte 2D-Punkte des Objektes in der Position B (siehe Abb. 3.10).

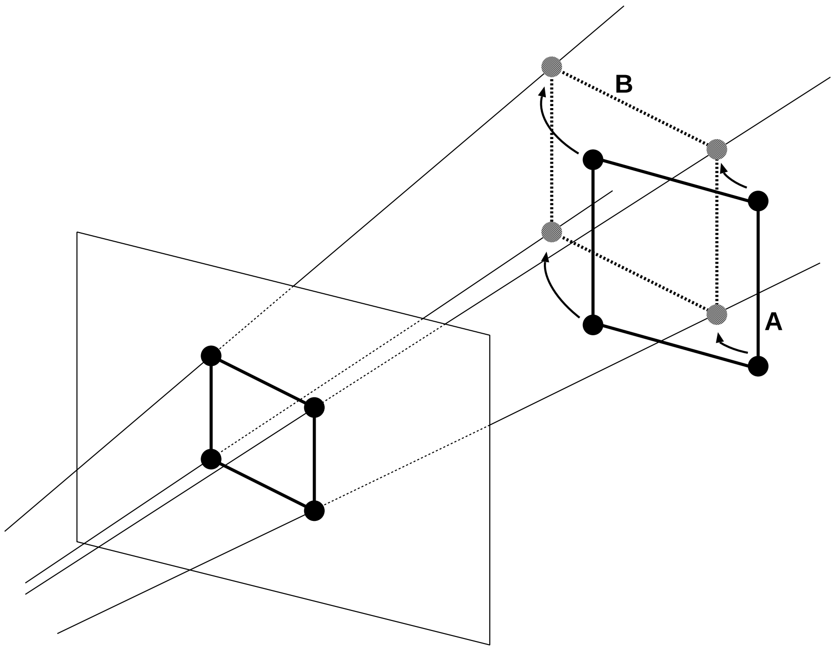


Abbildung 3.10: Idee der Poseschätzung

Aus den 2D-Punkten der Bildebene werden unter Verwendung der Projektionsmatrix die Projektionsgeraden berechnet. Um die Bewegungsparameter zu schätzen, werden die 3D-Modellpunkte mit den Projektionsgeraden in Beziehung gesetzt.

Die Kenntnis über die Korrespondenz zwischen Modellpunkten und Projektionsgeraden liefert der optische Fluß, der für das Verfolgen der 2D-Linien verwendet wird (siehe Kap. 3.4). Gesucht wird die Transformation, welche jeweils den 3D-Modellpunkt auf die korrespondierende Projektionsgerade abbildet (siehe Abb. 3.10).

3.5.1 Punkt zu Linie Zwangsbedingung

Gegeben sind für das Schätzen der Pose der Modellpunkt

$$X = \begin{pmatrix} \xi^1 \\ \xi^2 \\ \xi^3 \\ 1 \end{pmatrix}, \qquad (3.28)$$

und der projizierte Punkt Y_A des Modells in einer veränderten Lage (siehe Abb. 3.11). Da die Kamera kalibriert ist, läßt sich die Projektionsgerade L

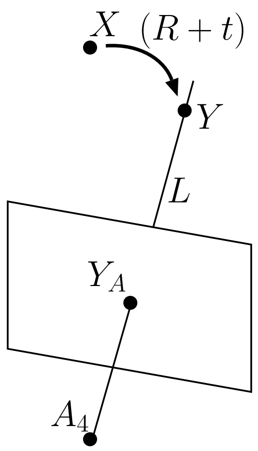


Abbildung 3.11: Grundlage der Poseschätzung

vom Punkt Y_A durch das optische Zentrum A_4 der Kamera angeben:

$$L = A_4 \wedge Y_A. \tag{3.29}$$

Gesucht ist nun die Transformation des Punktes X auf einen Punkt Y auf der Projektionsgeraden L. Die Transformation setzt sich zusammen aus Rotation R und Translation $t = (\tau^1, \tau^2, \tau^3, 0)$. Der Punkt Y berechnet sich damit durch:

$$Y = RX\tilde{R} + t. \tag{3.30}$$

Der Punkt Y soll auf der Linie L liegen. Dies ist die Zwangsbedingung für die Poseschätzung. Die Zwangsgleichung lautet also:

$$L \wedge Y = 0. \tag{3.31}$$

Einsetzen von Y aus Gleichung (3.30) und Umformen ergibt die Zwangsgleichung:

$$\begin{aligned} L \wedge (RX\tilde{R} + t) &= 0 \\ L \wedge RX\tilde{R} + L \wedge t &= 0. \end{aligned} \tag{3.32}$$

Eine Rotation R um einen Winkel θ in der normierten Rotationsebene $\hat{U}$ hat das folgende Aussehen:

$$R = \cos\frac{\theta}{2} - \sin\frac{\theta}{2}\hat{U}. \tag{3.33}$$

Die Taylorentwicklung von cos an der Stelle Null bis zum Grad eins ergibt:

$$
\begin{aligned}
\cos x &= \sum_{n=0}^{\infty} \frac{\cos^{(n)} 0}{n!} x^n \\
&= \frac{\cos 0}{0!} x^0 + \frac{-\cos 0}{1!} x^1 + \frac{-\cos 0}{2!} x^2 + \ldots \\
&= 1 + 0 - \frac{x^1}{2!} + \ldots \\
&\approx 1
\end{aligned} \tag{3.34}
$$

Die Taylorentwicklung von sin an der Stelle Null bis zum Grad eins ergibt:

$$
\begin{aligned}
\sin x &= \sum_{n=0}^{\infty} \frac{\sin^{(n)} 0}{n!} x^n \\
&= \frac{\sin 0}{0!} x^0 + \frac{\cos 0}{1!} x^1 + \frac{-\sin 0}{2!} x^2 + \ldots \\
&= 0 + x + 0 + \ldots \\
&\approx x
\end{aligned} \tag{3.35}
$$

Mit den Approximationen für sin (3.35) und cos (3.34) läßt sich die Rotation R aus Gleichung (3.33) für kleine Winkel annähern durch:

$$R = 1 - \frac{\theta}{2}\hat{U}. \tag{3.36}$$

Rotiert man den Vektor X mit dem approximierten R aus Gleichung (3.36), so erhält man

$$
\begin{aligned}
RX\tilde{R} &= (1 - \frac{\theta}{2}\hat{U})X(1 + \frac{\theta}{2}\hat{U}) \\
&= X + \frac{\theta}{2}x\hat{U} - \frac{\theta}{2}\hat{U}X - \underbrace{\frac{\theta^2}{2}\hat{U}X\hat{U}}_{\text{Grad 2}} \\
&= X + \frac{\theta}{2}(X\hat{U} - \hat{U}X) \\
RX\tilde{R} &= X + \theta X \cdot \hat{U},
\end{aligned} \tag{3.37}
$$

wobei der Term vom Grad zwei weggelassen werden kann, da nur die Approximation bis zum Grad eins betrachtet wird.

Einsetzen von (3.37) in die Zwangsgleichung (3.32) und Ersetzen der normierten Rotationsebene $\hat{U}$ durch $u = \theta\hat{U}$ ergibt die neue Zwangsgleichung:

$$
\begin{aligned}
L \wedge RX\tilde{R} + L \wedge t &= 0 \\
L \wedge (X + \theta X \cdot \hat{U}) + L \wedge t &= 0 \\
L \wedge X + L \wedge [(X \cdot u) + t] &= 0.
\end{aligned} \tag{3.38}
$$

Die Zwangsgleichung (3.38) läßt sich dualisieren, und es ergibt sich die folgende Form:

$$\begin{aligned} X \cdot L^* + [(X \cdot u) + t] \cdot L^* &= 0 \\ L^* \cdot [(X \cdot u) + t] &= X \cdot L^* \\ \underbrace{L^* \cdot (X \cdot u)}_{A} + \underbrace{L^* \cdot t}_{B} &= \underbrace{X \cdot L^*}_{C}. \end{aligned} \tag{3.39}$$

Die in Gleichung (3.39) mit A, B und C bezeichneten Teilterme sind Vektoren. Die Rotationsebene u ist gegeben durch

$$u = \theta \hat{U} = \theta(\sigma^1 e_2 e_3 + \sigma^2 e_3 e_1 + \sigma^3 e_1 e_2) = \theta \begin{pmatrix} \sigma^1 \\ \sigma^2 \\ \sigma^3 \end{pmatrix}, \tag{3.40}$$

der Translationsvektor t hat das folgende Aussehen:

$$t = (\tau^1 e_1 + \tau^2 e_2 + \tau^3 e_3) = \begin{pmatrix} \tau^1 \\ \tau^2 \\ \tau^3 \end{pmatrix}, \tag{3.41}$$

und das Dual der Projektionsgeraden, L^*, ist von folgender Gestalt:

$$L^* = \lambda^i e_4 e_i + \bar{\lambda}^{i_1} e_{i_2} e_{i_3}. \tag{3.42}$$

In Matrixform geschrieben, läßt sich der Vektor A aus Gleichung (3.39) auch berechnen durch:

$$A = \underbrace{\begin{bmatrix} -(\bar{\lambda}^2\xi^2 + \bar{\lambda}^3\xi^3) & \bar{\lambda}^2\xi^1 & \bar{\lambda}^3\xi^1 \\ \bar{\lambda}^1\xi^2 & -(\bar{\lambda}^3\xi^3 + \bar{\lambda}^1\xi^1) & \bar{\lambda}^3\xi^2 \\ \bar{\lambda}^1\xi^3 & \bar{\lambda}^2\xi^3 & -(\bar{\lambda}^1\xi^1 + \bar{\lambda}^2\xi^2) \\ \lambda^3\xi^2 - \lambda^3\xi^3 & \lambda^1\xi^3 - \lambda^3\xi^1 & \lambda^2\xi^1 - \lambda^1\xi^2 \end{bmatrix}}_{M} \begin{pmatrix} \theta\sigma^1 \\ \theta\sigma^2 \\ \theta\sigma^3 \end{pmatrix} \tag{3.43}$$

und der Vektor B wird durch die folgende Matrixgleichung berechnet:

$$B = \underbrace{\begin{bmatrix} 0 & \bar{\lambda}^3 & -\bar{\lambda}^2 \\ -\bar{\lambda}^3 & 0 & \bar{\lambda}^1 \\ \bar{\lambda}^2 & -\bar{\lambda}^1 & 0 \\ \lambda^1 & \lambda^2 & \lambda^3 \end{bmatrix}}_{T} \begin{pmatrix} \tau^1 \\ \tau^2 \\ \tau^3 \end{pmatrix}. \tag{3.44}$$

Mit den Bezeichnern M bzw. T für die Matrizen aus den Gleichungen (3.43) bzw. (3.44) läßt sich die Gleichung (3.39) folgendermaßen schreiben:

$$\begin{aligned} Mu + Tt &= C \\ \Leftrightarrow \begin{bmatrix} M & T \end{bmatrix} \begin{pmatrix} u \\ t \end{pmatrix} &= C. \end{aligned} \tag{3.45}$$

Die Matrizen M bzw. T aus den Gleichungen (3.43) bzw. (3.44) werden berechnet aus den gegebenen Punkten X (siehe Gleichung (3.28)) und Y_A (siehe Gleichung (3.29)). Gleiches gilt für den Vektor C. Damit ist Gleichung (3.45) ein unterbestimmtes Gleichungssystem mit den sechs unbekannten Transformationsparametern von u und t.

Zur Lösung von (3.45) müssen noch mindestens weitere zwei Objektpunkt-Bildpunkt-Korrespondenzen hinzugenommen werden. Das Gleichungssystem erweitert sich dadurch zu

$$\begin{bmatrix} M_1 & T_1 \\ M_2 & T_2 \\ M_3 & T_3 \end{bmatrix} \begin{pmatrix} u \\ t \end{pmatrix} = \begin{pmatrix} C_1 \\ C_2 \\ C_3 \end{pmatrix} \tag{3.46}$$

und kann damit gelöst werden. M_i, T_i bzw. C_i, $i \in \{1, 2, 3\}$ sind dabei die aus den drei Korrespondenzen berechneten Matrizen bzw. Vektoren. Der Lösungsvektor besteht aus den Bewegungsparametern u und t, welche die geschätzte 2D-3D-Bewegung beschreiben.

Durch die Approximationen (3.34) und (3.35) lassen sich nur Bewegungen mit kleinen Winkeln schätzen. Ein iteratives Verfahren kann auch Posen mit großen Bewegungswinkeln schätzen [44]. In dieser Arbeit wird die Bewegung mit Hilfe des optischen Flusses geschätzt, und daher sind nur Bewegungen mit kleinem Winkel gegeben.

Da für die Stereorekonstruktion in dieser Arbeit eine Stereo-Videosequenz verwendet wird, läßt sich die Poseschätzung auf die linke und die rechte Sequenz anwenden. Man erhält zwei Schätzungen der Bewegungsparameter für dieselbe beobachtete Bewegung. Der Abschnitt 3.7 beschreibt ein Verfahren, das eine Mittelung der beiden Bewegungsparameter durchführt.

3.6 Filterung des 3D-Modells

Aufgrund der Probleme bei der Extraktion von Informationen aus Bilder ist das mit der 3D-Stereorekonstruktion gewonnene 3D-Modell fehlerbehaftet. Es gibt Liniensegmente, welche aufgrund einer falschen Korrespondenz an der falschen Position im Raum rekonstriert wurden. Ebenso liegen bei vielen Segmenten ungenaue oder falsche Orientierungen vor.

Für die Schätzung der Pose ist ein möglichst genaues 3D-Modell zum Erzielen eines gutes Ergebnisses notwendig. Liniensegmente, die an falschen Stellen im Raum rekonstruiert wurden, wirken sich sehr störend aus. Es ist auf der anderen Seite nicht nötig, ein vollständiges 3D-Modell zu verwenden. Es genügt, eine ausreichende Anzahl von gut rekonstruierten und linear unabhängigen Segmenten zu verwenden, um eine gute Pose schätzen zu können.

Die Liniensegmente, für welche der Fehler durch eine Gaußsche Normalverteilung modelliert werden kann, ermöglichen aufgrund dieser Verteilung eine gute Schätzung der Pose.

3.6.1 Grundidee der Filterung

Für eine gute Poseschätzung auf der Basis fehlerhafter 3D-Modelle wurde als Vorverarbeitung ein Verfahren eingeführt, um das aus der Stereo-Rekonstruktion erhaltene fehlerhafte 3D-Modell zu filtern. Als Ergebnis der Filterung sollen nur möglichst gut rekonstruierte Liniensegmente des betrachteten Objektes übrig bleiben.

Jedes 3D-Liniensegment für sich alleine genommen enthält wenig Information darüber, wie gut es das Objekt repräsentiert. Es sind zwar Informationen über die einem Liniensegment zugrunde liegende Korrespondenz vorhanden, doch sind diese nicht ausreichend. Um sichere Information zu extrahieren, müssen die Segmente im Zusammenhang betrachtet werden. Die Menge der 3D-Liniensegmente wird daraufhin untersucht, inwieweit eine Anzahl von Segmenten sowohl von ihrer Lage im Raum als auch von ihrer Orientierung her auf einer Geraden liegen. Da Kollinearität häufig auftritt (siehe Krüger [27], Geisler et al. [15]), ist dies ein gutes Kriterium. Denn es ist unwahrscheinliche, daß zwei oder sogar mehrere 3D-Liniensegmente zufällig kollinear sind. Segmente, für die dieses Kollinearitätskriterium zutrifft, werden für die Poseschätzung beibehalten. Alle anderen fallen heraus. Die gefundenen Liniensegmente sind also zu einer Anzahl von anderen Segmenten kollinear. Je größer die Anzahl von jeweils kollinearen Segmenten, desto höher ist die Wahrscheinlichkeit, daß es sich um korrekt rekonstruierte 3D-Objektpunkte handelt. Die Qualität der Filterung des 3D-Modells läßt sich durch die Angabe der Mindestgröße einer Kollinearitätsgruppe von Liniensegmenten direkt beeinflussen.

Um kollineare 3D-Liniensegmente zu finden, wurde zunächst eine Metrik definiert, welche aussagt, ob zwei Segmente auf einer Geraden im Raum liegen. Ein iterativer Algorithmus sortiert dann auf Basis dieser Metrik die Liniensegmente in kollineare Gruppen, und verwirft die Segmente, die nicht zu ausreichend großen Gruppen kombiniert werden können. Das von uns benutzte Kollinearitätskriterium basiert auf einer heuristischen Definition von Kollinearität. Ziel aktueller Forschung ist es, diese heuristischen Annahmen durch statistische Messungen zu ersetzen (Krüger und Wörgötter [30], Sigman et al. [43], Geisler et al. [15], Brunswick und Kamiya [9]).

3.6.2 Abstandsmaße

Es wurden drei verschiedene Abstandsmaße zwischen 3D-Liniensegmenten definiert. Sie beruhen auf jeweils unterschiedlichen Parametern der Segmente, und eine geeignete Kombination erlaubt es, eine Aussage darüber zu machen, ob Segmente kollinear sind.

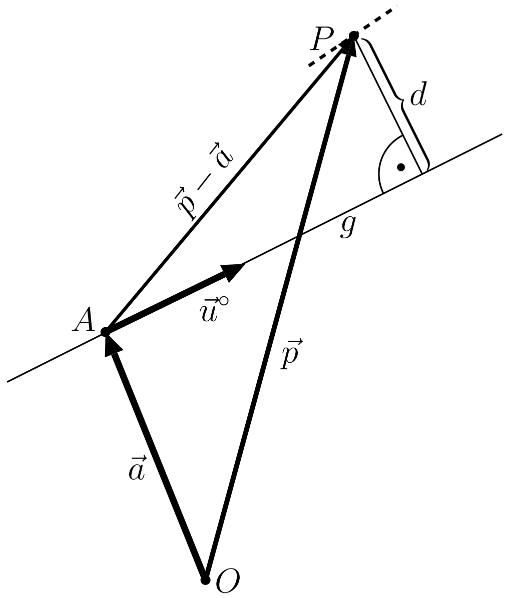


Abbildung 3.12: Abstand Punkt - Gerade im $\mathbb{R}^3$

Abstand zwischen Punkt und Gerade: Gegeben sind zwei 3D-Liniensegmente. Eines hat den Aufpunkt A und die Orientierungsrichtung $\vec{u}^\circ$, von dem anderen ist nur der Aufpunkt P von Bedeutung (siehe Abb. 3.12). Punkt A und Orientierung $\vec{u}^\circ$ geben die Gerade g im Raum vor. Berechnet wird der Abstand d, den Punkt P von der Geraden g hat. Die Metrik ist folgendermaßen definiert:

$$d = |\vec{u}^\circ \times (\vec{p} - \vec{a})|. \tag{3.47}$$

Anschaulich wird um die durch Aufpunkt und Orientierungsvektor gegebene Gerade g ein Zylinder mit Radius d konstruiert. Liniensegmente, deren Aufpunkt außerhalb des Zylinders liegen, sind nicht zum Ausgangssegment kollinear. Sie werden verworfen.

Unterschied in der Orientierung: Es wird der Unterschied in der Orientierung zweier 3D-Liniensegmente, gegeben durch Aufpunkt A und

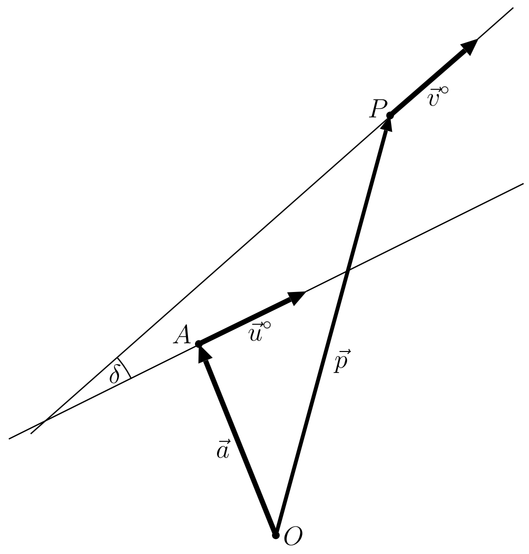


Abbildung 3.13: Orientierungsdifferenz zweier Liniensegmente

Orientierungsvektor $\vec{u}^{\circ}$ bzw. Aufpunkt P und Orientierung $\vec{v}^{\circ}$ (siehe Abb. 3.13) berechnet. Die Metrik ist wie folgt angegeben:

$$\delta = |\vec{u}^{\circ} - \vec{v}^{\circ}|. \tag{3.48}$$

Weichen die Orientierungen zu stark voneinander ab, so wird das Liniensegment als nicht kollinear angesehen und verworfen.

Unterschied zwischen Richtung und Orientierung: Um eine gewisse Variabilität in Bezug auf leicht gekrümmte Kanten zu erreichen, wurde diese dritte Metrik eingeführt. Sie dient ebenfalls dazu, parallele Liniensegmente zu erkennen und auszuschließen. Dazu wird der Vektor

$$\vec{r} = \vec{p} - \vec{a} \tag{3.49}$$

zwischen den Aufpunkten der zu vergleichenden Segmente berechnet (siehe Abb. 3.14). Der Vektor $\vec{r}$ wird normiert ($\vec{r}^{\circ}$) und mit dem Orientierungsvektor $\vec{u}^{\circ}$ des Liniensegments verglichen. Über den Winkel δ wird ein Doppelkegel definiert, in dem passende Liniensegmente liegen müssen, um das Kriterium der Kollinearität zu erfüllen.

Dieses Maß erlaubt es, daß weiter entfernt liegende Segmente einen größeren Abstand von der Geraden g haben dürfen. Umgekehrt kann es eingesetzt werden, um parallel liegende Liniensegmente zu detektieren und zu verwerfen. In Abb. 3.15 ist das Liniensegment l_2 parallel zu

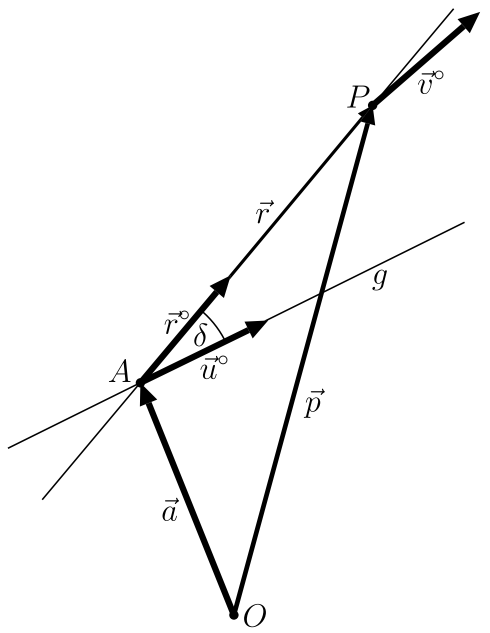


Abbildung 3.14: Differenz zwischen Orientierung des Liniensegments ($\vec{u}^\circ$) und Richtung zum zweiten Liniensegment ($\vec{r}^\circ$)

Segment l_1. Da es außerhalb des durch die Metrik definierten Kegels liegt, wird es verworfen. Die Segmente l_3 und l_4 liegen innerhalb des Kegels, und werden beibehalten.

3.6.3 Algorithmus

Auf der Basis der drei Abstandsmaße wurde ein Algorithmus entworfen, um die gewünschte Filterung der 3D-Liniensegmente zu erreichen. Der Algorithmus sortiert die Segmente zunächst in kollineare Gruppen. Dazu wird aus der Gesamtmenge LS der Liniensegmente ein noch nicht behandeltes Segment $l_1 \in LS$ genommen, und mit allen übrigen, ebenfalls noch nicht behandelten Segmenten $l_i \in LS \backslash \{l_1\}$, verglichen. Dabei wird folgendes überprüft:

1. Ob der Abstand zwischen dem Aufpunkt von l_i und der durch l_1 definierten Geraden einen gewissen Schwellwert nicht überschreitet.

2. Daß die Orientierungen der beiden Segmente sich nicht um mehr als einen vorgegebenen Betrag voneinander unterscheiden.

3. Daß der Vektor zwischen den Aufpunkten von l_1 und l_i und die Orientierung von l_1 nicht zu stark voneinander abweichen.

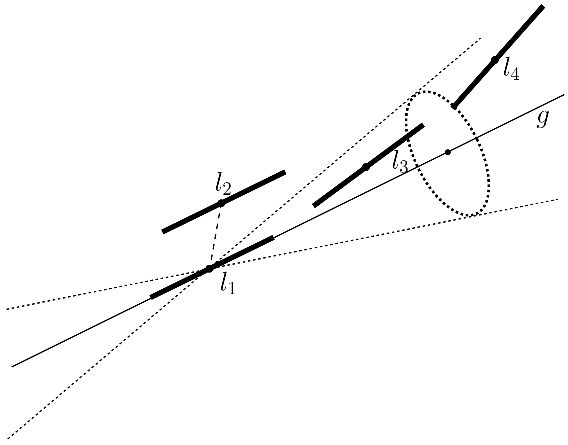


Abbildung 3.15: Detektierung parallel liegender Liniensegmente

4. Wenn schon eine Gruppe mit mehr als einem Liniensegment vorliegt, dann wird nochmals überprüft, ob das neue Segment l_i eventuel parallel zu einem der bereits vorhandenen Segmente ist. Nur wenn dies nicht zutrifft, wird das Liniensegment l_i in die Gruppe aufgenommen.

Treffen die Kriterien zu, daß ein Segment l_i zu l_1 kollinear ist, so wird dieses Segment also mit l_1 in eine Gruppe sortiert, sofern es nicht zu einem bereits gefundenen Segment parallel ist. Es ergeben sich bei diesem Verfahren Gruppen kollinearer Liniensegmente. Diese Gruppen können eine unterschiedliche Anzahl von Segmenten enthalten. Alle Gruppen mit Größe eins können sofort verworfen werden, denn sie enthalten isolierte Segmente, die zu keinem anderen kollinear sind. Je größer die Mindestanzahl gesetzt wird, desto zuverlässiger beschreiben die dort enthaltenen Liniensegmente das Objekt. Damit läßt sich direkt die Qualität der nach der Filterung übriggebliebenen Liniensegmente beeinflussen. Abgewägt werden muß an dieser Stelle, wie hoch die Mindestgruppengröße gewählt wird, denn es müssen für die Schätzung der Pose noch genügend Liniensegmente übrig bleiben.

3.7 Fusionieren der Bewegungsparameter

Dadurch, daß ein Stereosystem Verwendung findet, läßt sich die Poseschätzung für das linke und das rechte Bild durchführen. Das Ergebnis sind zwei Rotationsmatrizen A und B, und zwei Translationsvektoren t_l, t_r. Da nur eine Bewegung durchgeführt wird, bietet es sich an, das Mittel F, t_f der beiden geschätzten Bewegungsparameter zu bilden.

Für die Mittelung ist die Matrix zur Beschreibung von Rotationen nicht gut geeignet (Pennec [35]). Die gemittelte Rotationsmatrix wäre nicht mehr Element der speziellen orthogonalen Gruppe $SO(n)$. In der Darstellung durch Rotationsachse und -winkel läßt sich die Fusion durch eine Mittelung von Achse und Winkel erreichen, ohne $SO(n)$ zu verlassen. Für die Ausführung der fusionierten Rotation ist jedoch die Matrixform besser geeignet. Die folgenden Schritte müssen demzufolge durchgeführt werden:

1. Gewinnung der Rotationsachsen und der Rotationswinkel aus den Matrizen.

2. Bildung des Mittels der Achsen und der Winkel.

3. Rückformung der Darstellung durch Achse/Winkel in die Matrixform.

Der durch Rotationsmatrix A beschriebene Rotationswinkel θ_a läßt sich berechnen durch:

$$\theta_a = \arccos(\frac{(a_{11} + a_{22} + a_{33} - 1)}{2}), \tag{3.50}$$

und die normierte Rotationsachse $\hat{a}$ berechnet sich durch:

$$\hat{a} = \frac{1}{2\sin\theta_a} \begin{pmatrix} a_{32} - a_{23} \\ a_{13} - a_{31} \\ a_{21} - a_{12} \end{pmatrix}. \tag{3.51}$$

Der von Rotationsmatrix B beschriebene Rotationwinkel θ_b und die normierte Rotationsachse $\hat{b}$ werden analog gewonnen. Der mittlere Winkel θ_f ergibt sich durch

$$\theta_f = \frac{\theta_a + \theta_b}{2}, \tag{3.52}$$

und die mittlere Rotationsachse f durch

$$f = \frac{1}{2}(\hat{a} + \hat{b}). \tag{3.53}$$

Durch Normalisieren von f erhält man die normalisierte fusionierte Rotationsachse $\hat{f}$. Die Rotationsmatrix F berechnet sich aus der Achse $\hat{f}$ und dem Winkel θ_f wie folgt:

$$F = (1 - \frac{|p|^2}{2})I + \frac{1}{2}(pp^{\mathrm{T}} + \alpha S), \tag{3.54}$$

wobei

$$\begin{aligned} p &= 2\sin\frac{\theta_f}{2}\hat{f} = \left(\begin{array}{ccc} p_x & p_Y & p_z \end{array} \right)^{\mathrm{T}}, \\ \alpha &= \sqrt{4-|p|^2} \qquad \text{und} \\ S &= \begin{bmatrix} 0 & -p_z & p_y \\ p_z & 0 & -p_x \\ -p_y & p_x & 0 \end{bmatrix}. \end{aligned} \tag{3.55}$$

Die Mittelung der Translationen t_l, t_r geschieht durch einfache Mittelwertbildung der beiden Vektoren.

3.8 Akkumulieren der Objektrepräsentation

Verschiedene Ansichten eines Objektes führen zu unterschiedlich extrahierten Repräsentationen. Ist jeweils die Bewegung bekannt, welche aus den verschiedenen Ansichten zu der jeweils vorhergehenden, oder zur ursprünglichen Objektansicht führt, lassen sich die verschiedenen Repräsentationen aufakkumulieren (Krüger, Ackermann [28, 1]). Fehler, die in einzelnen Extraktionen auftreten, werden eliminiert. Dagegen stabilisieren sich korrekt rekonstruierte Teile der Repräsentation, je öfter sie in den unterschiedlichen Ansichten extrahiert wurden. Ackermann [28, 1] hat ein solches Verfahren entwickelt.

Es findet eine Initialisierung mit der ersten extrahierten 3D-Repräsentation statt. In einem iterativen Algorithmus werden weitere Repräsentationen in unterschiedlichen Ansichten jeweils so transformiert, daß sie der Ansicht entsprechen, in der die Initialisierung stattgefunden hat. Zwischen den 3D-Liniensegmenten der initialen Repräsentation und denen der neu gewonnenen und transformierten Repräsentation werden Korrespondenzen gesucht. Dazu wurden zwei Metriken definiert, die es erlauben, eine Aussage zu treffen, ob zwei Liniensegmente miteinander korrespondieren.

Es handelt sich um die Metrik d_o für die Orientierungsdifferenz, und d_p für die Abweichung der Position im Raum, welche einzeln betrachtet werden. Für zwei Liniensegmente s und s' und ihren zugehörigen Orientierungsvektoren $\vec{v}$ und $\vec{v'}$ ist die Metrik d_o wie folgt definiert:

$$d_o := \arccos(\vec{v} \cdot \vec{v'}). \tag{3.56}$$

Mittels eines Schwellwertes ξ wird entschieden, ob zwei Liniensegmente miteinander korrespondieren, oder ob ihre Orientierung zu stark voneinander abweicht, als daß sie sinnvoll zu einem neuen Segment zusammengefaßt werden könnten.

Zusätzlich wird mittels der Metrik d_p überprüft, ob die Position von s und s' im Raum übereinstimmt. Die angegebene Metrik ist in die Richtung des einen Liniensegmentes gestreckt, da korrespondierende Liniensegmente entlang ihrer Orientierungsachse verschoben sein können. Der Schwellwert ω darf senkrecht zur Orientierung nicht überschritten werden, in der Parallelen beträgt der Wert $\omega+\frac{l}{2}$, wobei l die Länge des Liniensegmentes s ist. Es ergibt sich ein elliptischer Suchraum.

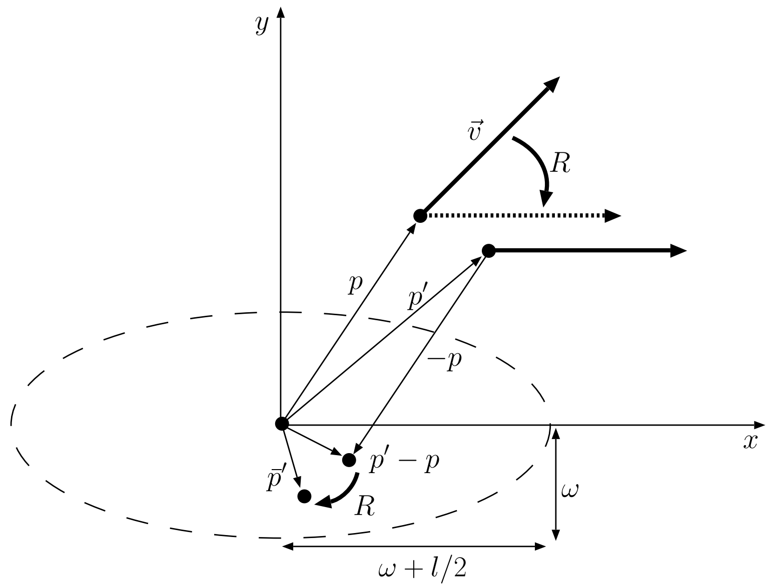


Abbildung 3.16: Elliptischer Suchraum für die Positionsdifferenz

Zu diesem Zweck wird das Liniensegment s mit seinem Mittelpunkt p auf den Ursprung des Koordinatensystems verschoben, und eine Rotation durchgeführt, so daß der Orientierungsvektor $\vec{v}$ von p mit der x-Achse des Koordinatensystems zusammenfällt (siehe Abb. 3.16):

$$\bar{p} = R(p-p)\tilde{R}(=\vec{0}), \tag{3.57}$$

wobei R der Rotor ist, welcher $\vec{v}$ auf die x-Achse e_1 abbildet:

$$R = \frac{(\vec{v}+e_1)\vec{v}}{|\vec{v}+e_1|} \tag{3.58}$$

Dieselbe Transformation wird auf den Punkt p' von Liniensegment s' angewendet:

$$\bar{p}' = R(p'-p)\tilde{R}. \tag{3.59}$$

Die Berechnung des Abstandes von p nach p' mit der geforderten Streckung in Richtung der Orientierung von s entspricht damit der Abstandsberechnung vom Ursprung des Koordinatensystems zum Transformierten Punkt $\bar{p}'$, und die Streckung erfolgt entlang der x-Achse. Die Abstandsfunktion $\mathbf{f}_p(s, s')$ ist dann definiert als

$$\mathbf{f}_p(s, s') := \sqrt{\left(\frac{\omega}{\omega + \frac{l}{2}} \bar{x}'\right)^2 + \bar{y}'^2 + \bar{z}'^2}, \tag{3.60}$$

wobei $\bar{x}', \bar{y}', \bar{z}'$ die Koordinaten des nach Gleichung (3.59) transformierten Punktes $\bar{p}'$ sind.

Die Metrik d_p ist mit Gleichung (3.60) wie folgt gegeben:

$$d_p(s, s') := \min\{\mathbf{f}_p(s, s'), \mathbf{f}_p(s', s)\}. \tag{3.61}$$

Eine Korrespondenz ist gegeben, wenn für zwei Liniensegmente s und s' die Kriterien $d_o(s, s') < \xi$ und $d_p(s, s') < \omega$ erfüllt sind. Erfüllen mehrere Segmente diese Kriterien, so wird das Paar gewählt, dessen Positionsdifferenz d_p minimal ist.

Wird eine Korrespondenz gefunden, so findet das Liniensegment der initialen Repräsentation eine Bestätigung durch ein Liniensegment, welches in einer anderen Ansicht extrahiert wurde. Diese Bestätigung wird dem Liniensegment in Form eines Konfidenzwertes zugeordnet. Bei gefundenen Bestätigungen wird dieser Wert erhöht, im anderen Fall erniedrigt.

Während der Iterationsschritte akkumuliert sich so die stabilere Objektrepräsentation anhand der Liniensegmente, die einen höheren Konfidenzwert erhalten. Damit werden die gleichverteilten Liniensegmente, die wegen zugrunde liegender falscher Korrespondenz einen großen Positionsfehler besitzen (siehe Kap. 3.2.3) mit eine niedrige Bewertung versehen.

Eine Mittelung über eng benachbarte Liniensegmente mit hoher Konfidenz wird abschließend durchgeführt. Da der Fehler dieser Segmente durch eine Gaußsche Normalverteilung modelliert werden kann, wird durch die Mittelung eine Stabilisierung erreicht.

Kapitel 4

Interaktion der Module

In diesem Kapitel wird die Interaktion der Module, die in Kapitel 3 vorgestellt wurden, erläutert. Durch diese Interaktion der einzelnen Module wird eine erhöhte Stabilitä der gewonnenen Informationen erreicht. Fehler der Stereorekonstruktion werden erkannt und können dadurch entfernt werden, Ungenauigkeiten erfahren eine Verbesserung. Beschrieben wird dabei das im Rahmen dieser Arbeit implementierte Verfahren, welches die in Kapitel 2 geschilderte Problemstellung löst.

Es werden zunächst die Rahmenbedingungen erläutert, die mit der gewählten Problemlösung einhergehen. Der entworfene Algorithmus wird dann anhand einer formalen mathematischen Beschreibung vorgestellt.

4.1 Rahmenbedingungen

Die Aufgabe des Algorithmus ist es, ohne explizites Vorwissen[1] ein verlässliches 3D-Modell eines bewegten Objektes zu erstellen und eine verlässliche Poseschätzung ([44]) durchzuführen. Folgende Anforderungen müssen allerdings von den aufgenomenen Videosequenzen erfüllt werden:

Es soll nur ein bewegtes Objekt in der Szene vorhanden sein:
Diese Anforderung vermeidet es, zuvor eine Segmentierung nach einzelnen Objekten in der rekonstruierten Szene vornehmen zu müssen.

Das Objekt soll sich stetig und langsam bewegen:
Diese Anforderung ist für die Berechnung des optischen Flusses nach dem hier benutzten Verfahren zwingend erforderlich.

[1] Voraussetzung ist allerdings, daß mit kalibrierten Kameras gearbeitet wird.

Das Objekt darf sich in sich nicht verformen, es muß starr sein: Das Akkumulationsverfahren und die Poseschätzung[2] können nur mit starren Objekten umgehen.

Das Objekt muß definierte Kanten besitzen[3]: Diese Anforderung ist aus drei Gründen erforderlich. Zum einen wurde ein Verfahren zur Berechnung des optischen Flusses gewählt, welches auf Objektkanten angewiesen ist. Aber auch die 3D-Stereorekonstruktion und die Poseschätzung benötigen Objekte mit klaren Kantenstrukturen.

Diese Bedingungen verletzten nicht das Kriterium, kein explizites Vorwissen in die Informationsextraktion aufzunehmen.

4.2 Bezeichnungen

Es werden zunächst die verwendeten Bezeichnungen für die im Rahmen des Algorithmus vorkommenden Objekte und Funktionen eingeführt.

4.2.1 Objekte

Die Einzelbilder sind Matrizen aus $\mathbb{R}^{n\times n}$. Die spezielle euklidische Gruppe $SE(3)$ beschreibt die Körperbewegungen durch Rotation und Translation. $\mathbf{M_{3D}}$ ist die Menge der 3D-Modelle, $\mathbf{M_{2D}}$ die Menge der 2D-Merkmale (2D-Liniensegmente). Ein 3D-Modell ist hier eine Menge von 3D-Liniensegmenten.

Betrachtet wird eine Stereo-Videosequenz, bestehend aus einer Bildfolge mit je zwei Bildern. Diese stammen von der linken bzw. rechten Kamera, und werden mit V_l bzw. V_r bezeichnet. Die Sequenzen laufen über die Zeit t, starten bei t_0, und enden bei t_n. Die n Einzelbilder, aus denen die Sequenzen bestehen, sind mit $\mathrm{I}_l(t) \in \mathbb{R}^{n\times n}$ für die linken bzw. $\mathrm{I}_r(t) \in \mathbb{R}^{n\times n}$ für die rechten Bilder benannt.

Es wird das 3D-Modell $\mathrm{M}_{3D} \in \mathbf{M_{3D}}$ des betrachteten Objektes berechnet. Für die Schätzung der Pose wird ein 2D-Featureset des Objektes benötigt, dieses ist mit $\mathrm{M}_{2D} \in \mathbf{M_{2D}}$ bezeichnet. Die Poseschätzung zwischen zwei Ansichten eines Objektes ergibt eine Körperbewegung $\Delta\mathrm{RBM} \in SE(3)$.

[2] Die Poseschätzung kann auch auf spezielle Objekte, die *kinematischen Ketten*, angewandt werden, siehe Granert [16, 38].

[3] Im momentanen Stadium sind Objekte ausschließlich durch Kanten mit assoziierter Strukturinformation repräsentiert. Unser Ziel ist es jedoch die Repräsentation durch andere Beschreibungsebenen wie Textur und Farbe [30] zu erweitern.

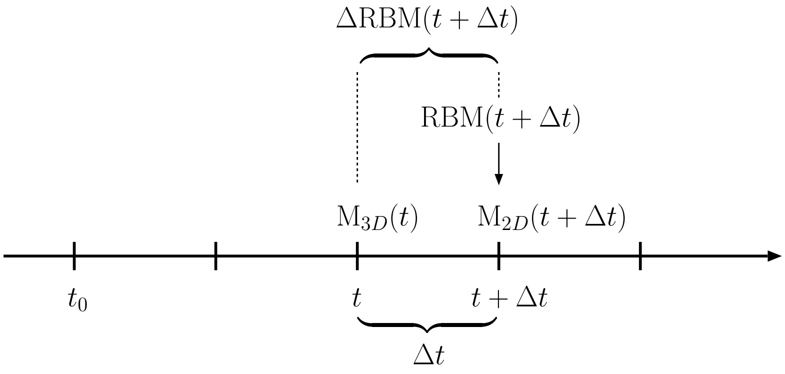


Abbildung 4.1: Zeitachse der Integration

4.2.2 Funktionen

Aus dem Stereobildpaar $\mathrm{I}_l(t)$ und $\mathrm{I}_r(t)$ wird das 3D-Modell $\mathrm{M}_{3D}(t) \in \mathbf{M_{3D}}$ berechnet. Die Berechnungsfunktion wird mit $\mathcal{S}$ bezeichnet:

$$\begin{aligned} \mathcal{S} : \mathbb{R}^{n \times n} \times \mathbb{R}^{n \times n} &\rightarrow \mathbf{M_{3D}} \\ \mathcal{S}(\mathrm{I}_l(t), \mathrm{I}_r(t)) &= \mathrm{M}_{3D}(t). \end{aligned} \tag{4.1}$$

Das so generierte 3D-Modell des Objektes ist gestört (siehe Kap. 2.2.1). Um für die Schätzung der Pose bessere Eingabedaten zu erhalten, wird das Modell gefiltert, um 3D-Liniensegmente auszuschließen, die mit höherer Wahrscheinlichkeit fehlerhaft sind und nicht das Objekt beschreiben. Die Filterfunktion (siehe Kap. 3.6) wird mit $\mathcal{F}$ bezeichnet:

$$\begin{aligned} \mathcal{F} : \mathbf{M_{3D}} &\rightarrow \mathbf{M_{3D}} \\ \mathcal{F}(\mathrm{M}_{3D}(t)) &= \mathrm{M}_{3D}{}'(t). \end{aligned} \tag{4.2}$$

Um aus dem 3D-Modell eine Darstellung des Objektes durch 2D-Features $\mathrm{M}_{2D}(t) \in \mathbf{M_{2D}}$ zu erhalten, wird das 3D-Modell auf die Bildebene projiziert. Die Projektionsfunktion $\mathcal{P}$ berechnet also aus den 3D-Liniensegmenten des 3D-Modells die projizierten 2D-Liniensegmente (siehe Faugeras [11]):

$$\begin{aligned} \mathcal{P} : \mathbf{M_{3D}} &\rightarrow \mathbf{M_{2D}} \\ \mathcal{P}(\mathrm{M}_{3D}(t)) &= \mathrm{M}_{2D}(t). \end{aligned} \tag{4.3}$$

Das Verfolgen von 2D-Liniensegmenten mit Hilfe des optischen Flusses geschieht mit der Trackingfunktion $\mathcal{T}$ (siehe Kap. 3.4). Die neue 2D-

Featuredarstellung des Objektes ist nun für den neuen Zeitpunkt $t + \Delta t$ gültig:

$$\begin{aligned} \mathcal{T} : \mathbf{M_{2D}} &\rightarrow \mathbf{M_{2D}} \\ \mathcal{T}(\mathrm{M}_{2D}(t)) &= \mathrm{M}_{2D}(t + \Delta t). \end{aligned} \tag{4.4}$$

Aus dem 3D-Modell zum Zeitpunkt t ($\mathrm{M}_{3D}(t)$) und der Darstellung durch 2D-Features zu einem Zeitpunkt $t+\Delta t$ ($\mathrm{M}_{2D}(t+\Delta t)$) kann mittels der Funktion zur Pose-Schätzung $\mathcal{E}$ die relative Körperbewegung $\Delta\mathrm{RBM}(t + \Delta t) \in SE(3)$ berechnet werden (siehe Granert, Rosenhahn [16, 38]). Sie beschreibt die Bewegung des Objektes zwischen den Zeitpunkten t und $t + \Delta t$ (siehe Abb. 4.1):

$$\begin{aligned} \mathcal{E} : \mathbf{M_{3D}} \times \mathbf{M_{2D}} &\rightarrow SE(3) \\ \mathcal{E}(\mathrm{M}_{3D}(t), \mathrm{M}_{2D}(t + \Delta t)) &= \Delta\mathrm{RBM}(t + \Delta t). \end{aligned} \tag{4.5}$$

Die Gesamtbewegung des Objektes vom Ausgangszeitpunkt t_0 bis zum aktuellen Zeitpunkt $t+\Delta t$ erhält man durch Verknüpfung der vorhergehenden Gesamtbewegung $\mathrm{RBM}(t)$ mit der aktuellen relativen Bewegung $\Delta\mathrm{RBM}(t + \Delta t)$. So wird sukzessive in jedem Schritt die Gesamtbewegung aktualisiert:

$$\mathrm{RBM}(t + \Delta t) = \mathrm{RBM}(t) * \Delta\mathrm{RBM}(t + \Delta t) \tag{4.6}$$

Das akkumulierte 3D-Modell ${\mathrm{M}_{3D}}'(t_0)$ wird durch die Akkumulationsfunktion $\mathcal{A}$ berechnet (siehe Krüger, Ackermann [28, 1]). Es wird auf den Zeitpunkt t_0 bezogen, da dort die Ausgangsposition des 3D-Modells gesetzt wird, auf die alle neu berechneten Modelle zurückgedreht werden. Zu jedem 3D-Modell muß jeweils die RBM angegeben werden, mit der das Modell in die Ausgangsposition zum Zeitpunkt t_0 zurückgedreht werden kann:

$$\begin{aligned} \mathcal{A} : (\mathbf{M_{3D}} \times SE(3)) \times \ldots \times (\mathbf{M_{3D}} \times SE(3)) &\rightarrow \mathbf{M_{3D}} \\ \mathcal{A}((\mathrm{M}_{3D}(t_0), \mathrm{RBM}(t_0)), \ldots, (\mathrm{M}_{3D}(t_n), \mathrm{RBM}(t_n))) &= {\mathrm{M}_{3D}}'(t_0). \end{aligned} \tag{4.7}$$

In der Praxis setzt sich die Akkumulationsfunktion aus mehreren Teilen zusammen. Die Akkumulation wird mit der *Initialisierungsfunktion* $\mathcal{AI}$ initialisiert.

$$\begin{aligned} \mathcal{AI} : \mathbf{M_{3D}} \times SE(3) &\rightarrow \mathbf{M_{3D}} \\ \mathcal{A}(\mathrm{M}_{3D}(t_0), \mathrm{RBM}(t_0)) &= \mathrm{M}_{3Dn}. \end{aligned} \tag{4.8}$$

Die *Updatefunktion* $\mathcal{AU}$ assoziiert ein in einer neuen Ansicht entstandenes 3D-Modell zu der Menge der für die Akkumulation zu berücksichtigten Modelle:

$$\begin{array}{rcl} \mathcal{AU} : \mathbf{M_{3D}} \times SE(3) & \rightarrow & \mathbf{M_{3D}} \\ \mathcal{AU}(\mathrm{M}_{3D}(t), \mathrm{RBM}(t)) & = & \mathrm{M}_{3Dn}. \end{array} \tag{4.9}$$

Das akkumulierte Modell wird durch die *Mergefunktion* $\mathcal{AM}$ berechnet. Alle durch die Updatefunktion $\mathcal{AU}$ hinzugefügten Modelle werden zusammengefaßt. Diese wurden auf die Ausgangsposition zurückgedreht. Damit sind sie gültig für den Zeitpunkt t_0, doch es existieren nun n verschiedene 3D-Modelle:

$$\begin{array}{rcl} \mathcal{AM} : \mathbf{M_{3D}} \times \ldots \times \mathbf{M_{3D}} & \rightarrow & \mathbf{M_{3D}} \\ \mathcal{AM}(\mathrm{M}_{3D1}, \ldots, \mathrm{M}_{3Dn}) & = & {\mathrm{M}_{3D}}'(t_0). \end{array} \tag{4.10}$$

4.3 Ablauf

Der Ablauf der Integration beginnt mit der Berechnung des ersten 3D-Modells $\mathrm{M}_{3D}(t_0)$ aus dem ersten Stereobild der Stereo-Videosequenz:

$$\mathrm{M}_{3D}(t_0) = \mathcal{S}(\mathrm{I}_l(t_0), \mathrm{I}_r(t_0)). \tag{4.11}$$

Mit diesem 3D-Modell wird die Akkumulation initialisiert. Es konnte noch keine Poseschätzung vorgenommen werden. Die Notwendigkeit zur Angabe einer Bewegung ist bei der Initialisierung noch gar nicht gegeben, denn die Position des Modells gibt gleichzeitig die Ausgangsposition vor. Als Bewegung wird somit die Identität I angegeben, es gilt folglich für die Initialisierung $\mathrm{RBM}(t_0) = I$:

$$\mathrm{M}_{3D0} = \mathcal{AI}(\mathrm{M}_{3D}(t_0), \mathrm{RBM}(t_0)). \tag{4.12}$$

Da das 3D-Modell $\mathrm{M}_{3D}(t_0)$ Fehler enthält, wird es vor der weiteren Verwendung für die Poseschätzung anhand des Kollinearitätskriteriums gefiltert (siehe Kapitel 3.6):

$${\mathrm{M}_{3D}}^f(t_0) = \mathcal{F}(\mathrm{M}_{3D}(t_0)). \tag{4.13}$$

Das resultierende 3D-Modell ${\mathrm{M}_{3D}}^f(t_0)$ ist eine Teilmenge des Ausgangsmodells $\mathrm{M}_{3D}(t_0)$. Es bleiben genau die 3D-Liniensegmente übrig, für die das Kollinearitätskriterium greift:

$${\mathrm{M}_{3D}}^f(t_0) \subseteq \mathrm{M}_{3D}(t_0). \tag{4.14}$$

Im nächsten Schritt wird das gefilterte 3D-Modell ${M_{3D}}^f(t_0)$ mit der Projektionsmatrix der kalibrierten Kamera auf die Bildebene projiziert. Als Ergebnis erhält man ein Set von 2D-Features des Objektes. Dies sind 2D-Liniensegmente. Durch die Projektion ist die Korrespondenz zwischen den 3D-Liniensegmenten und den daraus jeweils entstandenen 2D-Segmenten sofort gegeben. Das 2D-Featureset wird mit $M_{2D}(t_0)$ bezeichnet:

$$M_{2D}(t_0) = \mathcal{P}({M_{3D}}^f(t_0)). \tag{4.15}$$

Diese 2D-Liniensegmente werden jetzt in der Bildebene verschoben. Dazu wird die Bewegungsinformation des optischen Flusses benutzt. Die Verfolgung der Segmente kann über mehrere Bilder der Sequenz ausgedehnt werden. Für diese Betrachtung wird eine Zeitspanne Δt angenommen, ohne Konkret auf die tatsächliche Anzahl der Bilder einzugehen.Als Ergebnis erhält man ein 2D-Featureset $M_{2D}(t_0+\Delta t)$ des Objektes, welches für den Zeitpunkt $t_0 + \Delta t$ gültig ist:

$$M_{2D}(t_0 + \Delta t) = \mathcal{T}(M_{2D}(t_0)). \tag{4.16}$$

Die Verschiebung der 2D-Liniensegmente geschieht mit Hilfe des optischen Flusses, und die Zuordnung von originalem und verschobenem Segment bleibt erhalten. Die Korrespondenzen zwischen den 3D-Liniensegmenten von ${M_{3D}}^f(t_0)$ und den 2D-Segmenten von $M_{2D}(t_0 + \Delta t)$ sind daher weiterhin bekannt. Somit kann die Bewegung geschätzt werden, die das Objekt aus seiner Lage zum Zeitpunkt t_0 in die neue Lage zum Zeitpunkt $t_0 + \Delta t$ überführt. Die Poseschätzung liefert diese Körperbewegung $\Delta\text{RBM}(t_0 + \Delta t)$:

$$\Delta\text{RBM}(t_0 + \Delta t) = \mathcal{E}({M_{3D}}^f(t_0), M_{2D}(t_0 + \Delta t)). \tag{4.17}$$

Die Gesamtbewegung des Objektes von t_0 nach $t_1 = t_0 + \Delta t$ wird erreicht durch die Verknüpfung der bisherigen Bewegung $\text{RBM}(t_0)$ mit der neu geschätzten $\Delta\text{RBM}(t_0 + \Delta t) = \Delta\text{RBM}(t_1)$:

$$\text{RBM}(t_1) = \text{RBM}(t_0) * \Delta\text{RBM}(t_1). \tag{4.18}$$

Im folgenden wiederholt sich dieser Ablauf. Für den Zeitpunkt $t_1 = t_0 + \Delta t$ in der Videosequenz wird jetzt erneut das 3D-Modell berechnet:

$$M_{3D}(t_1) = \mathcal{S}(I_l(t_1), I_r(t_1)). \tag{4.19}$$

Als Änderung nach dem ersten Durchlauf wird in allen folgenden Schleifendurchgängen anstatt der Initialisierung der Akkumulation ein Update mit dem neuen 3D-Modell $M_{3D}(t_1)$ und der dazu gehörenden geschätzten Bewegung $\text{RBM}(t_1)$ durchgeführt:

$$M_{3D1} = \mathcal{AU}(M_{3D}(t_1), \text{RBM}(t_1)). \tag{4.20}$$

So wird bei jedem Durchlauf ein 3D-Modell, welches aus der neuen Ansicht rekonstruiert wurde, mit der zugehörigen Bewegung, die das Modell zurück zur Ausgangsposition bewegt, für das Akkumulieren hinzugenommen. Diese Modelle werden jeweils mit M_{3Di} bezeichnet. Die Anzahl der Schleifendurchläufe n ist begrenzt durch die Länge der Videosequenz. Sind alle Bilder der Sequenz abgearbeitet, so kann das akkumulierte 3D-Modell $\mathrm{M}_{3D}'(t_0)$ berechnet werden. Dazu wird die Mergefunktion aufgerufen, die aus den in den Update-Schritten übergebenen Einzelmodellen das akkumulierte Modell berechnet:

$$\mathrm{M}_{3D}'(t_0) = \mathcal{AM}(\mathrm{M}_{3D1}, \ldots, \mathrm{M}_{3Dn}). \tag{4.21}$$

4.4 Zusammenfassung des Algorithmus

Der Ablauf der Integration stellt sich folgendermaßen als Algorithmus dar:

`//Initialisierung`

$i = 0;$

$\mathrm{M}_{3D}(t_i) = \mathcal{S}(\mathrm{I}_l(t_i), \mathrm{I}_r(t_i));$

$\mathrm{RBM}(t_i) = I;$

$\mathrm{M}_{3Di} = \mathcal{AI}(\mathrm{M}_{3D}(t_i), \mathrm{RBM}(t_i));$

`//Schleife`

`DO`

$\quad \mathrm{M}_{3D}{}^f(t_i) = \mathcal{F}(\mathrm{M}_{3D}(t_i));$

$\quad \mathrm{M}_{2D}(t_i) = \mathcal{P}(\mathrm{M}_{3D}{}^f(t_i));$

$\quad \mathrm{M}_{2D}(t_i + \Delta t) = \mathcal{T}(\mathrm{M}_{2D}(t_i));$

$\quad \Delta\mathrm{RBM}(t_i + \Delta t) = \mathcal{E}(\mathrm{M}_{3D}{}^f(t_i), \mathrm{M}_{2D}(t_i + \Delta t));$

$\quad \mathrm{RBM}(t_{i+1}) = \mathrm{RBM}(t_i) * \Delta\mathrm{RBM}(t_i + \Delta t);$

$\quad i = i + 1;$

$\quad \mathrm{M}_{3D}(t_i) = \mathcal{S}(\mathrm{I}_l(t_i), \mathrm{I}_r(t_i));$

$\quad \mathrm{M}_{3Di} = \mathcal{AU}(\mathrm{M}_{3D}(t_i), \mathrm{RBM}(t_i));$

`WHILE (NOT END OF SEQUENCE)`

$\mathrm{M}_{3D}'(t_0) = \mathcal{AM}(\mathrm{M}_{3D0}, \ldots, \mathrm{M}_{3Di});$

Kapitel 5

Ergebnisse

In diesem Kapitel wird dargestellt, welche Ergebnisse mit dem vorgestellten Verfahren erreicht werden. Zu diesem Zweck wurden verschiedene Stereo-Videosequenzen aufgenommen, und auf diesen das Akkumulationsverfahren angewendet.

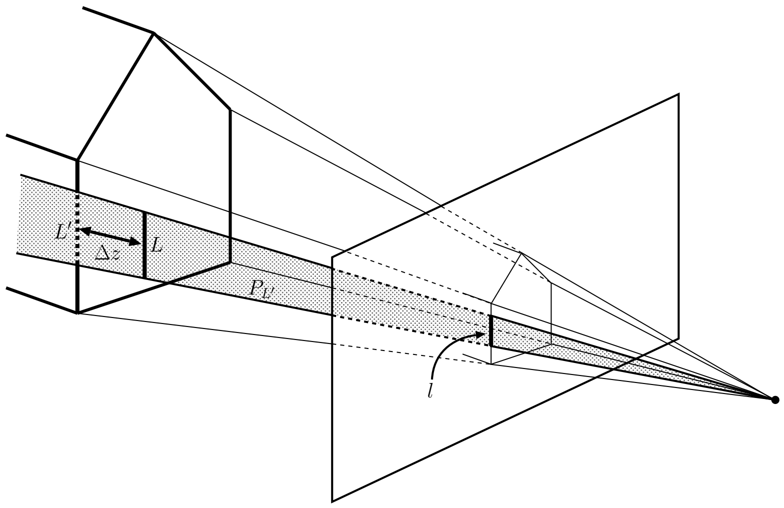


Abbildung 5.1: Falsch rekonstruiertes Liniensegment L erscheint in seiner Projektion l an der richtigen Position

Gezeigt werden die Ergebnisse in zwei verschiedenen Darstellungsweisen. Für eine Darstellung wird die berechnete 3D-Repräsentation auf die Bildebene projiziert und mit dem zugehörigen Kamerabild überlagert. In dieser Darstellungsweise kann überprüft werden, ob die berechnete 3D-Repräsentation tatsächlich dem aufgenommenem Objekt entspricht. Bestimmte Fehler in der

Rekonstruktion können durch diese Darstellung allerdings nicht erkannt werden. Wie Abb. 5.1 exemplarisch zeigt, ist das Liniensegment L in der Tiefe mit einem Fehler von Δz falsch rekonstruiert worden. Richtig wäre das Liniensegment L' gewesen. Da L auf den Projektionsebene $P_{L'}$ von L' liegt, ist die Projektion l korrekt, und der Fehler Δz bleibt unerkannt.

Um Fehler dieser Art zu erkennen, wird als zweite Darstellungsweise der berechneten 3D-Repräsentation zusätzlich noch die Projektion auf die 3 Koordinatenebenen hinzugenommen. Damit läßt sich erkennen, ob solche in der Tiefe falsch rekonstruierte Liniensegmente vorkommen.

Für ein Model wurde zusätzlich zur visuellen noch eine statistische Fehleranalyse durchgeführt.

5.1 Filterung

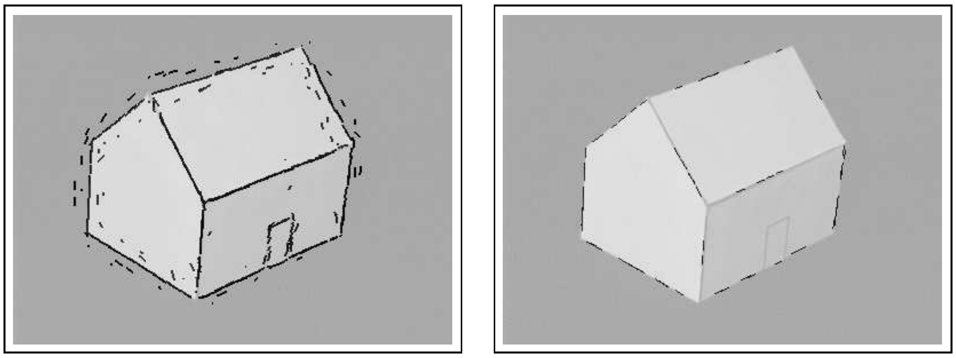


Abbildung 5.2: Vergleich ungefilterte (links) und gefilterte 3D-Repräsentation (rechts) in der 3D-Ansicht

Zunächst werden die Ergebnisse des Filterungsverfahrens (siehe Kap. 3.6) gezeigt. Die Abbildung 5.2 stellt links die ungefilterte 3D-Repräsentation eines Hauses, und rechts die nach der Filterung übrigbleibende Repräsentation dar. Zu erkennen ist, daß nach der Filterung nur noch solche 3D-Liniensegmente übriggeblieben sind, die das Modell des Hauses offenbar gut repräsentieren.

Ob Fehler in der Positionierung von 3D-Segmenten vorliegen (siehe Abb. 5.1), läßt sich mit der Darstellungsweise der Projektion der Repräsentation auf die drei Koordinatenebenen in Abbildung 5.3 überprüfen. Oben sind die Ansichten der ungefilterten Repräsentation abgebildet, unten die der gefilterten. Im Vergleich ist zu erkennen, daß die in der Tiefe falsch rekonstruierten 3D-Liniensegmente durch das Filterungsverfahren erfolgreich entfernt

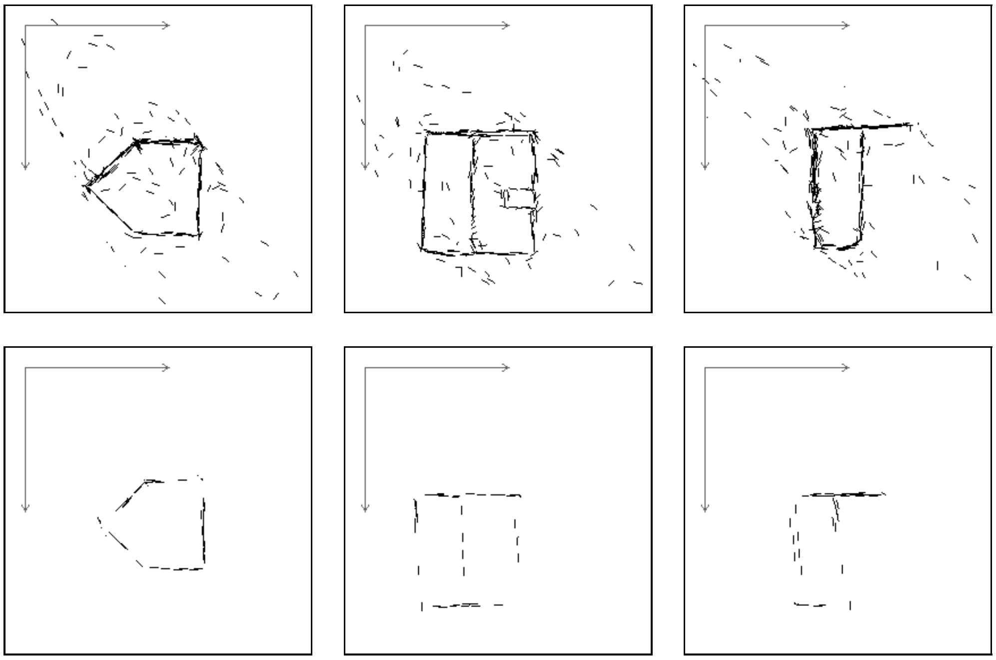


Abbildung 5.3: Vergleich ungefilterte (oben) und gefilterte Repräsentation (unten)

wurden. Auch diese Darstellung bestätigt das erfolgreiche Filtern der 3D-Repräsentation auf relevante Liniensegmente.

Die Ergebnisse des Filterungsverfahren hängen direkt mit den Einstellungen für die zulässigen Abstandsmaße zusammen, die in Kapitel 3.6 beschrieben wurden. Die Parameter, die eine gute Filterung erzielen, müssen in diesem Zusammenhang für ein verwendetes Objekt jeweils initial manuell ermittelt werden.

Um die Auswirkung der Filterung auf die Poseschätzung zu untersuchen, wurde für eine gefilterte und eine ungefilterte 3D-Repräsentation jeweils für einen Bewegungsschritt von Bild n nach Bild $n+1$ einer Videosequenz die Pose geschätzt. Verwendet wurde dafür die 3D-Repräsentation, die aus Bild n berechnet wurde, und ihre nach Bild $n+1$ getrackte 2D-Projektion. Für das Bild $n+1$ wurde nun eine neue Stereorekonstruktion vorgenommen. Diese ist dann unter Verwendung der aus der Poseschätzung erhaltenen Bewegungsparameter zurücktransformiert worden. Die Abb. 5.4 zeigt die Ergebnisse dieser Rücktransformation. Bei der Verwendung der ungefilterten Repräsentation (rechtes Bild) ist zu erkennen, daß die Poseschätzung fehlerhafte Bewegungparameter ergeben hat, die Repräsentation "paßt" nicht auf das Kamerabild. Nimmt man die gefilterte Repräsentation (linkes Bild), so ist zu erkennen,

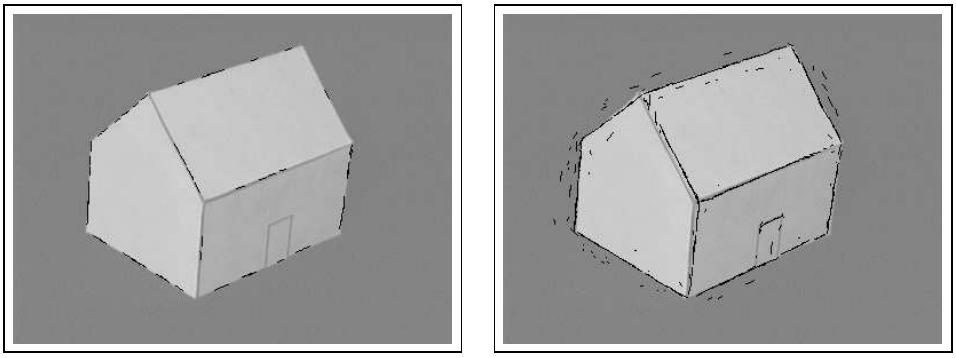


Abbildung 5.4: Vergleich korrekte Poseschätzung bei gefiltertem Repräsentation (links) und falsche Schätzung bei ungefiltertem Modell (rechts)

daß die Poseschätzung erfolgreich war.

5.2 Videosequenz "Haus"

Für diese Videosequenz wurde ein einfaches Modell eines Hauses verwendet. Die Beschaffenheit dieses Modells mit weißen Wänden und schwarzen Kanten ist für die Zwecke der 3D-Rekonstruktion als optimal anzusehen. Durch die Verwendung eines schwarzen Hintergrundes ohne jede Struktur wird vermieden, daß nicht zum bewegten Objekt gehörende 3D-Liniensegmente extrahiert werden. Das Haus dreht sich mit seiner Gibelseite von links nach rechts (siehe Abb. 5.5).

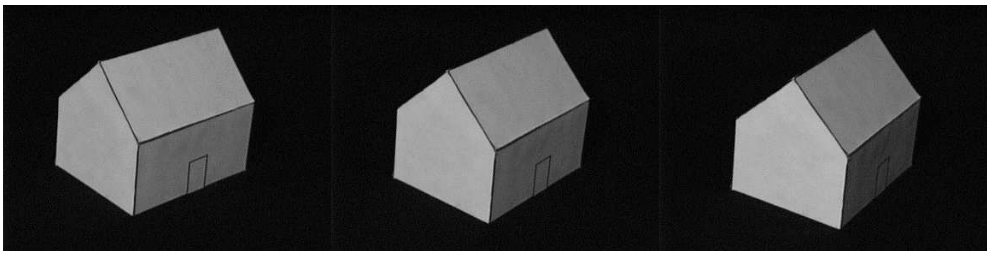


Abbildung 5.5: Videosequenz "Haus" (gezeigt ist hier das linke Kamerabild)

5.2.1 Visuelle Auswertung

Das Ergebnis der Akkumulation im Vergleich zur einfachen 3D-Stereorekonstruktion ist in Abb. 5.6 dargestellt. Dabei wird die Darstellungsweise

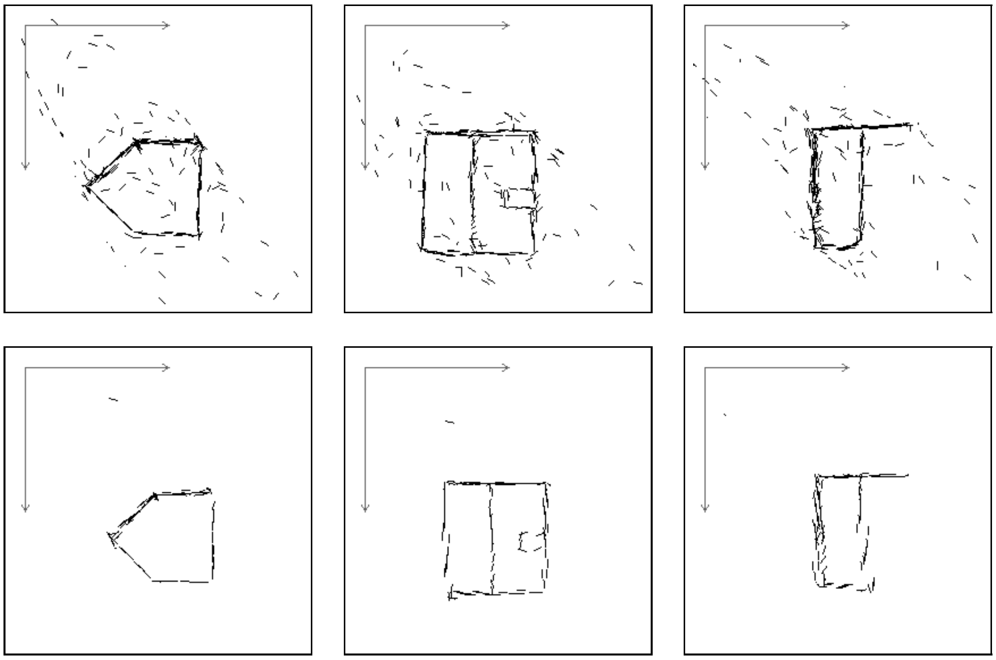


Abbildung 5.6: Vergleich 3D-Stereorekonstruktion (oben) und akkumulierte Repräsentation (unten)

der Projektion der Repräsentationen auf die drei Koordinatenebenen verwendet. Oben im Bild ist die 3D-Repräsentation ohne Akkumulation, unten die akkumulierte 3D-Repräsentation gezeigt. Erkennbar ist die durch die Akkumulation erzielte weitgehende Eliminierung falsch rekonstruierter 3D-Liniensegmente.

Während der einzelnen Akkumulationsschritte finden nach und nach immer mehr 3D-Liniensegmente eine Bestätigung, das Objekt gut zu repräsentieren. Daher vervollständigt sich die 3D-Repräsentation mit jedem Akkumulationsschritt immer mehr. Die Abb. 5.7 zeigt einzelne Schritte des Verfahrens. Oben beginnend mit einer lückenhaften Repräsentation, werden in den einzelnen Schritten nach unten immer mehr Liniensegmente hinzugenommen.

Die Projektion des akkumulierten Modells auf das Kamerabild ist in Abb. 5.8 dargestellt. Auch hier läßt sich vergleichen, wie aus einem anfänglich spärlich rekonstruiertem Objekt (Abb. 5.8 oben) nach mehreren Akkumulationsschritten eine vollständigere 3D-Repräsentation gewonnen wird (Abb. 5.8 mitte). In der Abb. 5.8 (unten) sind sämtliche Liniensegmente gezeigt, mit dem Grad ihrer Helligkeit in Abhängigkeit von ihrer Konfidenz. Dies macht deutlich, daß die Fehler der Rekonstruktion noch vorhanden sind, diese aber durch die Konfidenzbewertung beschrieben werden.

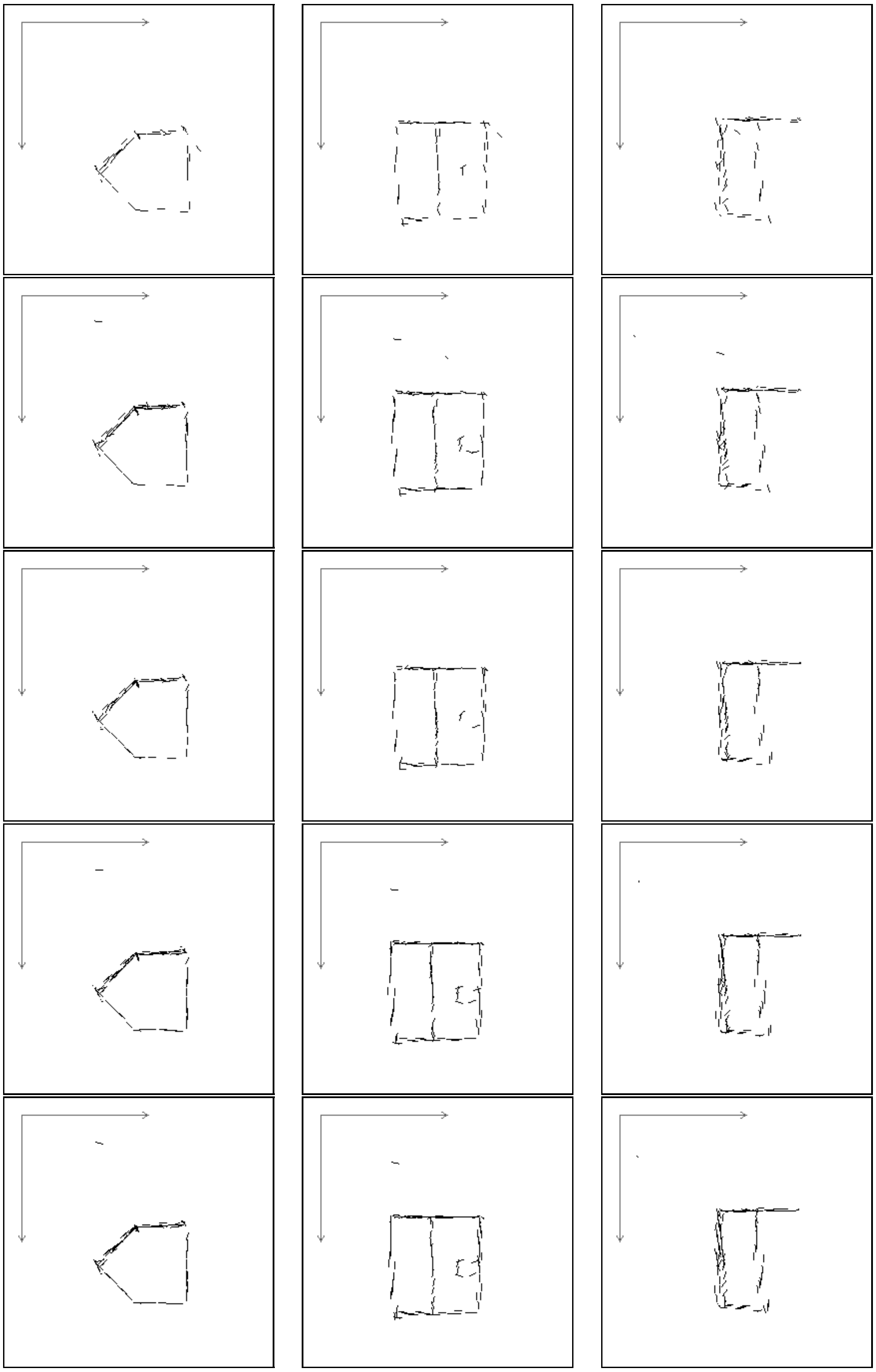


Abbildung 5.7: Akkumulation auf der Sequenz “Haus”

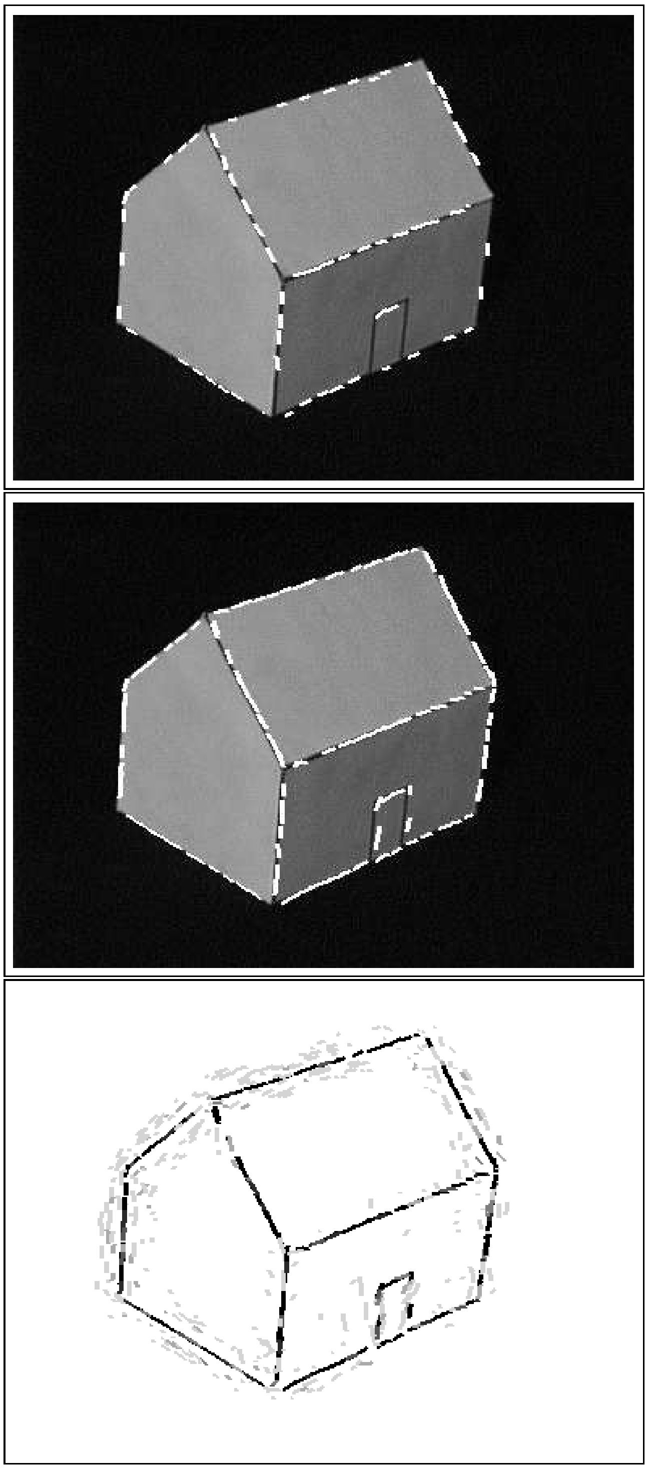


Abbildung 5.8: Vergleich unvollständige (oben), akkumulierte 3D-Repräsentation (mitte) und konfidenzbewertete Liniensegmente (unten)

5.2.2 Statistische Auswertung

Zusätzlich zur visuellen Kontrolle der Ergebnisse wurde für das Modell des Hauses eine statistische Auswertung vorgenommen. Aus den Vermessungsdaten des Modells wurde die korrekte 3D-Repräsentation erstellt. Für die Liniensegmente der berechneten 3D-Repräsentation wurde dann die Abweichung von dieser korrekten Repräsentation ermittelt. Dabei wurden sowohl die Positions- als auch die Orientierungsdifferenzen der Segmente berechnet.

5.2.2.1 Werteerfassung

Eine 3D-Repräsentation besteht aus einer Menge von 3D-Liniensegmenten. Um die Abweichung für jedes dieser berechneten 3D-Liniensegmente L_i mit Aufpunkt $\vec{a_i}$ und Orientierungsrichtung $\vec{u_i}^\circ$ von seinem korrekten Pendant K_j mit Aufpunkt $\vec{p_j}$ und Orientierung $\vec{v_j}^\circ$ bestimmen zu können, wurden zunächst die Korrespondenzen $i \leftrightarrow j$ zwischen berechneten und korrekten Liniensegmenten manuell ermittelt.

Mittels der für das Filterungsverfahren angegebenen Abstandsmaße wurde dann für jedes 3D-Liniensegment L_i das Abstandsmaß d_i in der Position und δ_i in der Orientierung von seinem korrespondierendem korrektem Segment K_j berechnet:

$$\begin{aligned} d_i &= |\vec{u_i}^\circ \times (\vec{p_j} - \vec{a_i})|, \\ \delta_i &= |\vec{u_i}^\circ - \vec{v_j}^\circ| \end{aligned}$$

(siehe Kap. 3.6.2, Gleichungen (3.47), (3.48), Abb. 3.12, 3.13). Erfaßt wurden diese Werte für die jeweiligen berechneten 3D-Repräsentationen in den durchgeführten Akkumulationsschritten.

5.2.2.2 Akkumulation

Die Ergebnisse der Akkumulation werden in den folgenden Diagrammen gezeigt. Die visuelle Auswertung der Ergebnisse ergab eine Erhöhung der Anzahl der rekonstruierten Liniensegmente mit zunehmenden Akkumulationsschritten. Die Abb. 5.9 zeigt die absoluten Zahlen für die Menge der Segmente in den einzelnen Iterationen. Beginnend mit 63 Liniensegmenten steigt die Anzahl nach sechs Iterationsschritten um das Doppelte auf 126 Segmente an. Die visuelle Beobachtung findet in diesen Zahlen ihre Bestätigung.

Auswertung Positionsfehler

Um zu überprüfen, ob das Akkumulationsverfahren erfolgreich die berechnete 3D-Repräsentation mit zunehmenden Iterationsschritten genauer an die kor-

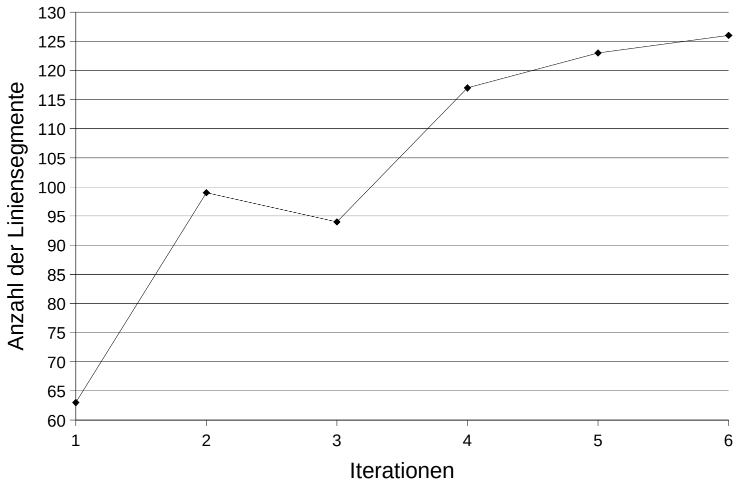


Abbildung 5.9: Anzahl der Liniensegmente pro Iterationsschritt

rekte Repräsentation annähert, wurde für jeden Iterationsschritt jeweils der mittlere Abstandsfehler über alle Fehlerwerte d_i berechnet. Die Ergebnisse sind in Abb. 5.10 dargestellt. Zu beobachten ist ein Rückgang des mittleren Fehlers Beginnend mit einem Wert von 5.6 fällt er auf 4.7 ab, eine Verbesserung um den Faktor 1.18 liegt vor. Eine Annäherung der berechneten 3D-Repräsentation an die korrekte Repräsentation wird durch die Akkumulation erzielt.

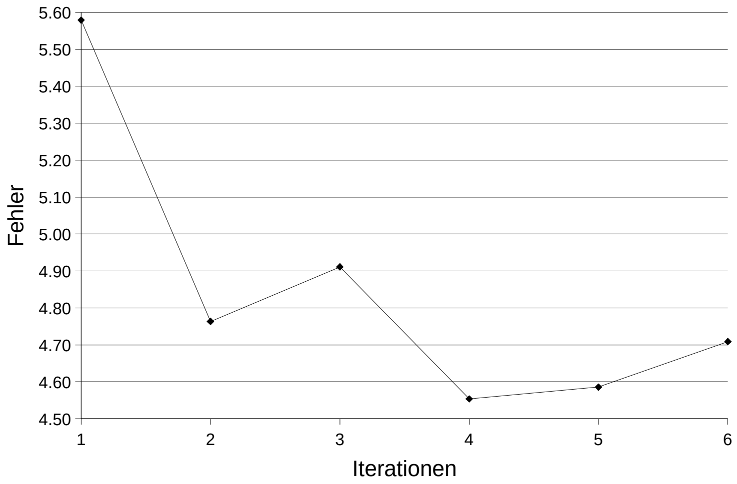


Abbildung 5.10: Mittelwert der Positionsfehler der Liniensegmente

Eine weitere Aussage läßt sich der Varianz der ermittelten Fehlerwerte d_i

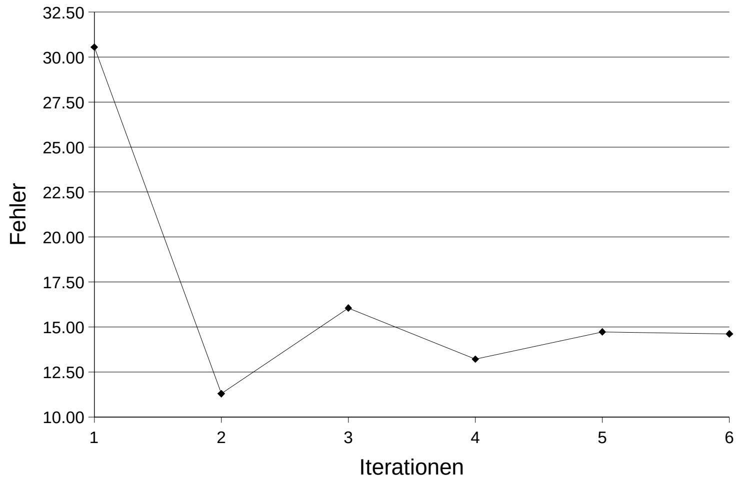


Abbildung 5.11: Varianz der Positionsfehler der Liniensegmente

entnehmen. Neben dem Mittelwert wurde deshalb auch die Varianz der Positionsfehler der berechneten 3D-Liniensegmente untersucht. Die Abb. 5.11 zeigt die Varianz der Positionsfehler für jeden Iterationsschritt der Akkumulation. Das Verfahren vermindert die Varianz deutlich, der Wert sinkt von 30.5 beinahe um die Hälfte auf 14.6 ab. Der genau Verbesserungsfaktor beträgt 2.09. Damit besitzt die akkumulierte Repräsentation eine geringere Streuung seiner 3D-Liniensegmente, eine Verbesserung kann somit durch diese Ergebnisse bestätigt werden.

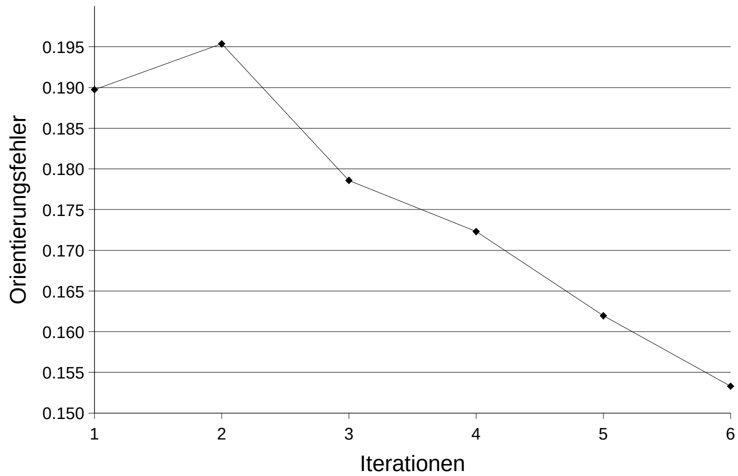


Abbildung 5.12: Mittelwert der Orientierungesfehler der Liniensegmente

Auswertung Orientierungsfehler

Für die Orientierungsfehler der berechneten 3D-Liniensegmente ist ebenfalls Mittelwert und Varianz untersucht worden. Die Mittelwerte über alle Orientierungsfehler δ_i für die Iterationsschritte 1 – 6 sind in Abb. 5.12 eingetragen. Auch für die Abweichung in den Orientierungen ergibt sich, wie bei den Positionsfehlern, eine Verminderung des Fehlers mit zunehmender Iterationszahl. Der Wert fällt von 0.19 auf 0.15 ab. Damit wird für die Verbesserung ein Faktor von 1.24 erzielt.

Deutlicher fällt die Verbesserung der Varianz in den Orientierungsfehlern aus. Die Abb. 5.13 zeigt den Verlauf für die Varianz in den einzelnen Iterationsschritten. Der Wert vermindert sich von 0.044 auf 0.018, dies ergibt einen Verbesserungsfaktor von 2.37. Demzufolge ist die Streuung in den Orientierungsfehlern mit Hilfe der Akkumulation deutlich vermindert worden.

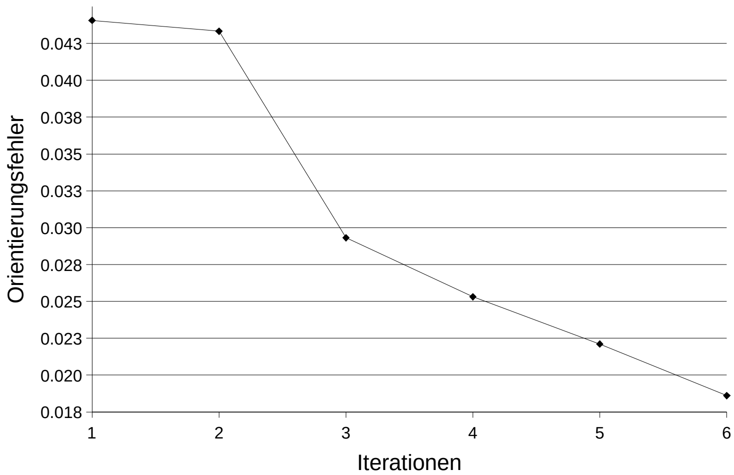


Abbildung 5.13: Varianz der Orientierungsfehler der Liniensegmente

Darstellung der Fehler als Histogramm

Aus der Verteilung der Fehlerwerte läßt sich ein Histogramm erstellen, welches für Gesamtbreite der vorkommenden Fehlerwerte jeweils aufzeigt, wieviele 3D-Liniensegmente mit dieser Fehlergröße in der 3D-Repräsentation auftreten. In Abb. 5.14 sind die Histogramme der Positionsfehler für die erste und die letzte durchgeführten Iteration aufgezeigt. Die Anzahl von 3D-Liniensegmenten, welche einen kleineren Fehler aufweisen, ist bei der letzten Iteration höher als bei der ersten.

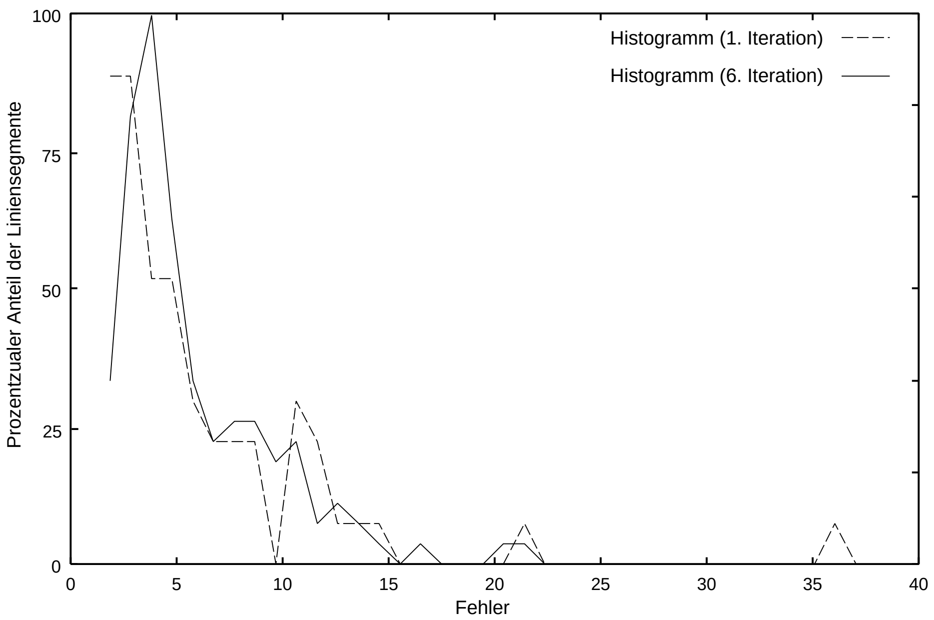


Abbildung 5.14: Histogramme der Distanzfehler für die 1. und die 6. Iteration

Auch für die Orientierungsfehler wurden die Histogramme für die erste und die letzte Iteration aufgezeichnet. Die Abb. 5.15 zeigt die Ergebnisse. Im Vergleich der beiden Histogramme zeigt sich, daß nach der ersten Iteration noch mehr Fehler mit großem Wert vorkommen, als nach der letzten Iteration.

Fazit

Die durch einfache visuelle Auswertung der Ergebnisse beobachtete Verbesserung der 3D-Rekonstruktion unter Anwendung des Akkumulationsverfahrens findet in der statistischen Auswertung ihre Bestätigung. Sowohl für die Position der berechneten 3D-Liniensegmente als auch für deren Orientierung nimmt der absolute Fehler ab. Eine weitere Bestätigung der Verbesserung der akkumulierten 3D-Repräsentation findet sich in der Abnahme der Varianzen mit zunehmenden Iterationsschritten.

Ein Rückgang der Fehler ganz auf Null wird nicht erreicht. Die Gründe hierfür liegen darin, daß mit manuell kalibrierten Kameras gearbeitet wurde. Hinzu kommt eine relativ geringe Bildauflösung im Viertel-PAL-Format mit 384×288 Bildpunkten, und ein im Verhältnis zum verwendeten Modell relativ großer Kameraabstand zwischen Modell und Kamera von ca. $2,5$m.

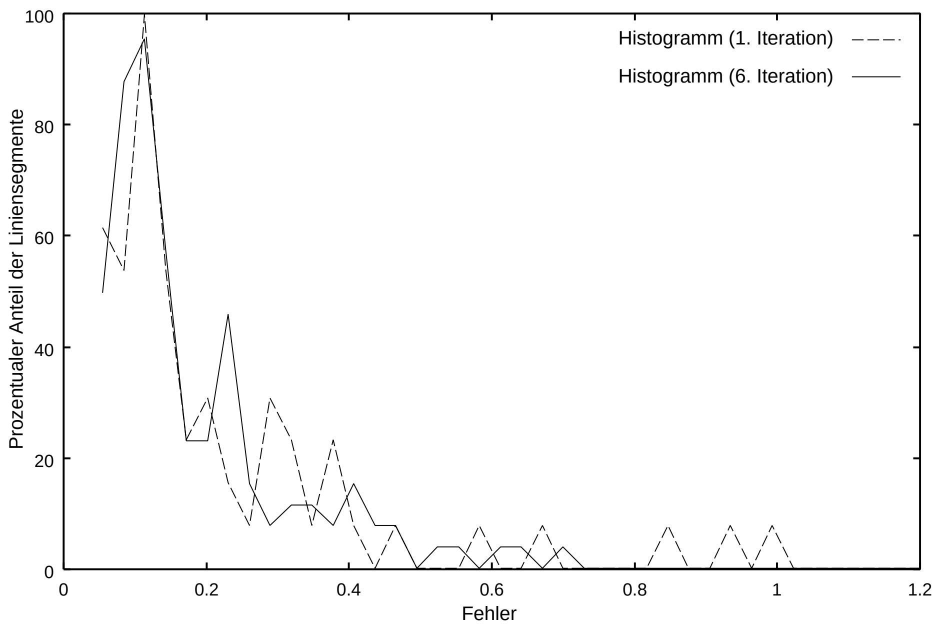


Abbildung 5.15: Histogramme der Orientierungsfehler für die 1. und die 6. Iteration

5.2.2.3 Verwendung fusionierter Bewegungsparameter

Es wurde statistisch die Auswirkung der Verwendung fusionierter Bewegungsparameter (siehe Kap. 3.7) untersucht. Die Abbildung 5.16 zeigt die Ergebnisse der ersten und der sechsten Iteration für den mittleren Fehler (links) und die Varianz (rechts) in den Positionierungen der Liniensegmente, jeweils unter Verwendung der Poseschätzung basierend auf dem linken oder dem rechten Bild, oder der Fusion beider Schätzungen.

Die Verwendung des rechten Bildes für die Poseschätzung führt zu einem kaum verkleinerten mittlere Fehler, wohingegen bei Verwendung des linken Bildes der Fehler deutlich zurückgeht. Im Ergebnis ist deshalb der mittlere Fehler der fusionierten Bewegungsschätzung größer als bei der nur linksseitig geschätzten Bewegung. Die Varianz nimmt auch bei der rechtsseitigen Bewegungsschätzung ab. Daher kann für die Varianz durch Verwenden der Fusionierten Bewegungsparameter die größtmögliche Verbesserung erzielt werden.

Betrachtet man den Orientierungsfehler in Abbildung 5.17, so ist zu erkennen, daß die rechtsseitige Schätzung weniger gute Ergebnisse erzielt als die linksseitige. Dies betrifft sowohl den mittleren Fehler (Abb. 5.17 links) als auch die Varianz (Abb. 5.17 rechts). Für die Verwendung der fusionierten Bewegungsparameter ergibt sich ein mittlerer Fehler und eine Varianz, die sich

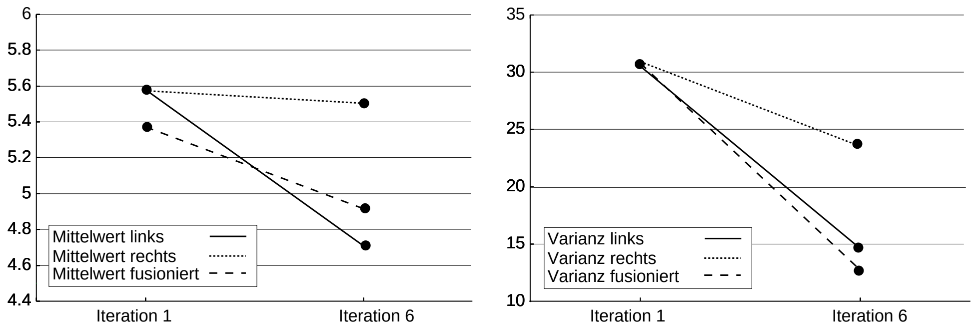


Abbildung 5.16: Vergleich der Mittelwerte (links) und Varianzen (rechts) der Positionsfehler bei Verwendung von linken, rechten und fusionierten Bewegungsparametern

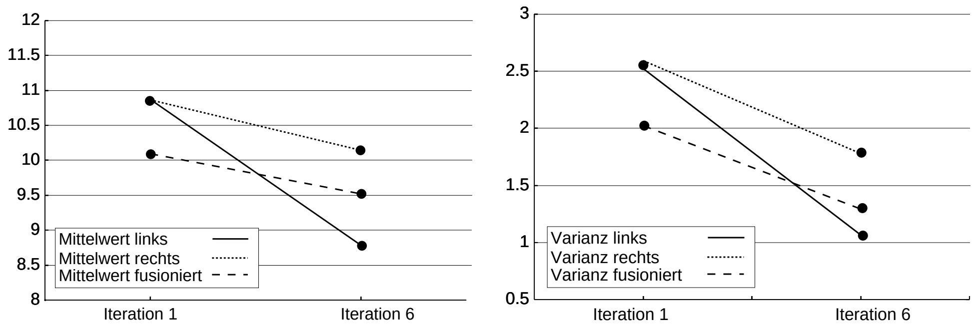


Abbildung 5.17: Vergleich der Mittelwerte (links) und Varianzen (rechts) der Orientierungsfehler bei Verwendung von linken, rechten und fusionierten Bewegungsparametern

in der Mitte zwischen links- und rechtsseitiger Bewegungsschätzung finden. Die Bewegungsschätzung basierend auf den linken Bildern erzielt bessere Ergebnisse als bei Verwendung der rechten Bilder, da die Stereorekonstruktion Punkte auf den Sehstrahlen der linken Bilder verwendet (siehe Kap. 3.2.2).

5.2.2.4 Filterung

Untersucht wurde die Auswirkung des Filterungsverfahrens auf die Ergebnisse der Poseschätzung. Dafür wurde eine Berechnung mit gefilterter 3D-Repräsentation durchgeführt, und dieselbe Berechnung unter Weglassung der Objektfilterung wiederholt. Die Fehlerwerte für die gefilterte und die ungefilterte Repräsentation wurden auf die oben angegebene Weise gewonnen (siehe

Kap. 5.2.2.1).

Auswertung Positionsfehler

Sowohl in der Berechnung mit Filterung, als auch in der ohne Filterung wurde jeweils die erste und die letzte Repräsentation ausgewertet. Die Tabelle 5.1 zeigt die erhaltenen Ergebnisse für Mittelwerte und Varianzen der Positionsfehler.

Mittelwert		
Iterationen	gefiltert	ungefiltert
1	5.4868	5.4171
6	4.6517	8.6010

Varianz		
Iterationen	gefiltert	ungefiltert
1	30.1260	29.524
6	9.9929	33.075

Tabelle 5.1: Vergleich gefiltertes und ungefiltertes Modell: Mittelwerte (links) und Varianzen (rechts) der Positionsfehler

Zu Erkennen ist, daß sich die Werte für das gefilterte und das ungefilterte Modell in der ersten Iteration kaum unterscheiden. Dies betrifft sowohl den Mittelwert als auch die Varianz. Aber nach sechs Iterationsschritten tritt ein deutlicher Unterschied auf. Während der Mittelwert und die Varianz für das gefilterte Modell erwartungsgemäß abnehmen[1], tritt für das ungefilterte Modell sogar noch einer Erhöhung dieser Werte auf. Abb. 5.18 stellt dies

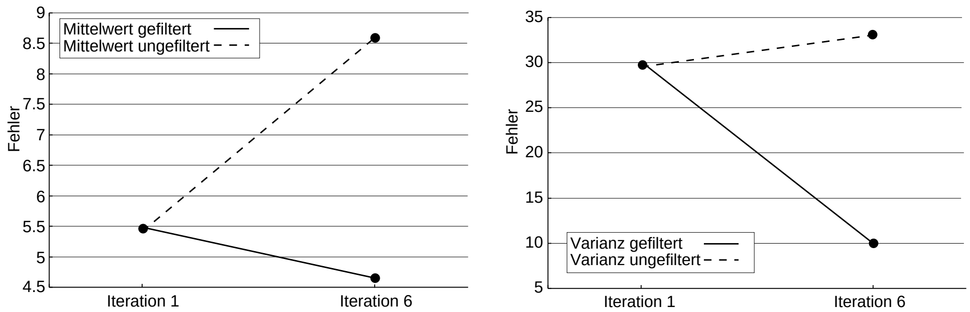


Abbildung 5.18: Vergleich Mittelwerte (links) und Varianzen (rechts) der Abstandsfehler für gefiltertes und ungefiltertes Modell (1. und 6. Iteration)

graphisch dar. Die Verbesserung der Werte beim gefilterten Modell steht in Gegensatz zur deutlichen Verschlechterung bei der Berechnung unter Verwendung des ungefilterten Modells.

[1]siehe Ergebnisse in Kap. 5.2.2.2

Auswertung Orientierungsfehler

Betrachtet man die Ergebnisse für Mittelwerte und Varianzen der Orientierungsfehler (siehe Tabelle 5.2), so läßt sich auch hier ablesen, daß die Verwendung des gefilterten Modells zu besseren Ergebnissen führt als die Verwendung des ungefilterten Modells. Die Ausgangswerte für Mittelwerte und

Mittelwert		
Iterationen	gefiltert	ungefiltert
1	0.19137	0.18231
6	0.12956	0.15151

Varianz		
Iterationen	gefiltert	ungefiltert
1	0.04404	0.04260
6	0.01694	0.02311

Tabelle 5.2: Vergleich gefiltertes und ungefiltertes Modell: Mittelwerte (links) und Varianzen (rechts) der Orientierungsfehler

Varianzen der Orientierungsfehler sind, wie bei den Positionsfehlern, ungefähr gleich groß. Im Unterschied zu den Positionsfehlern tritt hier aber eine Verbesserung sowohl bei gefiltertem als auch bei ungefiltertem Modell auf. Lediglich der Grad der Verbesserung unterscheidet sich, mit gefiltertem Modell fällt diese stärker aus als bei der Verwendung des ungefilterten Modells. In der Abb. 5.19 sind diese Ergebnisse anschaulich graphisch dargestellt.

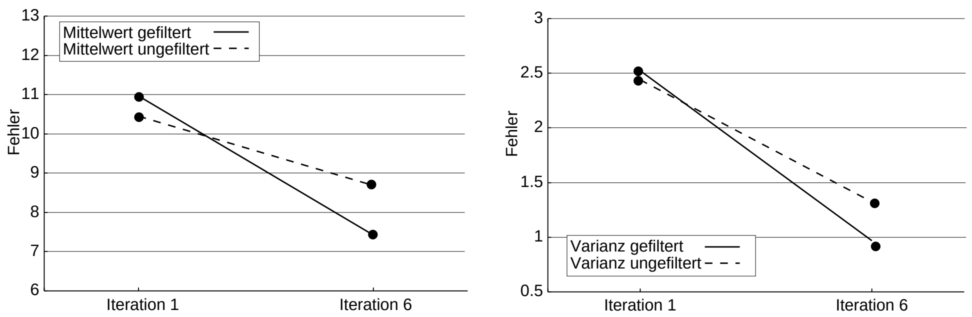


Abbildung 5.19: Vergleich Mittelwerte (links) und Varianzen (rechts) der Orientierungsfehler für gefiltertes und ungefiltertes Modell (1. und 6. Iteration)

Darstellung der Positionsfehler als Histogramm

Betrachtet man die Histogramme der Fehlerverteilung von gefiltertem und ungefiltertem Modell nach der sechsten Iteration in Abb. 5.20, so ist zu erkennen, wie bei Verwendung des ungefilterten Modells deutlich mehr Segmente

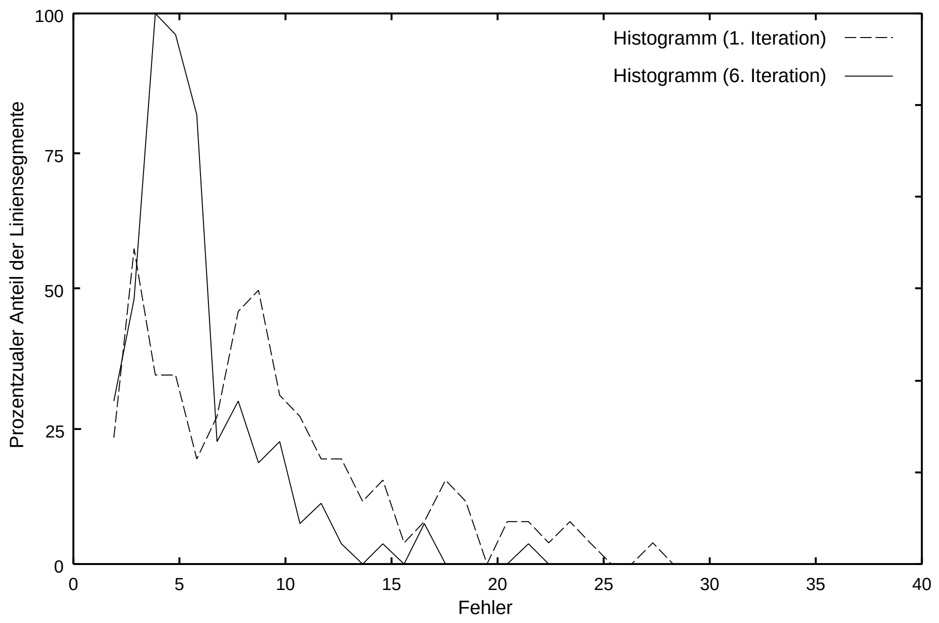


Abbildung 5.20: Histogramme der Abstandsfehler für das gefilterte und das ungefilterte Modell (beides 6. Iteration)

mit einem höheren Fehler auftreten, als das beim gefilterten Modell der Fall ist. Dort liegt eine Häufung von Liniensegmenten mit einem kleinen Fehler vor.

Fazit

Untersucht wurde die Auswirkung des Filterungsverfahrens auf die Akkumulation. Die Ergebnisse dieser Untersuchung lassen die folgenden Aussagen zu:

- Die Positionsfehler der berechneten 3D-Repräsentation nehmen bei der Verwendung des ungefilterten Modells im Laufe des Akkumulationsverfahrens zu. Bei Einsatz des Filterungsverfahrens wird dagegen die erwartete Abnahme der Fehler und damit eine Verbesserung der 3D-Repräsentation erreicht. Dies betrifft sowohl den mittleren Fehler, als auch die Varianz.

- Bei den Orientierungsfehlern wird auch bei Verwendung des ungefilterten Modells eine Verbesserung erzielt. Diese fällt allerdings nicht so hoch aus wie beim Einsatz des gefilterten Modells.

Ohne das Filterungsverfahren ist es nicht möglich, eine stabile Akkumulation der 3D-Repräsentation zu erzielen. Dies zeigt die deutliche Verschlechterung der Positionsfehler bei Verwendung des ungefilterten Modells.

Die Filterung ist eine direkte Vorverarbeitung für die Poseschätzung. Mit ungefiltertem Modell ist es nicht möglich, eine gute Schätzung der Pose vorzunehmen. Das Filterungsverfahren gewährleistet eine gute Poseschätzung, und damit ein erfolgreiches Akkumulieren der 3D-Repräsentation.

5.2.3 Fazit

Die aus der visuellen Auswertung gewonnene Ansicht, daß das Akkumulationsverfahren erfolgreich die Anzahl und die Qualität der rekonstruierten Liniensegmente in der berechneten 3D-Repräsentation erhöht, findet in der statistischen Untersuchung ihre Bestätigung.

Desweiteren konnte mit Hilfe der statistischen Auswertung belegt werden, welche wichtige Rolle das Filterungsverfahren als Vorverarbeitung der Poseschätzung für eine erfolgreiche Akkumulation einnimmt.

5.3 Videosequenz "Duplo"

Für diese Videosequenz wurde ein mauerähnliches Objekt verwendet (siehe Abb. 5.21). Durch den Aufbau aus einzelnen Steinen besitzt dieses Objekt nicht nur äußere Kanten, sondern auch auf seiner Oberfläche sind Linienstrukturen zu finden. Damit sollten bei diesem Objekt nicht nur an den Außenkanten, sondern auch auf der Oberfläche Liniensegmente rekonstruierbar sein. Das Objekt dreht sich von links nach rechts.

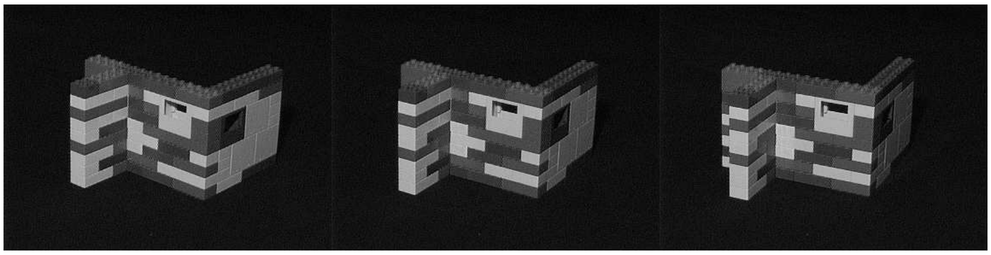


Abbildung 5.21: Videosequenz "Duplo"

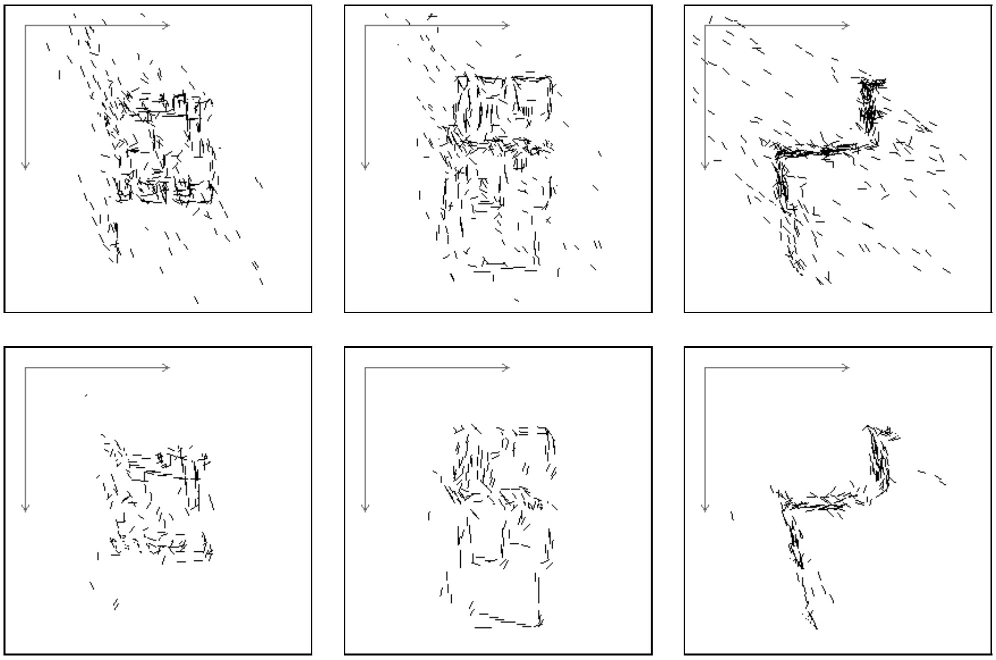


Abbildung 5.22: Vergleich 3D-Stereorekonstruktion (oben)und akkumulierte Repräsentation (unten)

5.3.1 Visuelle Auswertung

Der Vergleich zwischen der reinen 3D-Rekonstruktion und der durch das Akkumulationsverfahren gewonnenen 3D-Repräsentation zeigt die Abb. 5.22. Oben ist die nicht akkumulierte Repräsentation gezeigt, jeweils in den drei Projektionen auf die Koordinatenebenen. Unten sind dieselben Projektionen für die akkumulierte Repräsentation abgebildet. Deutlich zu erkennen sind in der oberen Bildreihe die falsch rekonstruierten Liniensegmente, die durch das Akkumulationsverfahren fast vollständig entfernt wurden. Die untere Bildreihe zeigt bis auf sehr wenige Fehler das akkumulierte Mauerobjekt.

Wie sich in den Akkumulationsschritten die Repräsentation nach und nach vervollständigt, ist in der Abb. 5.23 dargestellt. Von oben nach unten sind jeweils die drei Projektionen auf die Koordinatenebenen für einzelne Akkumulationsschritte abgebildet. Die anfänglich dünn besetzte 3D-Repräsentation wird in jedem Schritt durch weitere Liniensegmente erweitert.

In der Abb. 5.24 sind die Projektionen der Repräsentationen auf das Kamerabild für die erste (oben) und letzte (mitte) Iteration der Akkumulation gegenübergestellt. Das mittlere Bild des letzten Iterationsschrittes weist im

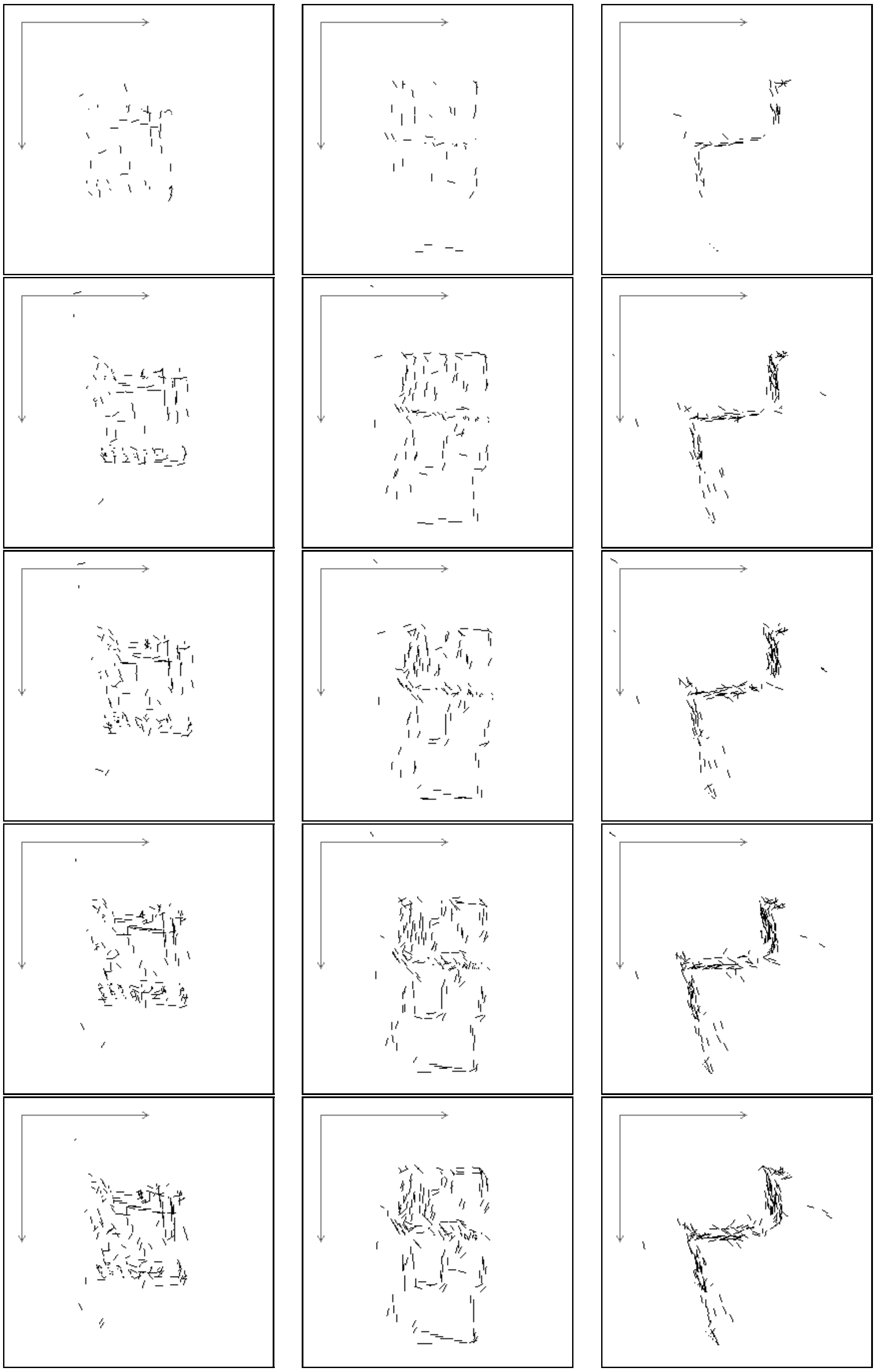


Abbildung 5.23: Akkumulation auf der Sequenz "Duplo"

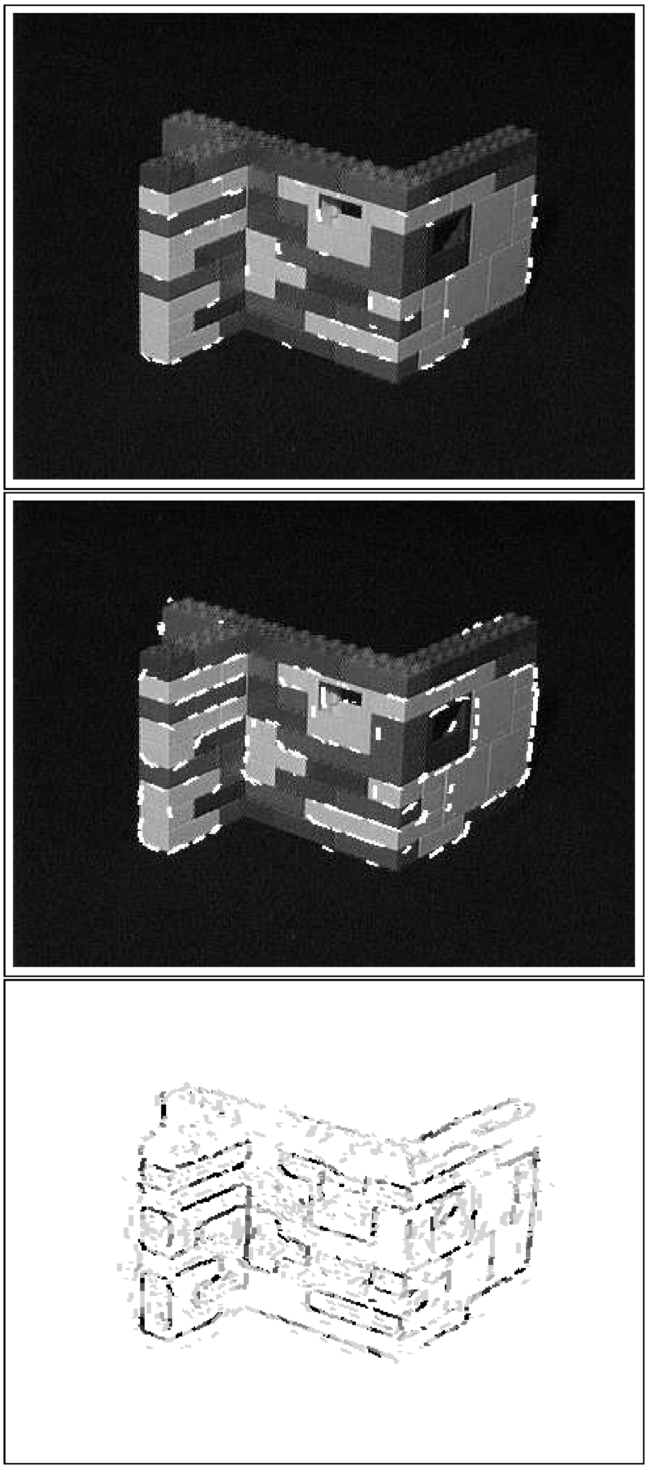


Abbildung 5.24: Vergleich unvollständige (oben), akkumulierte 3D-Repräsentation (mitte) und konfidenzbewertete Liniensegmente (unten)

Unterschied zum oberen Bild, welches den Beginn darstellt, eine deutlich höhere Anzahl von 3D-Liniensegmenten auf.

Unten in Abb. 5.24 sind alle Liniensegmente, eingefärbt nach ihrer Konfidenzbewertung gezeigt. Dies zeigt, das fehlerhafte Segmente weiterhin vorhanden, aber mit einer geringen Bewertung versehen sind.

5.3.2 Fazit

Die Darstellung der 3D-Repräsentation als Projektion auf die drei Koordinatenebenen zeigen, daß das Akkumulationsverfahren auch auf dieser Videosequenz erfolgreich fehlerhaft rekonstruierte Liniensegmente eliminiert. Die Gesamtzahl der Segmente erhöht sich. Die Projektion der Repräsentation auf das Kamerabild zeigt, daß das Objekt gut, wenn auch bei weitem nicht vollständig rekonstruiert wurde.

5.4 Videosequenz “Lego”

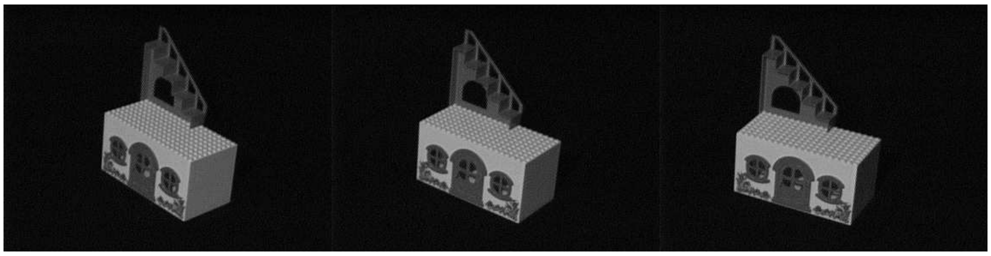


Abbildung 5.25: Videosequenz “Lego”

In dieser Sequenz findet ein weiteres Modell eines Hauses Verwendung. Es besitzt eine kompliziertere Struktur als das obige Modell “Haus”. Die Vorderseite ist besetzt mit zwei Fenstern und einer Tür, welche runde Formen aufweisen. Das Dach ist nicht glatt, sondern regelmäßig genoppt.

5.4.1 Visuelle Auswertung

Die Abb. 5.26 zeigt den Vergleich zwischen einfacher 3D-Stereorekonstruktion (oben) und der akkumulierten Repräsentation (unten) in der Darstellungsweise der Projektion auf die drei Koordinatenebenen. Zu beobachten ist, daß die einfache Stereorekonstruktion viele fehlerhaft rekonstruierte Liniensegmente enthält.

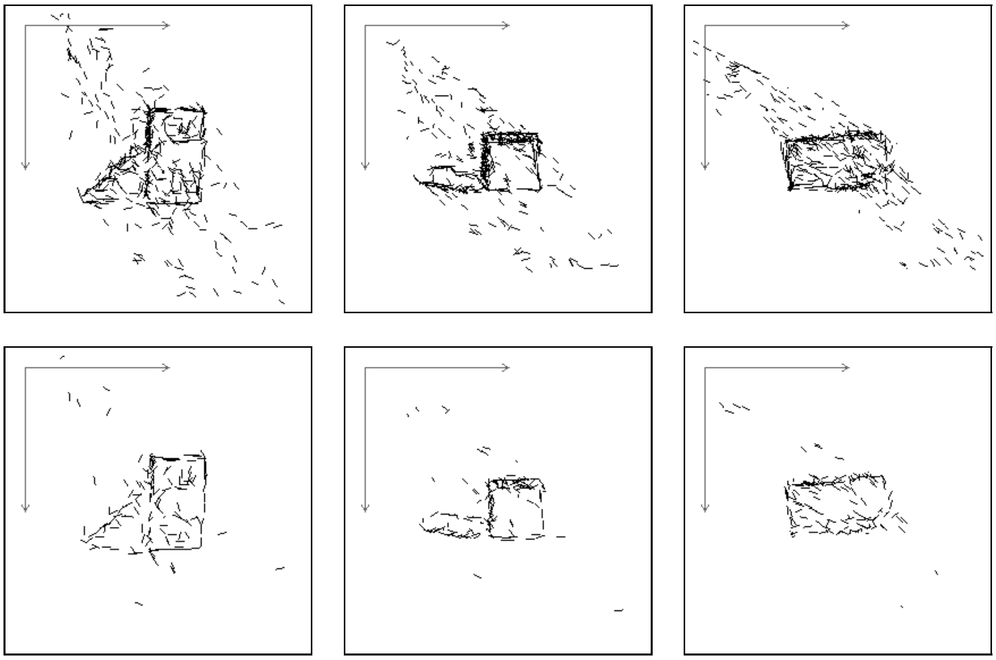


Abbildung 5.26: Vergleich 3D-Stereorekonstruktion (oben) und akkumulierte Repräsentation (unten)

In der Aufsicht (Abb. oben rechts) ist deutlich die Art und Weise sichtbar, in der die fehlerhaft rekonstruierten Segmente verteilt sind. Sie befinden sich alle in einem Kegel, der seine Spitze in der linken oberen Ecke des Bildes besitzt, und der durch die äußeren Begrenzungen des Hauses aufgespannt wird. Die linke obere Ecke ist die Position der Kamera. Der Kegel wird also durch die Sehstrahlen zwischen der Kamera und dem Objekt gebildet.

Die Abb. 5.27 zeigt von oben nach unten die schrittweise Vervollständigung der 3D-Repräsentation. Dies wird auch in der Darstellungsweise der Projektion auf das Kamerabild in Abb. 5.28 deutlich. Fehler treten auf dem Dach des Hauses auf. Die genoppte Struktur ist zu gleichförmig, und erschwert demzufolge eine Korrespondenzfindung für die Stereorekonstruktion, so daß es zu Fehlern kommt. In Abb. 5.28 (unten) sind alle Liniensegmente gezeigt, eingefärbt nach ihrer Konfidenzbewertung.

5.4.2 Fazit

Auch auf der Videosequenz “Lego” konnte das Akkumulationsverfahren erfolgreich angewandt werden. Die Menge der Liniensegmente in der 3D-Repräsentation ist erhöht worden. Die Anzahl der fehlerhaft rekonstruierten

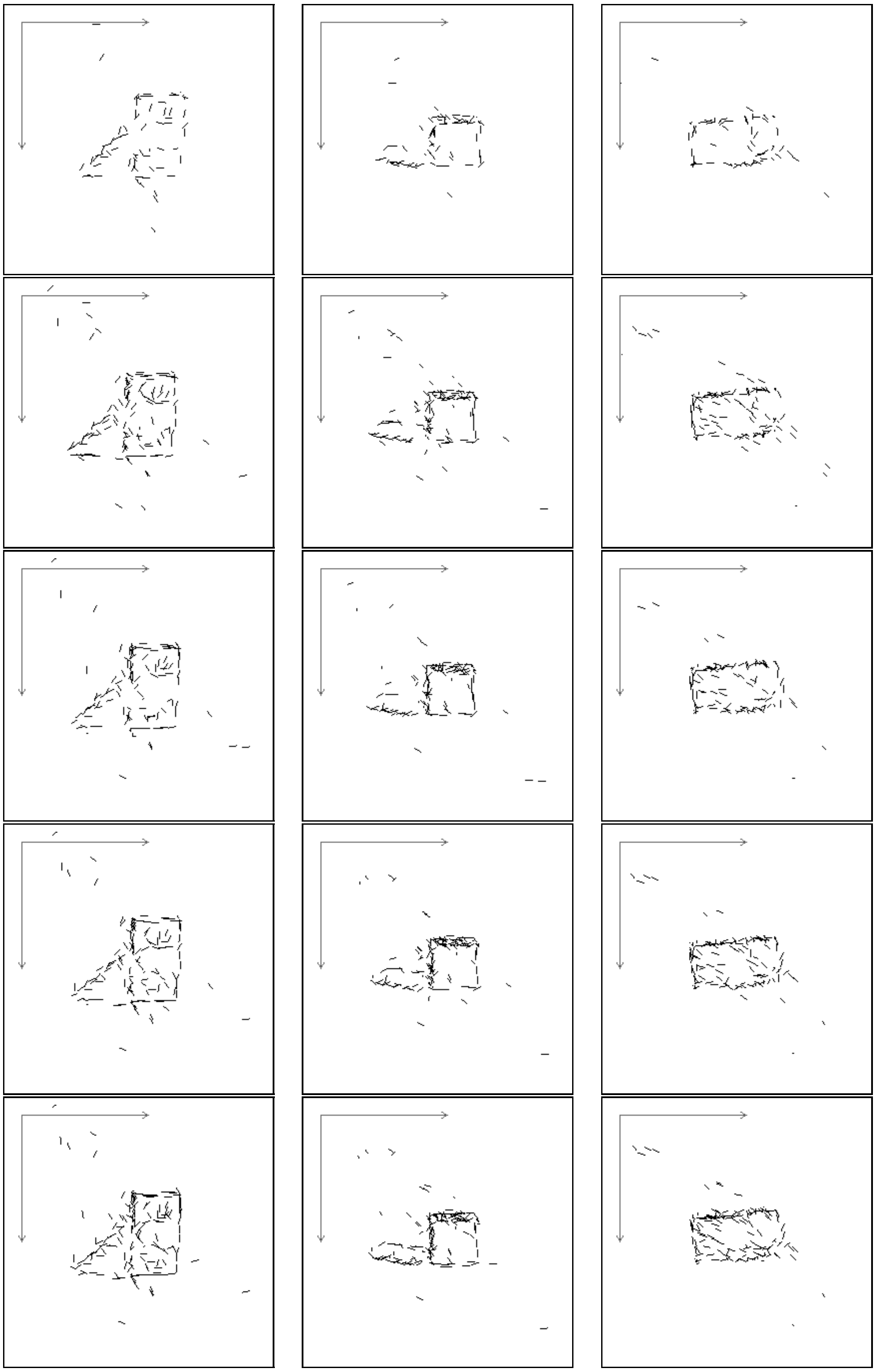


Abbildung 5.27: Akkumulation auf der Sequenz "Lego"

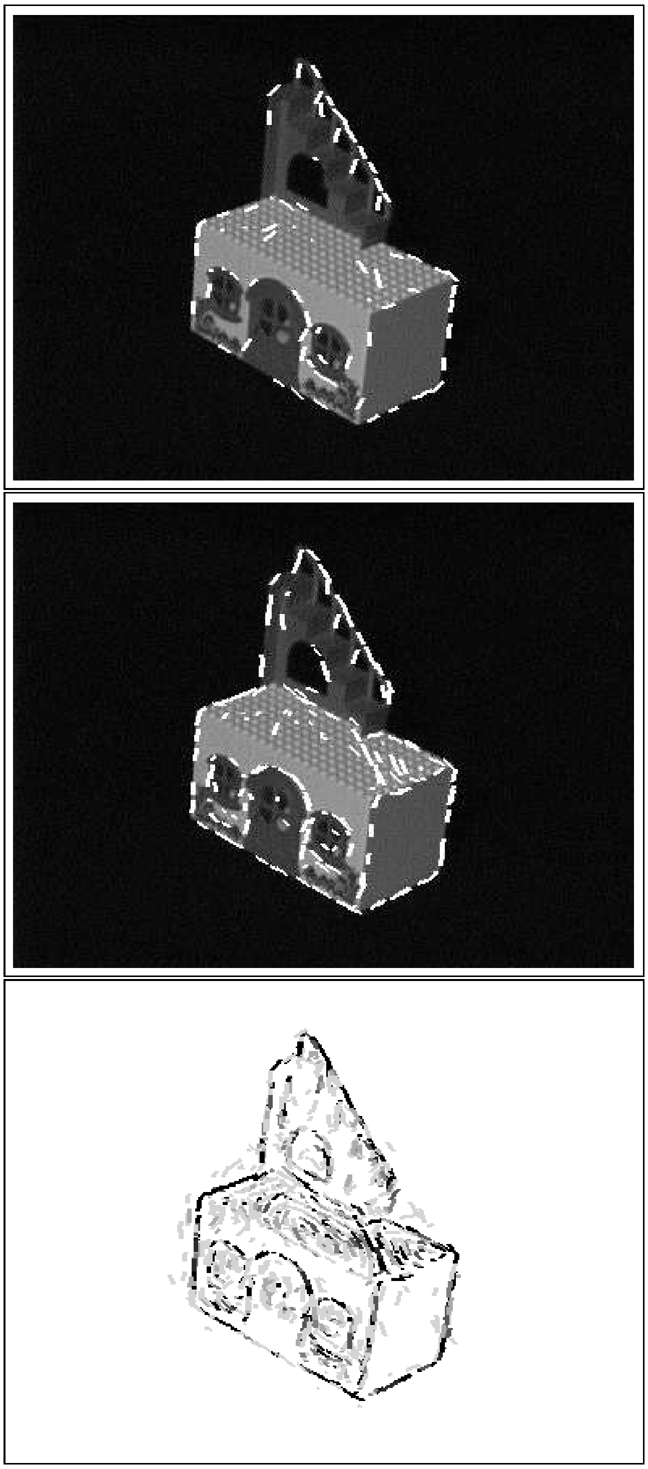


Abbildung 5.28: Vergleich unvollständige (oben), akkumulierte 3D-Repräsentation (mitte) und konfidenzbewertete Liniensegmente (unten)

Segmente konnte stark reduziert werden.

5.5 Laufzeit

Berechnungsschritt		**Zeit (Sekunden)**		**Anteil (%)**	
Stereorekonstruktion					
Teilschritte	Strukturmultivektor		40.0		71.4%
	2D-Liniensegmente		0.5		0.9%
	3D-Liniensegmente		3.5		6.2%
Stereo insgesamt		44.0		78.5%	
Filterung		1.5		2.7%	
Optischer Fluß		9.5		17.0%	
Poseschätzung		0.5		0.9%	
Akkumulation		0.5		0.9%	
Gesamt		56.0		100%	

Tabelle 5.3: Laufzeiten für die einzelnen Berechnungsschritte und Gesamtlaufzeit

Die Laufzeiten der einzelnen Verfahren zeigt die Tabelle 5.5. Ermittelt wurden diese Zeiten auf einer CPU mit 700 MHz unter Verwendung von Bildern mit einer Größe von 384×288 (Viertel-PAL-Auflösung). Ein Akkumulationsschritt benötigt also insgesamt etwas unter einer Minute. Durch die Verwendung kleinerer Bilder würde eine erhebliche Verringerung der Laufzeit erreicht werden. So benötigt etwa die Rekonstruktion auf Bildern der Größe 256×256 nur noch 15 Sekunden. Auch durch eine Parallelisierung des Stereo- und des Flußalgorithums und Einsatz eines Multiprozessorsystems wäre eine weitere Beschleunigung erzielbar. Vor allem ließe sich dadurch die Laufzeit für die Berechnung des Strukturmultivektors, der mit über 70% an der Gesamtzeit beteiligt ist, verringern.

5.6 Fazit

Auf den drei untersuchten Videosequenzen wurde erfolgreich der Integrationsalgorithmus angewendet. Unter Ausnutzung des Statistik- und des Bewegungszusammenhanges konnten stabile 3D-Repräsentationen berechnet und zuverlässige Poseschätzungen durchgeführt werden.

Kapitel 6

Zusammenfassung und Ausblick

In der vorliegenden Arbeit wurde die Interaktion verschiedener visueller Modalitäten mit dem Ziel der Gewinnung einer stabilen 3D-Repräsentation untersucht. Verwendung fanden die visuellen Modalitäten 3D-Stereorekonstruktion, optischer Fluß, Poseschätzung und Akkumulation. Ein Algorithmus wurde entwickelt, innerhalb dessen diese verschiedenen Modalitäten in geeigneter Weise miteinander interagieren. Ergebnis dieser Interaktion ist die Extraktion einer stabilen dreidimensionalen Repräsentation in Form von 3D-Liniensegmenten eines beobachteten und sich in Bewegung befindlichen Objektes.

In der Literatur wurden Verfahren gefunden, welche ebenfalls unterschiedliche Modalitäten wie Flußverfahren oder Stereo zur Gewinnung der dreidimensionalen Struktur und Bewegung einer Szene bzw. eines Objektes benutzen ([8, 42, 46, 21]). Gewonnen werden dabei teils Tiefenkarten der betrachteten Szene, oder die 3D-Positionen und -Bewegungen von feature points, d.h. von charakteristischen Szene- oder Objektpunkten. Die vorliegende Arbeit unterscheidet sich dadurch, daß sie versucht, das beobachtete Objekt mittels einer dreidimensionalen Rekonstruktion seiner Kanten möglichst vollständig zu repräsentieren, und die Stabilität dieser Repräsentation durch die iterative Anwendung des Akkumulationsverfahrens zu erhöhen.

Im Unterschied zu Ackermann [28, 1] werden sämtliche Informationen zur Berechnung der stabilen 3D-Repräsentation auf visuellem Weg gewonnen, während in [28, 1] die Bewegungen kontrolliert mit einem Roboterarm durchgeführt werden. Diese aktive Kontrolle der Bewegung bietet den Vorteil, daß die zur Akkumulation nötigen Bewegungsparameter sehr genau und damit verlässlich sind.

Die vorliegende Arbeit zeigt, daß es möglich ist, auf diesen Kontrollmechanismus zu verzichten, und weiterhin eine Stabilisierung der 3D-Repräsentation zu erreichen, wenn schon eine Eigenbewegung der Objekte vorliegt.

Damit wird gezeigt, daß zuverlässiges Weltwissen durch die alleinige Beobachtung der Welt gewonnen werden kann, wobei die Bewegung nach wie vor in einem Wahrnehmungs-Handlungs-Zyklus generiert werden kann, jedoch ohne die exakten Parameter der generierten Bewegung zu kennen. Da eine Laborumgebung mit einem Manipulationsroboter nicht mehr unbedingt notwendig ist, ist das Verfahren flexibler. Auch weiter entfernte, oder größere Objekte können erfaßt werden. Die Vorteile in der Verwendung der Robotik liegen in der Kontrolle über das beobachtete Objekt. Es ist dadurch z.B. möglich, sich bewußt bestimmte Ansichten zu generieren. Mit dem in dieser Arbeit entwickelten Verfahren ist kein genaues Wissen um die Bewegungsparameter dieser generierten Ansichten mehr nötig. Durch Hinzufügen einer haptischen Modalität bei Verwendung der Robotik ließe sich eine weitere Informationsquelle erschließen.

Ein Teilaspekt dieser Arbeit ist, daß eine erfolgreiche Schätzung der Pose auch auf der Basis unzuverlässiger und fehlerbehafteter 3D-Repräsentationen durchgeführt werden kann, wenn gewisse heuristische Annahmen über die statistischen Regularitäten des betrachteten Objekts zugrundegelegt werden. Die Ausprägung dieser Annahmen findet sich in dem entworfenen Filterungsverfahren, welches darauf basiert, daß Kollinearität ein häufig anzutreffendes Merkmal ist.

In dieser Arbeit werden als erscheinungsbasierte Merkmale Orientierung und Phase verwendet. Erweiterungsmöglichkeiten sind gegeben durch die Hinzunahme weiterer beschreibender Merkmale wie z.B. Farbe ([14]) oder Textur. Das Verfahren zur Filterung der Repräsentationen basiert in dieser Arbeit auf einer heuristischen Parameterfindung. Hier wäre es für zukünfige Arbeiten interessant und wichtig, diese Heuristik durch ein statistisch begründetes Modell zu ersetzen. Von Relevanz für die Poseschätzung, das Akkumulationsverfahren und die Filterung ist die Inhomogenität des projizierten dreidimensionalen Raumes. Tiefer im Raum liegende Strukturen erscheinen in der Projektion kleiner als solche, die sich näher an der Kamera befinden. So haben Fehler oder Ungenauigkeiten in der zweidimensionalen Ortsbestimmung von Strukturen einen umso größeren Folgefehler für den dreidimensionalen Ort, je weiter die Strukturen von der Kamera entfernt liegen. Eine Modellierung dieser Zusammenhänge sollte es möglich machen, die genannten Verfahren qualitativ zu verbessern ([31]).

Abbildungsverzeichnis

Literaturverzeichnis

[1] M. Ackermann. Akkumulieren von Objektrepräsentationen im Wahrnehmungs-Handlungs-Zyklus. Diplomarbeit am Lehrstuhl für Kognitive Syteme der Christian-Albrechts-Universität zu Kiel, 2000.

[2] John (Yiannis) Aloimonos and David Shulman. *Integration of Visual Modules: An Extension of the Marr Paradigm.* Academic Press Inc., 1989.

[3] P. Anandan. *Measuring Visual Motion from Image Sequences.* PhD thesis, Univ. of Massachusetts, Amherst, MA, 1987.

[4] P. Anandan. A computational framework and an algorithm for the measurement of visual motion. *International Journal of Computer Vision*, 2(3):283–310, January 1989.

[5] N. Ayache. *Stereovision and Sensor Fusion.* MIT Press, 1990.

[6] J. L. Barron, D. J. Fleet, and S. S. Beauchemin. Performance of optical flow techniques. *IJCV*, 12(1):43–77, 1994.

[7] C. Bregler and J. Malik. Video motion capture, 1998.

[8] T. Brodský, C. Fermueller, and Y. Aloimonos. Beyond the epipolar constraint: Integrating 3D motion and structure estimation. *Lecture Notes in Computer Science*, 1506:109–??, 1998.

[9] E. Brunswick and J. Kamiya. Ecological cue–validity of 'proximity' and of other Gestalt factors. *American Journal of Psychologie*, LXVI:20–32, 1953.

[10] W. K. Clifford. Applications of Grassmann's extensive algebra. *Am. J. Math.*, 1:350, 1878.

[11] O. Faugeras. *Three-dimensional computer vision. A geometric viewpoint.* MIT Press, Cambridge, MA, 3. edition, 1999.

[12] M. Felsberg and G. Sommer. Structure multivector for local analysis of images. In *Proc. Dagstuhl Workshop on Theoretical Foundations of Computer Vision*, 2000. accepted.

[13] David J. Fleet and Allan D. Jepson. Computation of component image velocity from local phase information. *International Journal of Computer Vision*, 5(1):77–104, 1990.

[14] C. Gebken. Stero-Korrespondenzanalyse mittels Farbinformationen. Studienarbeit am Lehrstuhl für Kognitive Syteme der Christian-Albrechts-Universität zu Kiel, 2001.

[15] W. S. Geisler, J. S. Perry, B. J. Super, and D. P. Gallogly. Edge co-occurrence in natural images predicts contour grouping performance. *Vision Research*, 41:711–724, 2001.

[16] O. Granert. Posenschätzung von kinematischen Ketten. Diplomarbeit am Lehrstuhl für Kognitive Syteme der Christian-Albrechts-Universität zu Kiel, 2001.

[17] H. Grassmann. Der Ort der Hamilton'schen Quaternionen in der Ausdehnungslehre. *Math. Ann.*, 12:375, 1877.

[18] S. Gull, A. Lasenby, and C. Doran. Imaginary Numbers are not Real - the Geometric Algebra of Spacetime. *Found. Phys.*, 23(9):1175–1201, 1993.

[19] R. E. Harke. An Introduction to the Mathematics of the Space-Time Algebra. http://home.earthlink.net/~sueharke/rharke/ps/intro.ps.gz, 1998.

[20] D. J. Heeger. Optical flow using spatiotemporal filters. *International Journal of Computer Vision*, 1:279–302, 1988.

[21] A. Y. Ho and T. Pong. Cooperative fusion of stereo and motion. *Pattern Recognition*, 29(1):121–130, 1996.

[22] Berthold K. P. Horn and Brian G. Schunck. Determining optical flow. *Artificial Intelligence*, 17(1-3):185–203, August 1981.

[23] Katsushi Ikeuchi and Berthold K. P. Horn. Numerical shape from shading and occluding boundaries. *Artificial Intelligence*, 17(1–3):141–184, 1981.

[24] B. Jähne. *Digitale Bildverarbeitung.* Springer-Verlag, Berlin, 1997.

[25] R. E. Kalman. A new approach to linear filtering and predictive problems. *Transactions ASME, Journal of basic engineering*, (82):34–45, 1960.

[26] R. Klette, K. Schlüns, and A. Koschan. *Computer Vision - Three-Dimensional Data from Images*. Springer, 1998.

[27] N. Krüger. Collinearity and parallism are statistically significant second order relations of complex cell responses. *Neural Processing Letters*, 8(2):117–129, 1998.

[28] N. Krüger, M. Ackermann, and G. Sommer. Accumulation of object representations utilizing interaction of robot action and perception. *Knowledge Based Systems*, (13), 2002.

[29] N. Krüger, M. Felsberg, M. Pörksen, and G. Sommer. A new stereo algorithm which integrates geometrical and structural information. *to be submitted.*

[30] N. Krüger and F. Wörgötter. Multi modal estimation of collinearity and parallelism in natural image sequences. *Submitted.*

[31] Norbert Krüger and Bodo Rosenhahn. These to the unisotropy of the perceptive space and its consequences for stereo, pose esrimation, grouping and learning. *in progress.*

[32] H. A. Mallot. *Sehen und die Verarbeitung visueller Information.* Vieweg, Braunschweig/Wiesbaden, 2. edition, 2000.

[33] H. Nagel. Displacement vectors derived from second-order intensity variations in image sequence. *Computer Vision, Graphics, and Image Processing*, 21:85–117, 1983.

[34] Hans-Hellmut Nagel. On the estimation of optical flow: Relations between different approaches and some new results. *Artificial Intelligence*, 33(3):299–324, November 1987.

[35] Xavier Pennec. Computing the mean of geometric features application to the mean rotation. Technical Report RR-3371, Inria, Institut National de Recherche en Informatique et en Automatique.

[36] C. Perwass. *Applications of Geometric Algebra in Computer Vision.* PhD thesis, Cambridge University, 2000.

[37] M. Pörksen. 3D Objektrepräsentation basierend auf dem Strukturmultivektor. Diplomarbeit am Lehrstuhl für Kognitive Syteme der Christian-Albrechts-Universität zu Kiel, 2001.

[38] B. Rosenhahn, O. Granert, and G. Sommer. Monocular pose estimation of kinematic chains. In *Applied Geometric Algebra for Computer Science and Engineering*. Birkhäuser Verlag, to appear 2001.

[39] B. Rosenhahn, N. Krüger, T. Rabsch, and G. Sommer. Tracking with a novel pose estimation algorithm. In R. Klette and G. Sommer, editors, *International Workshop on Robot Vision; Auckland, New-Zealand*, LNCS 1998, pages 9–19, Heidelberg, 2001. Springer-Verlag.

[40] B. Rosenhahn and J. Lasenby. Constraint Equations for 2D-3D Pose Estimation in Conformal Geometric Algebra. Technical Report CUED/F - INFENG/TR.396, Cambridge University Engineering Department, 2000.

[41] B. Rosenhahn, Y. Zhang, and G. Sommer. Pose estimation in the language of kinematics. In G. Sommer and Y. Zeevi, editors, *2nd International Workshop on Algebraic Frames for the Perception-Action Cycle, AFPAC 2000, Kiel*, volume 1888 of *LNCS*, pages 284–293, Heidelberg, 2000. Springer-Verlag.

[42] Y. Q. Shi, C. Q. Shu, and J. N. Pan. Unified optical flow field approach to motion analysis from a sequence of stereo images. *Pattern Recognition*, 27(12):1577–1590, 1994.

[43] M. Sigman, G.A. Cecchi, C.D. Gilbert, and M.O. Magnasco. On a common circle: Natural Scenes and Gestalt Rules. *PNAS*, 98(4):1935–1949, 2001.

[44] G. Sommer, B. Rosenhahn, and Y. Zhang. Pose estimation using geometric constraints. In R. Klette, Th. Huang, and G. Gimmel'farb, editors, *Multi-Image Search and Analysis*, volume 2032 of *LNCS*, pages 153–170, Heidelberg, 2001. Springer-Verlag.

[45] R. Y. Tsai. An efficient and accurate camera calibration technique for 3-D machine vision. In *Proc. IEEE Computer Vision and Pattern Recognition*, pages 364–374, June 1986. Miami Beach, FL.

[46] Wendong Wang and James H. Duncan. Recovering the three-dimensional motion and structure of multiple moving objects from binocular image flows. *Computer Vision and Image Understanding: CVIU*, 63(3):430–446, May 1996.

[47] R. Zavodnik and H. Kopp. *Graphische Datenverarbeitung, Grundzüge und Anwendungen.* Hanser-Verlag, München, 1995.

[48] C. Zetsche and E. Barth. Fundamental limits of linear filters in the viusal processing of two dimensional signals. *Vision Research*, 30, 1990.

Danksagung

Bedanken möchte ich mich bei Dr. Norbert Krüger und Dr. Christian Perwass für die hilfreiche Betreuung meiner Diplomarbeit. Mein Dank gilt allen Beteiligten am KiVi-Projekt. Ohne dieses Projekt wäre diese Arbeit nicht möglich gewesen. Ganz besonderer Dank gilt meinen Eltern für ihre materielle und moralische Unterstützung meines Studiums und dieser Diplomarbeit.

Erklärung

Hiermit versichere ich, daß ich die vorliegende Arbeit selbständig verfaßt und außschließlich die angegebenen Hilfsmittel und Quellen verwendet habe.

Kiel, den 16. März 2002