Diplomarbeit

Integration visuell-aktorischer Kompetenzen im Wahrnehmungs-Handlungs-Zyklus

am Lehrstuhl für Kognitive Systeme
Institut für Informatik und Praktische Mathematik
Christian–Albrechts–Universität zu Kiel
Prof. Dr. G. Sommer

ausgearbeitet von:

Daniel Wendorff

Betreuer: Dr. Norbert Krüger

April 2001

Inhaltsverzeichnis

1 Einleitung

Die folgende Sichtweise über die Funktionsweise sehender Roboter war lange Zeit populär: Ein aus Bildern extrahiertes oder a priori gegebenes Weltmodell ist Grundlage für die Generierung eines Plans, der dann von den aktorischen Komponenten des Roboters ausgeführt wird. Dieses Paradigma wurde von Brooks [5] als 'sense-model-plan-act (SMPA)' Schema bezeichnet und ausführlich kritisiert. Es ist gekennzeichnet von einer starken Trennung bzw. einer großen Distanz zwischen Handlungs– und Wahrnehmungskomponenten. Das SMPA Paradigma weist wesentliche, konzeptionelle Schwächen auf und ist auch in biologischen Systemen nicht in dieser Form realisiert. Es hat sich gezeigt, daß Wahrnehmung und Handlung sehr viel enger miteinander verknüpft sind. Sie können in der Tat so miteinander verwoben sein, daß Wahrnehmung und Handlung als zwei Seiten derselben Medaille betrachtet werden müssen [12, 34].

Granlund [12] stellt diesem SMPA Konzept die Idee einer Wahrnehmungs-Handlungspyramide entgegen, indem Sensordaten über Abstraktionsebenen verschiedener Komplexitäten mit Handlungen verknüpft werden. Ein wesentlicher Aspekt einer solchen Struktur ist die Verkoppelung von Wahrnehmungs-Handlungskompetenzen verschiedener Qualität, die sich zu der jeweiligen Situation in angepaßten Strategien ergänzen. So kann zum Beispiel bei dem Greifen eines Objektes erst eine schnelle grobe Bewegung auf das Objekt hin stattfinden. Wenn das System den Roboterarm hinreichend nahe an das Objekt und es dieses in sein Sichtfeld gebracht hat, kann es relative Abstände zwischen Objekt und Greifer benutzen, um eine präzisere Annäherung zu erreichen und schließlich das Greifen nach einer Berührung unter zusätzlicher Ausnutzung haptischer Information durchzuführen.

In der hier vorgestellten Diplomarbeit werden verschiedene Kompetenzen bzw. Handlungsmodule miteinander zu komplexeren Kompetenzen verknüpft. Dabei geht es einerseits um die technische Durchführung zweier wichtiger Fähigkeiten eines visuell gesteuerten Roboters: Dem Herstellen eines physikalischen Kontaktes von einem Objekt mit einem Greifer und der Kalibration des Systems ohne künstliche Kalibrationsmuster. Andererseits ist die Bedeutung dieser Kompetenzen als Einleitung noch komplexerer Handlungsmöglichkeiten innerhalb eines sich mit seiner Umwelt auseinandersetzenden Gesamtsystems eine zentrale Motivation dieser Arbeit.

Ein definiertes Anfangsverhalten kann Lernprozesse einleiten beziehungsweise effizienter gestalten. So sind zum Beispiel Babies keine passiven Geschöpfe [17, 28]. Sie explorieren ihre Umwelt in unterschiedlichen Entwicklungsphasen mit Hilfe von Handlungen unterschiedlicher Komplexitätsstufen. Dazu gehören unter anderem die zielgerichtete Augenbewegung, später das Berühren, Betasten und Greifen. In einem späteren Stadium explorieren Babies ihre Umwelt zusätzlich mittels Eigenbewegung (Krabbeln). Diese Anfangsverhalten bauen aufeinander auf und sind voneinander abhängig [17].

Für ein künstliches System stellt sich analog die Frage nach möglichen *Anfangsverhalten*, welche komplexe *Verhaltensweisen* einleiten. Diese hängen von den Möglichkeiten der verwendeten Hardware und dem Ziel des Systems ab. In dieser Arbeit wird ein aktiver Stereokamerakopf und ein Roboterarm verwendet (siehe Abbildung 1). Ein mögliches Ziel des Systems und eine wichtige Fähigkeit menschlicher Perzeption ist die multimodale Exploration von Objekten.

Abbildung 1: Versuchsszene der verwendeten Hardware

In dieser Arbeit wird ein Verhaltensmodul beschrieben, welches innerhalb eines komplexeren Systems die Aufgabe hat, ein Verhalten zur Exploration eines Objektes einzuleiten: Durch die Veränderung der betrachteten Szene durch das Hinzufügen eines neuen Objekts wird die visuelle Aufmerksamkeit des Systems auf dieses Objekt gelenkt, woraufhin der Greifer sich dem Objekt annähert, so daß ein taktiler Kontakt zum Objekt entstehen kann. Die gegenseitige Unterstützung von Wahrnehmung und Handlung als grundlegende Strategie wird innerhalb vieler Kompetenzen verwendet.

Die Annäherung des Greifers zum Objekt ist eine gute Ausgangsbasis für eine Exploration des Objektes. Die Vorteile hierfür sind: Es können relative Abstände für eine Greifstrategie verwendet werden und keine eventuell ungenauen, absoluten Entfernungen. Das Objekt kann durch die Verwendung des Zooms mit einer hohen Auflösung untersucht werden. Hierdurch wird eine höhere Genauigkeit der extrahierten Informationen erzielt. In einem weiteren Schritt kann eine Greifstrategie entwickelt werden, so daß mittels des Greifens die physische Kontrolle über das Objekt erlangt werden kann. Im folgenden könnte dann eine komplexere Objektrepräsentation erzeugt werden, welche die physische Kontrolle über das Objekt voraussetzt [1].

Dieses *einfache* Verhaltensmodul setzt bereits eine Menge von Fähigkeiten voraus, die geordnet miteinander interagieren müssen. Diese Fähigkeiten sind unter anderem: Kontrolle der aktiven Kompetenzen (Binokularkopf und Roboterarm mit Greifer), 3D-Schätzung, Fixierung, Berechnung der Abbildungsparameter nach Eigenbewegung, Kalibration von Kamera- und Robotersystem und eine rudimentäre 2D-Objektrepäsentation. Eine weitere wichtige Kompetenz ist die Sicherheit und der Schutz der Hardware und anderer Objekte, wie z.B. ein Mensch, der sich in der Szene bewegt.

Eine wesentliche Voraussetzung für viele Module ist die Berechnung der Abbildungsparameter des Kamerakopfes nach einer Eigenbewegung. Weiterhin ist die Approximation der Transformation der Koordinatensysteme Welt in Roboter bzw. Roboter in Welt wichtig. Hierfür wurde ein Verfahren entwickelt, welches auf der visuellen Erkennung des Greifers und einer Gütekontrolle basiert. Dieses Verfahren kann während des *normalen* Betriebes innerhalb von anderen Handlungen eingebettet werden und basiert nicht auf künstliche Kalibrationsmerkmale.

Kalibrationsverfahren reagieren in der Regel sehr empfindlich auf falsche Korrespondenzen von 2D-Bildpunkten und 3D-Objektpunkten. Dieses Problem kann durch einfach zu erkennende Kalibrationsmerkmale gemindert werden. Dies stellt jedoch einen schwerwiegenden Eingriff dar, da die Kalibration in einem getrennten Schritt durchgeführt werden muß. Die Kalibration ohne künstliche Kalibrationsmerkmale wird hier möglich, indem das System in der Lage ist, Korrespondenzen mit Hilfe von unabhängigen Modulen zu finden, die zudem jeweils ein Gütemaß für die Qualität der Korrespondenz liefern. Nur wenn mehrere Module miteinander kompatible Korrespondenzen mit hohen Gütemaßen finden, wird eine wirkliche Korrespondenz angenommen und für die Berechnung der Kalibrationsparameter benutzt.

Diese Arbeit gliedert sich wie folgt: In Kapitel 2 werden die Handlungsmodule, welche in dieser Arbeit entwickelt wurden, vorgestellt und in Bezug zu einer Gesamtaufgabe eines möglichen Systems gestellt. In Kapitel 3 wird das Kamerakopfmodell eines aktiven binokularen Systems beschrieben. Es wird darauf eingegangen, wie die Abbildungsparameter bestimmt werden. In Kapitel 4 werden die Grundkompetenzen des Kamerasystems erläutert, die auf dem Modell des Kamerasystems basieren. Diese Module sind wiederum die Grundlage für die Interaktion mit dem Roboterarm. Es werden aufbauend auf den in Kapitel 4 beschriebenen Kompetenzen zwei komplexere Kompetenzen implementiert: 1. Körperlicher Kontakt durch Integration verschiedener Kompetenzen im Wahrnehmungs-Handlungs-Zyklus und 2. Kalibration im Wahrnehmungs-Handlungs-Zyklus. Diese werden in Kapitel 5 beschrieben. Kapitel 6 liefert einen Ausblick.

2 Einordnung des Handlungsmoduls in einen systemischen Zusammenhang

In Abschnitt 2.1 wird die Aufgabe von dieser Arbeit – das Erlangen des körperlichen Kontakts mit einem Objekt – beschrieben. Es wird gezeigt, daß viele einzelne Kompetenzen und deren Interaktionen für diese Aufgabe benötigt werden. Weiterhin wird das häufig verwendete Konzept des Wahrnehmungs-Handlungs-Zykluses näher beschrieben.

2.1 Aufgabenbeschreibung

Das verwendete visuell-aktorische System beinhaltet einen aktiven Stereokamerakopf (TRC Bisight, siehe Abb. 2 [15]) und einen Roboterarm (Stäubli RX-90 [30]) mit einem 2-Backen-Greifer als Manipulator. Das Ziel dieser Arbeit besteht in der Entwicklung eines Systems zum Erlangen des taktilen Kontakts mit einem Objekt.

Abbildung 2: TRC-Kopf

Das anfangs abwartende System reagiert auf eine Veränderung der Umwelt: dem Hinzufügen eines Objekts in die Szene. Die folgenden Handlungen werden durch diese äußere Einflußnahme aktiviert. In der ersten Handlungsphase wird die Aufmerksamkeit der visuellen Komponenten auf die Veränderung gerichtet. Das neue Objekt wird fixiert (siehe Abb. 3). Es wird versucht das Objekt zu detektieren. In der zweiten Handlungsphase versucht das System zum Objekt einen taktilen Kontakt herzustellen (siehe Abb. 4). Hierzu wird der Greifer visuell zum Ziel gesteuert.

Dieses System hat ein bestimmtes Grundwissen von seiner Umwelt, z.B. es kennt eine rudimentäre Objektrepräsentation des Greifers und die ungefähre Entfernung von Roboter und Kamera, wodurch der Aktionsradius vom Greifer festgelegt ist. Ein Problem bei der Verwendung eines aktiven Stereokamerakopfes ist die ständige Berechnung der Abbildungsparameter. Daher sind für die genannten

Abbildung 3: Ablauf Detektion eines neuen Objektes

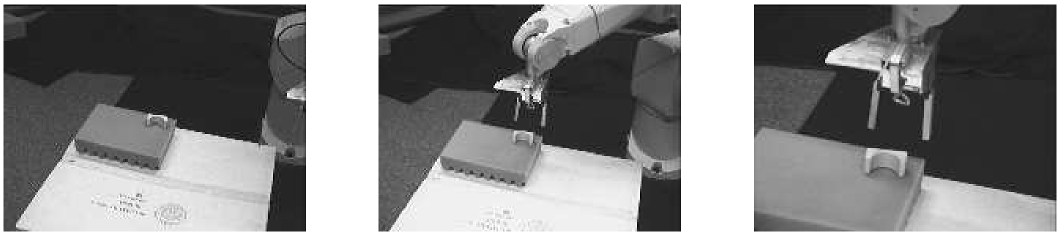


Abbildung 4: Ablauf Aufmerksamkeit und Annäherung zum neuen Objekt

Verhaltensmuster weitere Fähigkeiten notwendig, wie z.B. die ständige Kalibration des Kamerakopfes nach Eigenbewegungen und die Objektwiedererkennung nach Eigenbewegung des Kamerakopfes.

Die Feststellung des köperlichen Kontakts erfolgt durch andere Handlungsmodule. Hierfür können z.B. taktile Sensoren (siehe Abb. 8) verwendet werden. Das Heranführen eines mit taktilen Sensoren ausgestatteten Greifers zum Objekt könnte auch als Ausgangslage für eine taktile Exploration des Objekts genutzt werden. Zur Zeit wird an einer Diplomarbeit am Lehrstuhl 'Kognitive Systeme' gearbeitet, die als Thema unter anderem die Exploration mit einen taktilen Sensor beinhaltet [8].

2.1.1 Herstellung des taktilen Kontakts zu einem neuen Objekt

Für die Handlung des Systems zur Erlangung des taktilen Kontakts zu einem neuen Objekt werden eine Menge von Kompetenzen benötigt. Die Interaktion dieser Fähigkeiten miteinander ergeben die Gesamthandlung.

Die Kompetenzen können in zwei verschiedene Klassen aufgeteilt werden: Basiskompetenzen und zusammengesetzte Kompetenzen. Die Basiskompetenzen sind sehr einfache Fähigkeiten, die keine anderen Kompetenzen verwenden. Zusammengesetzte Kompetenzen fügen andere Kompetenzen zusammen. Die zusammengesetzten Kompetenzen können zudem in der Regel ein Gütemaß über ihre Handlung ermitteln, welches für eine Rückkopplung verwendet werden kann. In der folgenden Auflistung der Basiskompetenzen und der zusammengesetzte Kompetenzen sind jeweils die einfließenden Kompetenzen in Klammern { } vermerkt. In der Abbildung 5 sind die Abhängigkeiten von den einzelnen Kompetenzen mit einem Graph dargestellt.

Basiskompetenzen:

1. Zoom
2. Stereokopfaktionen
3. Aquirierung von Stereobildern
4. Roboterarmbewegung
5. Modellwissen über den Kamerakopf
6. Rudimentäre Objektrepräsentation

Zusammengesetzte Kompetenzen:

a) Detektion eines neuen Objekts {3}
b) 2D-Positionsbestimmung eines neuen Objekts {a}
c) Detektion eines geeigneten Zeitpunktes für eine Roboteraktion {3}
d) Koordinatentransformation Roboter-Kamera {5}
e) Fixieren des visuellen Systems auf einen 3D-Punkt {2, 5}
f) Irissteuerung {2, 3}
g) Schärfesteuerung {2, 3}
h) Berechnung der Abbildungsparameter des visuellen Systems {2, 5}
i) Tiefenschätzung {h}
j) Korrespondenzfindung in Stereobildern {6, b}
k) Wiederfinden eines Objektes in unterschiedlichen Bildsequenzen {3, 6, i}
l) Berechnung der 3D-Position eines Objektes mittels visueller Daten {i, j}
m) Bestimmung der 3D-Position des Greifers mittels visueller Daten {j, k}
n) Fixierung eines Objektes durch das visuelle System {e, k, l}
o) Kalibration Roboter-Kamera im PAC {4, d, m}
p) Annäherung des Greifers zum Objekt {c, n, o}

Zusätzlich haben die verschiedenen Kompetenzen unterschiedliche Prioritäten. So sollte zum Beispiel die Sicherheit der Hardware und der Umwelt die oberste Priorität haben. Die Kalibration Roboter-Kamera sollte für die Interaktion zwischen Roboter und Kamera eine bestimmte Güte K besitzen. Je besser die Kalibrations Roboter-Kamera ist, um so größer ist dieser Wert K. Die Struktur der unterschiedlichen Prioritäten ist in der Tabelle 1 dargestellt. Je weiter oben in der Tabelle eine Kompetenz steht, um so höher ist ihre Priorität für das gesamte System. Wenn mehrere Spalten für einen Bereich z.B. Kalibration Roboter-Kamera vorhanden sind, dann können nur die Handlungen aus

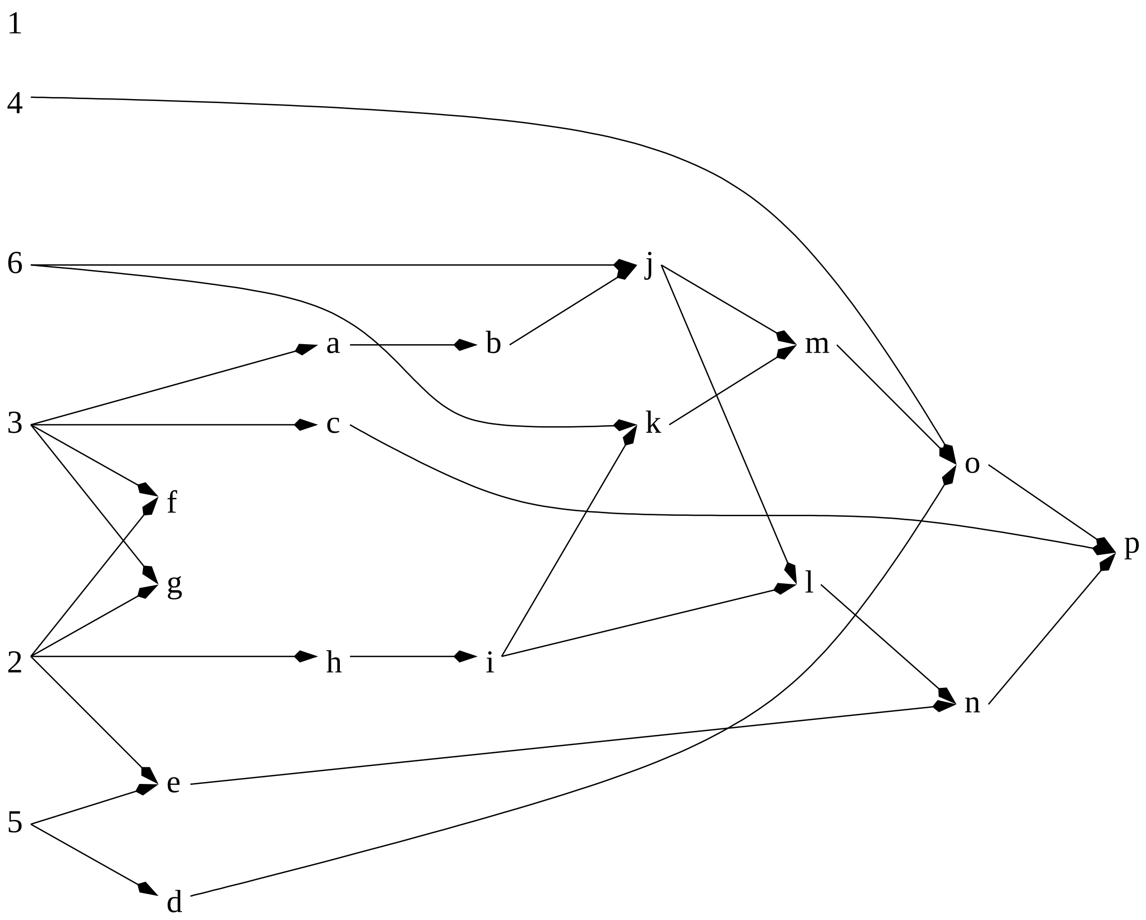


Abbildung 5: Graph mit den Abhängigkeiten der verschiedenen Kompetenzen

einer Spalte durchgeführt werden. Wenn z. B. das Sicherheitskriterium des Gesamtsystems positiv und der Wert $K << 0$ ist, dann kann das System nur die *extra* Kalibration Roboter-Kamera durchführen. Falls zwischenzeitlich das Sicherheitskriterium negativ ausfällt, sollte das System keine weiteren Handlungen durchführen.

Für die Herstellung des taktilen Kontakts zu einem neuen Objekt verfügt das System über ein Grundverhalten. Das System beobachtet die Szene und beurteilt laufend, ob sich die Szene verändert oder ob sie konstant bleibt. Wenn eine Dynamik in der Szene entsteht (wenn z.B. eine Person die Szene betritt), wartet das System, bis die Szenendynamik wieder konstant wird. Das System wird nicht früher aktiv, damit keine Gefahren z.B. für Personen, die sich in der Szene bewegen, entstehen können.

Anschließend versucht das System ein Objekt zu detektieren, welches innerhalb der dynamischen Phase der Szene hinzugefügt sein könnte. Falls kein Objekt gefunden wird, beobachtet das System die Szene weiter. Wenn ein neues Objekt hoch

Priorität

niedrig

<table>
<tr><td colspan="4">Sicherheitskriterium Gesamtsystem</td></tr>
<tr><td>Negativ</td><td colspan="3">Positiv</td></tr>
<tr><td rowspan="3">Keine Aktionen durchführen</td><td colspan="3">Kalibration Roboter-Kamera
Gütemaß K</td></tr>
<tr><td>$K << 0$</td><td>$0 < K$</td><td>$0 << K$</td></tr>
<tr><td>*Extra* Kalibration Roboter-Kamera durchführen bis $0 < K$</td><td>Kalibration Roboter-Kamera während anderer Aktionen durchführen</td><td>Anfangsverhalten Szene beobachten Auf Veränderungen reagieren</td></tr>
</table>

Tabelle 1: Prioritäten des Verhaltens

vermutet wird, wird die 3D-Position des Objekts berechnet und autonom eine rudimentäre Objektrepräsentation (OR) erstellt (siehe Abschnitt 4.2.2). Als nächstes wird das Objekt fixiert. Der Verlauf dieser Aufmerksamkeitssteuerung und der daraus resultierenden Handlung ist in der Abbildung 6 dargestellt.

Bei der Objektfixierung erhält das System nach einer Wiedererkennung des Objekts mit einer vorher erstellten Objektrepräsentation ein Gütemaß über die Qualität der durchgeführten Aktion. In Abhängigkeit der Güte führt das System eine genauere Fixierung durch. In jedem Schritt wird die 3D-Position des Objekts neu bestimmt. Hierfür wird die vorhandene Objektrepräsentation verwendet. Wenn das System die Entscheidung trifft, daß die Fixierung gut genug ist, bewegt es den Greifer zu den 3D-Objekt-Koordinaten.

Da das Systemmodell die verwendete Hardware nicht vollständig nachbildet, sind Fehlerquellen vorhanden. Die Transformation Kamerakoordinatensystem und Roboterkoordinatensystem ist nur eine Approximation. Somit wird der Greifer in der Regel nur in die Nähe des Objekts geführt. Die weitere Annäherung erfolgt innerhalb eines Wahrnehmungs-Handlungs-Zykluses.

Als nächstes erkennt das visuelle System den Greifer und berechnet aus diesen visuellen Daten eine 3D-Position. Aus dieser Position und der Objektposition wird der relative Abstand ermittelt. Dieser Wert ist das Gütemaß für die nächste Aktion des Greifers. In Abhängigkeit von diesem Gütewert wird der Greifer um den relativen Abstand zum Objekt geführt. Der Greifer wird also visuell kontrolliert zum Objekt angenähert. Dieser Verlauf ist in der Abbildung 7 dargestellt.

In weiteren Handlungsmodulen könnte das Objekt zum einen mit dem Zoom näher betrachtet und untersucht werden und zum anderen mit taktilen Sensoren

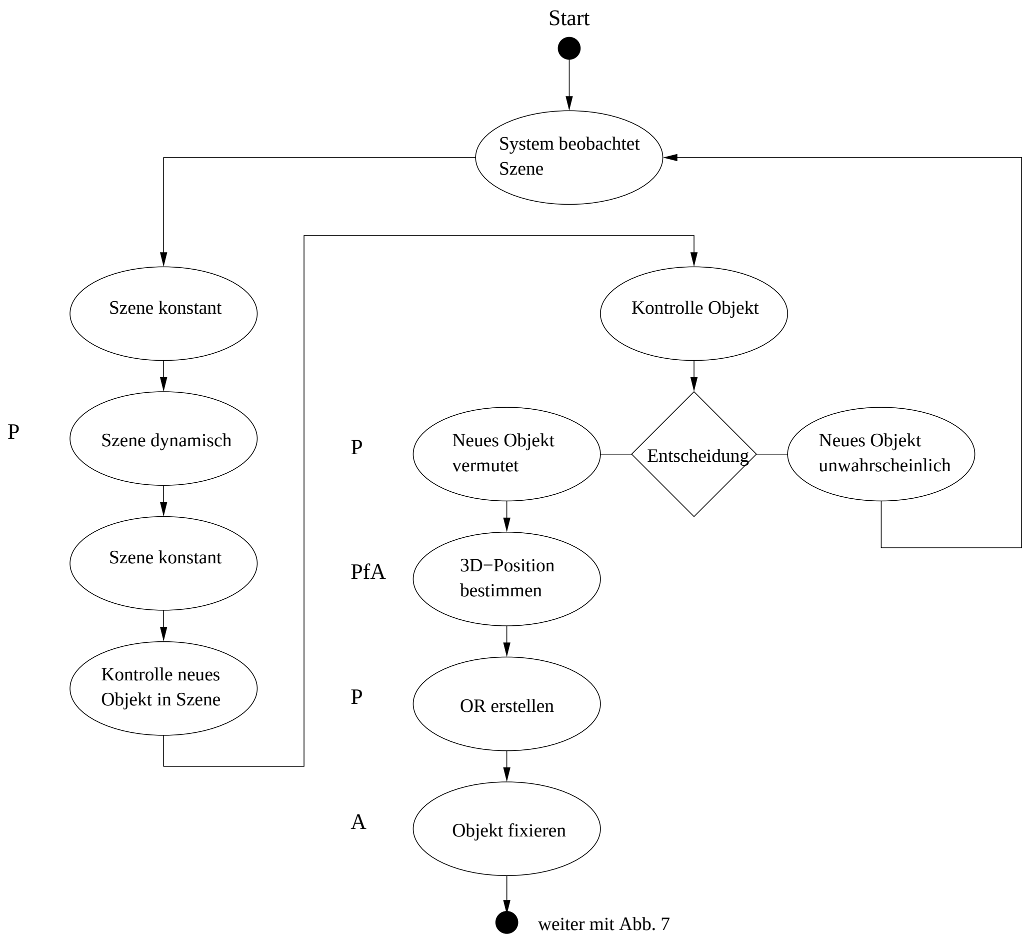


Abbildung 6: Beobachtung der Szene; Entscheidung, ob in der veränderten Szene ein Objekt vermutet wird. Legende: AfP Action for Perception, PfA Perception for Action, PAC Perception Action Cycle

(siehe Abb. 8) am Greifer taktil exploriert werden. Mit den taktilen Sensoren könnte auch der taktile Kontakt vom Greifer zum Objekt überprüft werden.

Dieses Vorgehen des *Ansehens* und *Anfassens* entspricht dem analogen Vorgehen von Babies, wenn sie Objekte explorieren wollen (siehe Abb. 9). Durch diese neuen Informationen kann eine weitere Objektrepräsentation erstellt werden. Aufgrund dieser neuen Objektrepräsentationen könnte das System eine Greifstrategie entwickeln. Beim Ansetzen des Griffes könnte das Greifen durch neue haptische Objektinformationen weiter optimiert werden. Nachdem das Objekt nun erfolgreich gegriffen werden könnte, würde das System die physikalische Kontrolle über das Objekt erhalten. Nun wäre eine neue Basis geschaffen worden, um weitere Informationen über das Objekt innerhalb eines Wahrnehmungs-Handlungs-

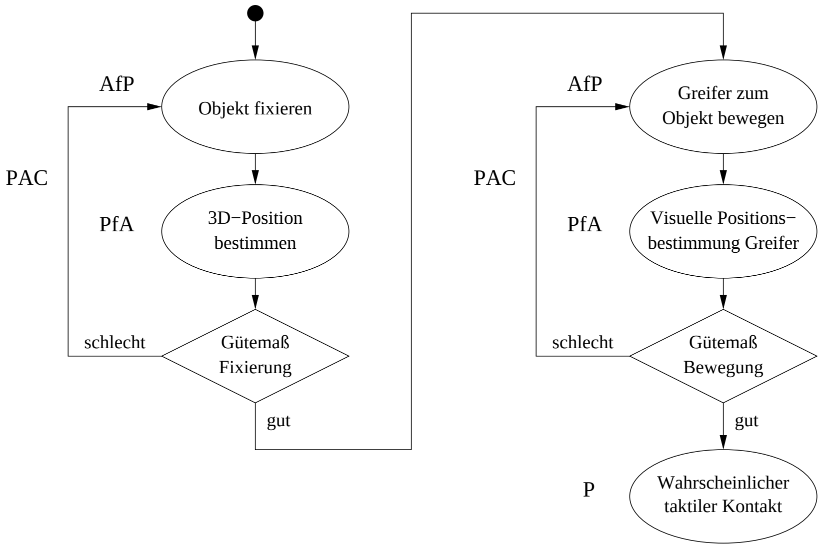


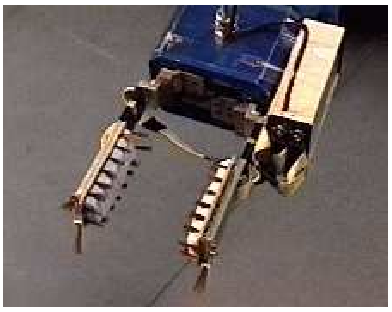
Abbildung 7: Fixierung des Objektes mit Nachfixierung gemäß einem Gütemaß; Annäherung Greifer zum Objekt. Legende: AfP Action for Perception, PfA Perception for Action, PAC Perception Action Cycle

Zykluses zu erlangen. Es kann z.B. eine Objektrepräsentation aufakkumuliert werden [1].

2.1.2 Gegenseitige Unterstützung von Wahrnehmung und Handlung

Zur Beschreibung des Zusammenspiels von Wahrnehmung und Handlung in der handlungbasierten Wahrnehmung erscheint der Wahrnehmungs-Handlungs-Zyklus als ein schlüssiges und notwendiges Instrument (z.B. [19, 34]). Im Gegensatz zum klassischen Ansatz, nach dem die Handlungsplanung erst nach einer vollständigen Elaboration einer Szene beginnt, werden nach dem Konzept des Wahrnehmungs-Handlungs-Zykluses die mit den Sensoren aufgenommenen Signale kontinuierlich verarbeitet und an die aktorischen Zentren weitergeleitet. Sobald eine bestimmte Schwelle überschritten ist, bildet sich ein Verhaltensmuster aus und der Wahrnehmungs-Handlungs-Zyklus setzt sich fort [2, 14]. Für bestimmte Handlungen sind nur einige bestimmte Wahrnehmungsinformationen notwendig. Somit sollten nur die notwendigen Informationen extrahiert und verarbeitet werden (Wahrnehmung unterstützt eine konkrete Handlung). Auf der anderen Seite erleichtern wahrnehmungsorientierte Handlungen die Extraktion der notwendigen Informationen (Handlung unterstützt Wahrnehmung).

Bei einer handlungsbasierten Wahrnehmung ergeben sich mehrere zu klärende Aufgaben. Für welche Handlung muß ich was wahrnehmen? Welche Handlung



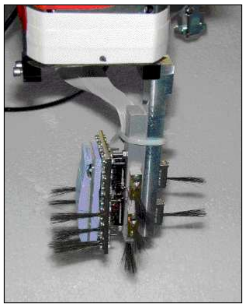


Abbildung 8: Zwei taktile Sensoren. Der rechte Sensor wird gerade am Lehrstuhl erprobt.

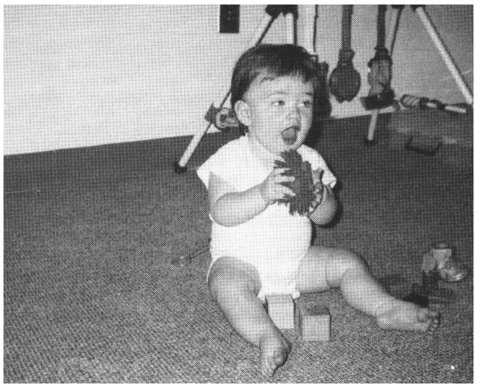


Abbildung 9: Ein 12 Monate altes Kind exploriert ein neues Objekt visuell und manuell [17].

sollte das System durchführen, um ein bestimmtes Ziel besser wahrnehmen zu können? Kann das System die Handlung mittels der Wahrnehmung kontrollieren? Wie gut war die Handlung und wie gut ist die Wahrnehmung? Durch einen Gütewert von der Handlung und Wahrnehmung besteht die Möglichkeit einer Rückkopplung für das System, wodurch der gesamte Handlungsablauf optimiert werden kann.

Bei einem Gütewert kann die nächste Aktion in Abhängigkeit von dem beurteilten Maß durchgeführt werden. Bei einem Ja/Nein–Kriterium wären weitere Unterscheidungen nicht möglich. Bei einem Ja/Nein Kriterium kann z.B. entschieden werden, ob eine Aktion durchgeführt wird oder nicht. Wenn ein Gütewert bekannt ist, kann eine Aktion schneller oder langsamer durchgeführt werden. Dies ermöglicht dem System eine größere Handlungsvielfalt. Somit sind Gütewerte Ja/Nein–Kriterien vorzuziehen. Ein Ja/Nein–Kriterium wird zum Beispiel bei der Irissteuerung verwendet (siehe unten). Bei der Roboter–Kamera–Kalibration basieren die weiteren Kalibrationsaktionen auf einen jeweils vorher berechneten Gütewert.

Die Interaktion zwischen Wahrnehmung und Handlung ist ein grundlegendes Konzept, welches in dieser Arbeit angewandt wird. Das Konzept wird auf verschiedenen Ebenen verwendet. Zum Beispiel gehört die Fähigkeit der Iriseinstellung zu den Basiskompetenzen. Um die Helligkeitsverteilung im Bild zu optimieren, wird die Iris in Abhängigkeit von der Bildgüte geregelt. Dies passiert im Wahrnehmungs-Handlungs-Zyklus (PAC) [38]. Das Fixieren von Objekten gehört zu den zusammengesetzten Kompetenzen. Das System kennt eine Positionsschätzung und besitzt eine rudimentäre Objektrepräsentation vom zu fixierenden Objekt. Das System fixiert das Objekt und sucht in den neuen Bilddaten das Objekt, um die Objektposition zu bestimmen. Hieraus kann ein Gütewert berechnet werden. In Abhängigkeit vom Gütewert kann eine weitere Fixierung durchgeführt werden, bis die Qualität der Fixierung ausreichend ist. Das Fixieren eines Objektes erfolgt also im Wahrnehmungs-Handlungs-Zyklus.

In den Abbildungen 6 und 7 sind die verschiedenen Abläufe von Handlungen skizziert. Dort sind auch die verwendeten Kompetenzen klassifiziert, ob sie in die Kategorie Wahrnehmung für Handlung (PfA), Handlung für Wahrnehmung (AfP) oder Wahrnehmungs-Handlungs-Zyklus (PAC) gehören.

2.2 Einordnung der Arbeit in einem möglichen komplexen System

Mehrere Studenten und Mitarbeiter vom Lehrstuhl 'Kognitive Systeme' arbeiten und entwickeln Teile eines visuell-aktorischen Systems. Ein mögliches Konzept eines solchen Gesamtsystems wird in [22] näher ausgeführt.

Dieses Handlungsmodul leitet mittels einfacher und reflektorischer Verhaltensmuster eine Handlung ein, so daß ein Roboter die physische Kontrolle über ein Objekt erlangen kann. Die Kontrolle über das Objekt kann verwendet werden, um eine Objektrepräsentation vom Objekt auf zu akkumulieren [1]. Diese Repräsentationen könnten verwendet werden für Tracken, Greifen und Wiedererkennen in komplexen Szenen. Für das Tracken ist die Handlung erfolgreich durchgeführt worden [29].

Ein wichtiger Aspekt bei der Entwicklung eines komplexen, verhaltensbasierten visuell-aktorischen Systems ist die Interaktion der verschiedenen Module, die von verschiedenen Personen entwickelt werden oder wurden. Die Interaktion wird mittels eines gemeinsamen Softwarepakets ermöglicht. Komplexere Kompetenzen entstehen durch die Kombination mehrerer einfacher Kompetenzen. Das verwendete Softwarepaket ist eine C++–Bibliothek (**KiViGraP**, **Ki**eler **Vi**sion and **Gra**sping **P**roject), welches fortlaufend weiterentwickelt wird und welches die Interaktion der Module ermöglicht (weitere Details siehe [18]).

2.3 Einordnung in die Wahrnehmungs- und Handlungsproblematik

Aus Wahrnehmungen und Handlungen resultieren viele Probleme, die zum Teil mit unterschiedlichen Methoden gelöst werden können. Diese Probleme sind sowohl bei biologischen als auch künstlichen Systemen vorhanden. In Bezug auf den Berührungsvorgang werden im nächsten Abschnitt Studien über Babies gezeigt, in denen deren Probleme und die daraus resultierende Lösungsstrategie, die Kopplung von Wahrnehmung und Handlung aufgeführt werden. Im weiteren wird aufgeführt, welche Lösungsstrategie sich daraus für künstliche Systeme ergeben. Hierbei wird im besonderen auf die Kalibration im Wahrnehmungs-Handlungs-Zyklus und der aufgabenbezogene Wandel von rudimentären zu komplexen Objektrepräsentationen eingegangen.

2.3.1 Annahmen zur Kopplung von Sehen und Greifen bei Säuglingen

Im folgenden Abschnitt werden einige Studien von Kellman [17] bezüglich der Wahrnehmung von Babies und Kleinkindern vorgestellt: Nach Kellman hängt die Wahrnehmung von Babies und Kleinkindern mit einigen primitiven Aktionen zusammen. Es wurde festgestellt, daß die Fähigkeit zur Wahrnehmung von Babies besser entwickelt ist, als ihre Durchführungssicherheit von motorischen Handlungen.

Ein wichtiger Schritt in der Wahrnehmungs-Handlungs-Koordination ist die Entwicklung des Greifens mit einem Alter von 4 – 5 Monaten. Das erfolgreiche Greifen von einem Objekt benötigt die Wahrnehmung zur Objekt-Lokalisation und einer approximativen Generierung von Bewegungen zum Ort. Im Alter von 0 – 4 Monaten verwenden die Babies eine Vorstufe von der mit der Zeit entwickelten Greifhandlung. Diese Vorstufe ist weniger genau als das Greifen bei älteren Babies, welches visuell gesteuert wird. Bei der Vorstufe des Greifens wird die *geführte* Hand zum gesehenen Ziel gebracht. Bei dem visuell geführten Greifen wird die gesehene Hand mit dem gesehenen Ziel in Verbindung gebracht.

Neugeborene haben bei der Berührung von Objekten eine Trefferrate von 9%. 17-Wochen alte Babies treffen das Objekt mit einer Rate von 93%. Mit der Zeit erlernen die Babies somit eine bessere Greifstrategie. Ältere Kleinkinder und Erwachsene verwenden beim Greifen zwei typische Bewegungsphasen. In der ersten Phase ist die Bewegung groß. Die Hand nähert sich dem Objekt. In der zweiten Phase werden kleinere Korrekturbewegungen durchgeführt. Die Strategie der Unterteilung der Bewegung in diese zwei Bewegungsschritte begründet sich darin, daß dem Menschen vor der Handlung die Objektposition nur ungefähr bekannt ist. Die absolute Tiefenwahrnehmung des menschlichen visuellen Systems ist wesentlich schlechter als die relative Tiefenwahrnehmung [24]. In der zweiten Bewegungphase basiert die Bewegung unter anderem auf relativen Entfernungsschätzungen, die aus den visuellen Daten gewonnen werden können, und der Rückkopplung zwischen Zielobjekt und Hand.

2.3.2 Primitive Handlungen in einem künstlichen System als Einleitung von komplexen Prozessen

Gemäß dem vorherigen Abschnitt sind Babies mit Grundfähigkeiten ausgestattet, um ihre Umwelt zu explorieren. Diese werden verfeinert bzw. wandeln sich zu komplexen Fähigkeiten. Zum Beispiel hat die Bewegung mittels einer absoluten 3D–Schätzung schon einen gewissen Erfolg, welcher aber durch die visuelle Handführung (also der Kopplung der Handbewegung und des visuellen Apparats) deutlich verbessert werden kann.

Für ein künstliches System stellt sich dementsprechend die Frage, ob das System sich Voraussetzungen für die gewünschte komplexe Tätigkeit verschaffen kann, so daß diese Tätigkeit, wie zum Beispiel das Aufakkumulieren von Objektrepräsentationen eines Objektes, besser ausgeführt werden können. Da das System dieser Arbeit viele aktive Komponenten besitzt, ist es potentiell in der Lage, sich bessere Voraussetztungen für bestimmte Handlungen zu verschaffen. Dem System sollten analog zum Baby primitive Handlungen und Kompetenzen von Anfang an zur Verfügung stehen. Um den Prozess zügig einzuleiten, ist es von Vorteil, wenn diese primitiven Kompetenzen durch bestimmte Wahrnehmungen und Handlungserfolge eindeutig ausgelöst werden. Wenn also das System ein neues Objekt *sieht*, fixiert das System das Objekt. Wenn die Fixierung erfolgreich war, dann bewegt das System den Greifer zum Objekt usw. Der nächsten Abschnitt behandelt das Thema Handlungserfolg und die daraus mögliche Rückkopplung zur Systemoptimierung.

2.3.3 Rückkopplung zur Systemoptimierung

Ein visuell-aktorisches System kann in der Regel kein abgeschlossenes System sein, da sich z.B. interne Parameter ändern können. Es ist also von einer permanenten Rückkopplung über seinen Erfolg abhängig und sollte daher die Möglichkeit haben, sich in jeder Situation neu anpassen zu können (Adaption). Insofern ist es wünschenswert, für möglichst viele Handlungen ein Maß für den Erfolg mit zu evaluieren. In dem System, welches in dieser Arbeit entwickelt wurde, sind für verschiedene Handlungen Gütemaße vorhanden , wie zum Beispiel bei der

Objektfixierung oder die Greiferposition nach visuellen und aktorischen Informationen.

2.3.4 Kalibration im PAC

Eine wichtige Systemvoraussetzung ist die Kalibration des Roboter-Kamera-Koordinatensystems. Beide Teilsysteme sind nicht miteinander physikalisch verbunden. Dem System stehen nur wenige Informationen bezüglich Translation und Rotation der beiden Objekte als Grundwissen zur Verfügung. Damit beide Teilsysteme miteinander interagieren können und somit eine Einheit bilden, müssen dem System diese Informationen relativ genau bekannt sein. Hierbei ist es wieder von Vorteil, wenn das System dies anhand eines Gütemaßes selbst überprüfen kann. Falls das Gütemaß schlecht ist, sollte das System in der Lage sein, dieses Wissen durch geeignete Handlungen zu erlangen.

Das in dieser Arbeit beschriebene System kann diese Informationen bis zu einer bestimmten Qualität innerhalb eines Wahrnehmungs-Handlungs-Zykluses extrahieren und optimieren. Das System hat als Grundwissen 2D-Objektrepräsentationen von verschiedenen Posen des Greifers. Wenn sich der Roboterarm zu verschiedenen Positionen bewegt, kann das System jeweils den Greifer wiedererkennen. Aus dem Wissen von der Greiferposition im Roboterkoordinatensystem und der visuellen Positionsbestimmung kann eine Liste von Korrespondenzen erstellt werden. Aus dieser Liste kann eine erste Schätzung der Transformationen der beiden Koordinatensysteme berechnet werden. Im weiteren werden neue Korrespondenzlisten erstellt, die in Abhängigkeit von der bisherigen Güte in das Modell einfließen.

2.3.5 Von rudimentärer Objektrepräsentation zur komplexen Objektrepräsentation

Da die Art der Wahrnehmung und der Handlung im Verlauf des Wahrnehmungs-Handlungs-Zykluses komplexer werden kann, bedeutet dies, daß die Art der Repräsentationen vom Objekt sich mit der Zeit verändern kann. In die neue Objektrepräsentation können die bisherigen Erfahrungen und aufgesammelten Informationen einfließen (siehe Abb. 10). Das Konzept, daß sich die Art der Objektrepräsentation mit der Zeit und mit der Erfahrung verändern, wurde auch bei Kleinkindern festgestellt [21]. Kleinkinder, jünger als sechs Monate, nehmen Objekte als *etwas mit einer bestimmten Position und einer bestimmten Geschwindigkeit* wahr. Objekte besitzen kein *darüber*, *darunter*, *links*, *rechts*, *davor* oder *dahinter*. Kleinkinder über sechs Monate beginnen Objekte durch Form und Farbe zu repräsentieren [3]. Ab diesem Alter fangen sie auch an zu verstehen, daß

Objekte weiterexistieren, wenn diese verdeckt werden (Objektpermanenz) [26].

In dieser Arbeit wird zum Beispiel am Anfang ein Objekt mittels einem Grauwertdifferenzbild abgebildet, welches einer nur sehr rudimentären Repräsentation entspricht. In einer anderen Stufe des PAC wird ein Objekt als ein Gitternetzwerk von Jets von Gaborwavelets verwendet [23]. Zu einem späteren Zeitpunkt kann das gleiche Objekt mittels 3D-Liniensegmente repräsentiert werden [1]. Der Wandel der Objektrepräsentationen wird in der Abbildung 10 dargestellt. Wichtig hierbei ist, daß das System in der Lage ist, sich durch autonome Interaktionen mit seiner Umgebung, ausgehend von rudimentären Objektrepräsentationen, komplexere und abstraktere Objektrepräsentationen anzueignen. Der Wandel der Art der Objektrepräsentation ist in der Abbildung 10 dargestellt.

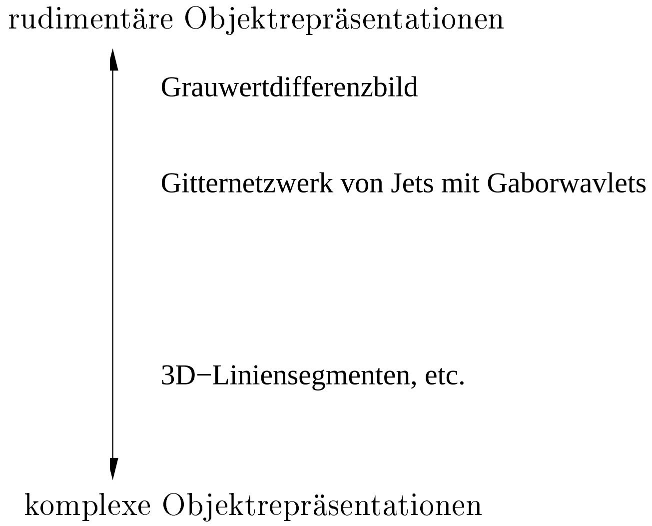


Abbildung 10: Verlauf der Komplexität von Objektrepräsentationen

3 Modell eines binokularen Kamerasystems

Die visuellen Kompetenzen sind wichtige Bestandteile des entwickelten visuell-aktorischen Systems. Für diese visuellen Kompetenzen ist ein geeignetes Modell von einem aktiven Kamerakopf notwendig. Binokulare Kamerakopfsysteme sind in ihrer Gestaltung an die Anatomie des menschlichen visuellen Systems angelehnt [10].

Das Modell muß mehrere Grundkompetenzen besitzen, um die komplexeren Kompetenzen, wie zum Beispiel die Extraktion von 3D-Informationen aus Stereobildern, zu ermöglichen. Zu den Grundkompetenzen gehören unter anderem der Bezug des Kamerasystems zum Weltkoordinatensystem, die Abbildungsparameter und deren Neuberechnung nach einer Eigenbewegung. Dies liefert die Grundlagen für die Tiefenschätzung.

In Abschnitt 3.1 wird der Aufbau des Binokularkopfes beschrieben. Es wird ein Bezug vom Kamerakoordinatensystem zum Weltkoordinatensytem hergestellt. Dies erfolgt einerseits durch die geometrische Beschreibung des speziellen Kamerasystems und andererseits durch die analytische Beschreibung der aktiven Komponenten.

In Abschnitt 3.2 wird eine Einführung in die Abbildungsparameter von Kameras gegeben. Weiterhin wird auf ein Standardverfahren zur Bestimmung der Abbildungsparameter eingegangen. Die Abbildungsparameter können in die intrinsischen und extrinsischen Parametern aufgeteilt werden [11]. Die intrinsischen Parameter wurden in dieser Arbeit für einen speziellen Stereokamerakopf (TRC-Kopf) durch eine Statistik bestimmt. Die externen Parameter werden über das Wissen über den Aufbau des Kamerakopfes und den Zustand seiner aktiven Komponenten berechnet. Mit der Kenntnis der Abbildungsparameter der Kameras können metrische Messungen mittels Stereobilder durchgeführt werden [11].

3.1 Aufbau des Binokularkopfes

Das in dieser Arbeit verwendete aktive, binokulare Kamerasystem besteht aus mehreren Bauteilen (siehe Abb. 11) [6, 10]. Die Bauteile sind die Plattformbasis, der Hals, ein Verbindungsstück zwischen den beiden Kameras und die zwei Kameras. Die Basis hat keine Freiheitsgrade. Der Hals ist auf der Basis montiert und kann um eine Achse gedreht werden. Diese Achse wird als Schwenkachse (pan axis) χ_H bezeichnet (y-Achse). Ein Verbindungsstück ist auf dem Hals befestigt und kann um die x-Achse gedreht werden. Es hat die Länge b.

Die technischen Daten und die Grenzwerte der aktiven Komponenten des TRC-Kopfes sind in [6, 38] beschrieben. Beim TRC-Kopf ist $b = 25cm$. Diese Achse wird als Neigungsachse (tilt axis) ϕ_H bezeichnet. Die Schwenk- und Neigungsachse besitzen einen Schnittpunkt.

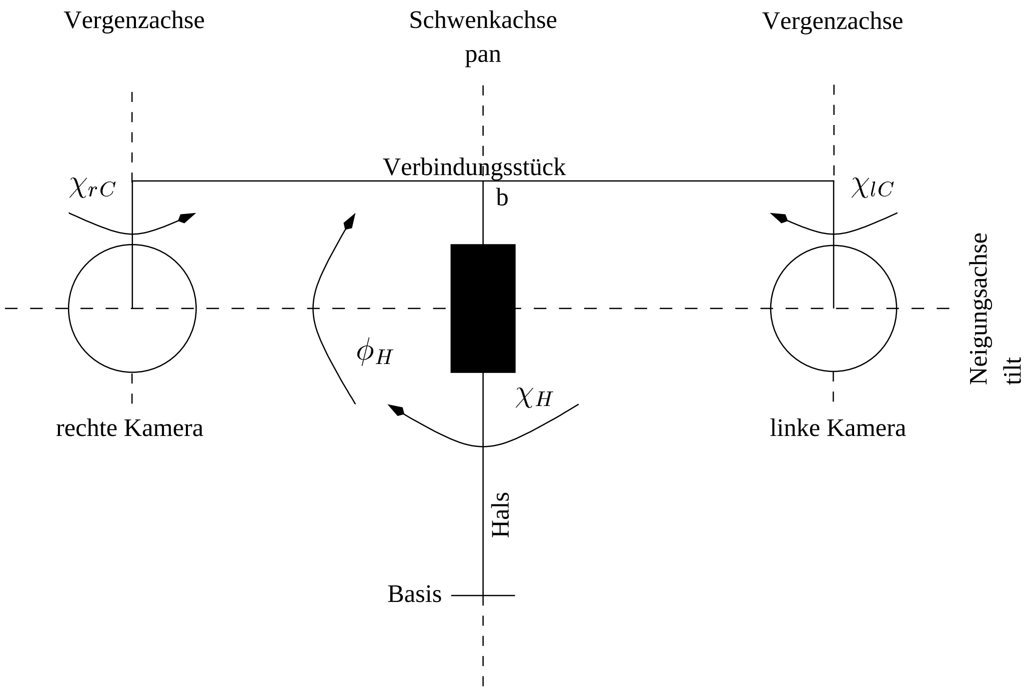


Abbildung 11: Grafik von den möglichen mechanischen Freiheitsgraden [6], rechte Kamera Pan-Winkel χ_{rC}, linke Kamera Pan-Winkel χ_{lC}, Kamerakopf Pan-Winkel χ_H, Kamerakopf Tilt-Winkel ϕ_H

Beide Kameras sind am Verbindungsstück montiert und besitzen je einen Freiheitsgrad. Sie können um die Y-Achse (vergenz) $\chi_{l/rC}$ gedreht werden. Die Vergenzachsen schneiden sich auf der Neigungsachse des Kamerakopfes. Dieser Schnittpunkt ist auch das optische Zentrum der jeweiligen Kamera [15]. Insgesamt hat also das Modell des Kamerasystems vier mechanische Freiheitsgrade. Die beiden Kameras haben zudem jeweils als optische Freiheitsgrade die Parameter Fokus (Entfernungseinstellung), Zoom (Brennweite) und Iris (Blendenöffnung). Objektive mit veränderbarer Brennweite werden als Varioobjektive bezeichnet. In [25] werden Varioobjektive näher erläutert. Der prinzipielle Aufbau eines Varioobjektives ist in Abbildung 12 dargestellt. Die Brennweitenänderung geschieht durch axiales Verschieben von Linsengruppen in einem optischen System. Prinzipiell besteht ein Varioobjektiv aus einem feststehenden Grundobjektiv G und dem Variator V. Der Variator bewirkt im wesentlichen die Brennweitenänderung. Damit die Bildlage erhalten bleibt, ist eine weitere Linsengruppe K (Kompensator) erforderlich. Weitere vor dem Variator angeordnete

Frontlinsengruppen F ermöglichen unter anderem die Entfernungseinstellung zum betrachteten Objekt.

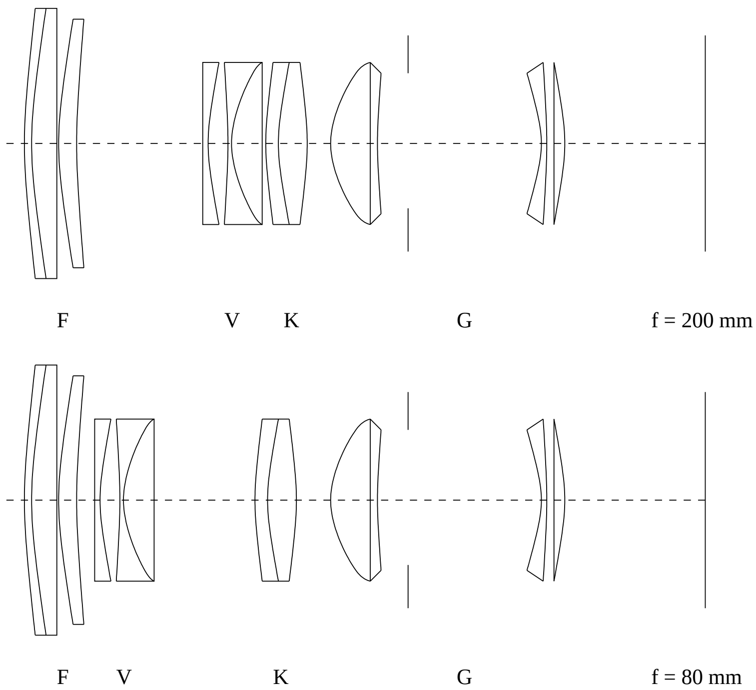


Abbildung 12: Prinzipieller Aufbau eines Varioobjektivs $f = 80 - 200$ mm mit einem Grundobjektiv G, einem Variator V, einem Kompensator K und einem Frontglied F zur Entfernungseinstellung (Grafik aus [25])

3.1.1 Koordinatensysteme

Um das Kamerasystem mit Koordinatensystemen genau zu beschreiben, sind verschiedene Koordinatensysteme notwendig. In diesem Abschnitt werden nur die beiden wichtigen Systeme in Kamera- und Weltkoordinaten und deren Transformation ineinander erläutert. Diese stimmen mit den Definitionen von [10] überein. In dieser Diplomarbeit werden auch weitere Koordinatensysteme beschrieben. Alle verwendeten Koordinatensysteme sind kartesisch. Der Ursprung des Weltkoordinatensystems $\mathbf{O}$ ist der Schnittpunkt der Schwenk- und Neigungsachse des Kamerakopfes. Das optische Zentrum der linken/rechten Kamera ist der Ursprung des linken/rechten Kamerakoordinatensystemes $\mathbf{K}_{l/r}$. Das Kamerakoordinatensystem wird mit der Rotation $\mathbf{R}_{l/r}$ und der Translation $\mathbf{t}_{l/r}$ in das Weltkoordinatensystem überführt.

$$\mathbf{O} = \mathbf{R}\mathbf{K} + \mathbf{t} \quad \text{bzw.} \quad \mathbf{K} = \mathbf{R}^{-1}(\mathbf{O} - \mathbf{t}) \tag{3.1}$$

Da die beiden Kameras beim TRC-Kopf die gleiche Entfernung auf der X-Achse zum Schnittpunkt der X-Achse mit der Y-Achse haben, ist der Ursprung des Weltkoordinatensystems somit auch der Mittelpunkt der beiden optischen Zentren. Die optische Achse der beiden Kameras ist in der Anfangsposition (d.h. nach der Initialisierung des Kamerakopfes) parallel zur Z-Achse. Für den Vektor $\mathbf{t}_{l/r}$ kann angenommen werden: $\mathbf{t}_{l/r} = \left(\mp\frac{b}{2}, 0, 0\right)^T$

Die Ausrichtung des Weltkoordinatensystems ist in Abbildung 13 dargestellt.

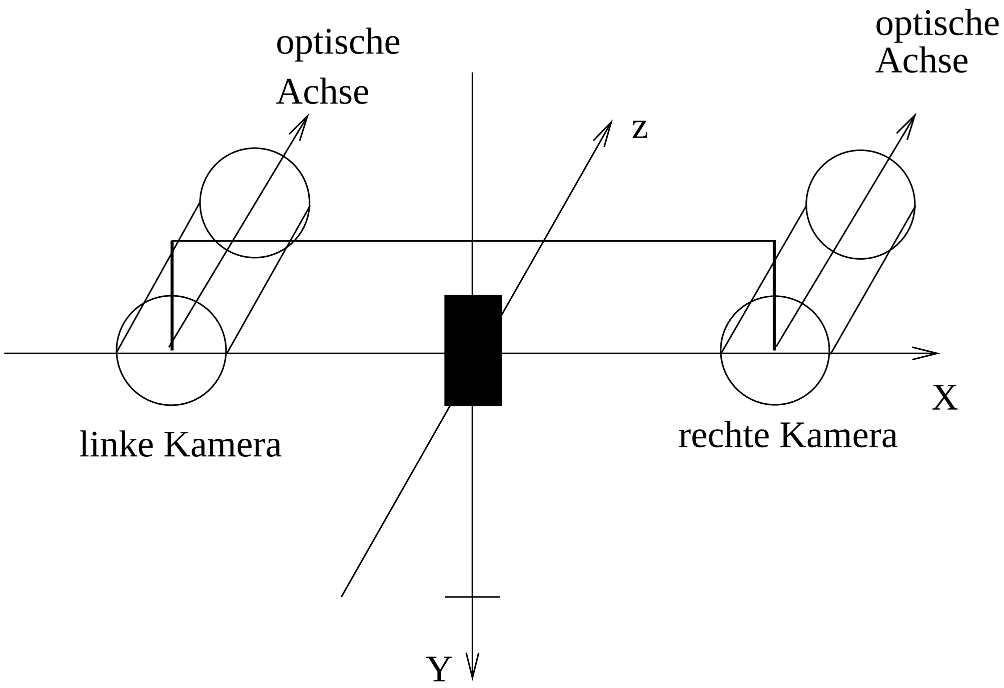


Abbildung 13: Die optischen Achsen der Kameras sind in der Anfangsposition parallel zur Z-Achse. Die Pfeilrichtung der Achsen markieren die positiven Zahlen der jeweiligen Achse.

3.2 Die Abbildungsparameter

Eine dreidimensionale Szene wird durch eine Kamera auf ein zweidimensionales Bild abgebildet. Jeder Bildpunkt repräsentiert einen Sehstrahl in der dreidimensionalen Szene. Der Sehstrahl führt von einem 3D-Szenenpunkt durch das optische Zentrum der Kamera zum Bildpunkt. Die Beziehung zwischen einem Bildpunkt und der entsprechenden Geraden kann durch eine Projektionsmatrix angegeben werden. Die Projektionsmatrix hängt von den intrinsischen Parametern (Aufbau der Kamera) und den extrinsischen Parametern (Orientierung und Position der Kamera zu einem Weltkoordinatensystem) ab.

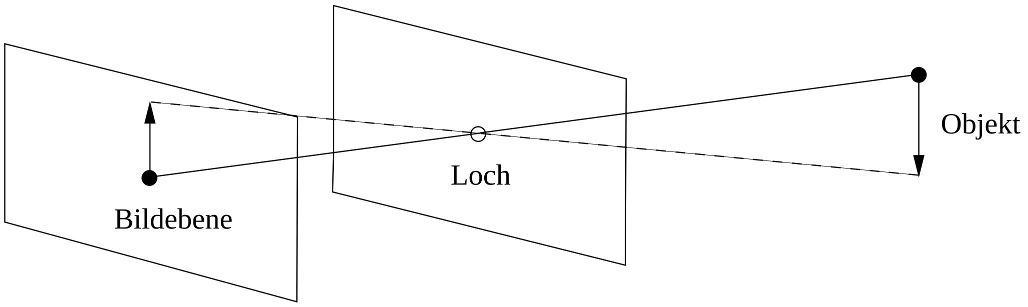


Abbildung 14: Das Lochkameramodell.

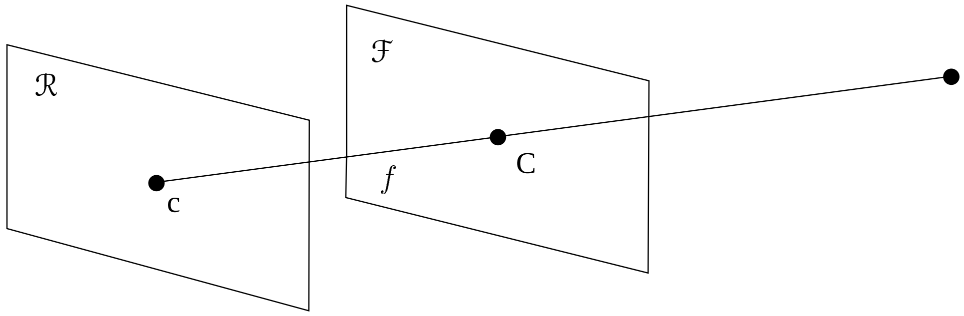


Abbildung 15: Das optisches Zentrum C, die Bildebene $\mathcal{R}$, die fokalen Ebene $\mathcal{F}$ und die Brennweite f

Im folgenden werden die theoretischen Grundlagen für die verschiedenen Parameter geliefert [11].

Definition der intrinsischen und extrinsischen Parameter

Die Geometrie des Strahlenganges einer Kamera kann durch das Lochkameramodell (siehe Abb. 14) angenähert werden. In der Abbildung 15 werden mehrere optische Größen (das optisches Zentrum C, die Bildebene $\mathcal{R}$, die fokalen Ebene $\mathcal{F}$ und die Brennweite f) dargestellt. Die Fokalebene $\mathcal{F}$ hat das optische Zentrum C. Das optische Zentrum C wird auf den Punkt c auf der Bildebene $\mathcal{R}$ abgebildet. Dieser Strahl beschreibt die optische Achse. Die Bildebene hat als Abstand zur Fokalebene die Brennweite f.

Die vier intrinsischen Parameter k_u, k_v, u_0 und v_0 (siehe Abb. 16) beschreiben den inneren Aufbau der Kamera. Der Vektor $(u_0, v_0)^T$ ist die Translation in Pixelkoordinaten vom Schnittpunkt der optischen Achse mit der Bildebenen zum Bildursprung. Die Abmessung eines Pixels in Meter in der horizontalen und vertikalen Richtung wird durch k_u^{-1} und k_v^{-1} angegeben. Mittels der Brennweite f der Kamera wird der Wert Brennweite in Pixel $\alpha_u = -fk_u$ und $\alpha_v = -fk_v$ berechnet. Diese Parameter sind unabhängig von der Position und Orientierung

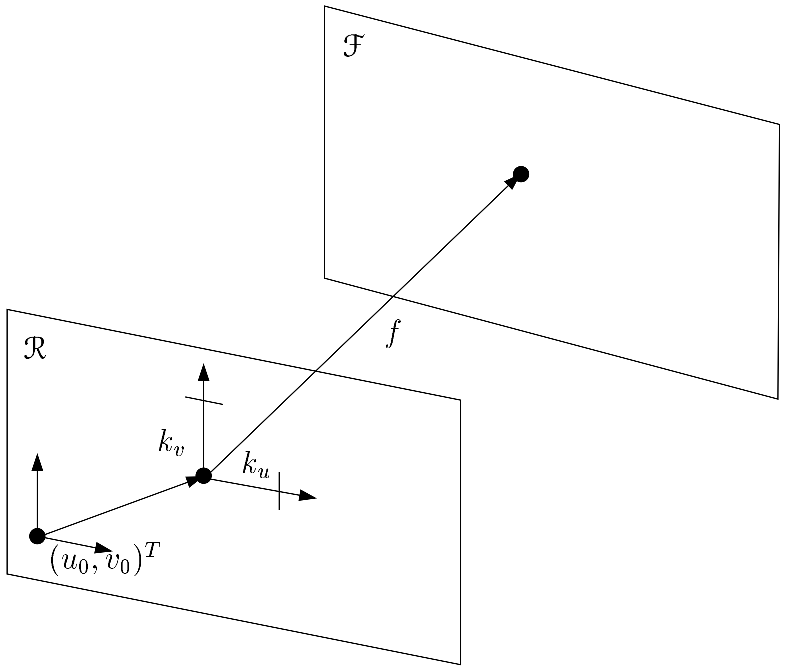


Abbildung 16: Die intrinsischen Kameraparameter

der Kamera. Mit ihrer Kenntnis können metrische Messungen durchgeführt werden, zum Beispiel die Winkelberechnung zwischen zwei Lichtstrahlen, die jeweils einen Pixel bestimmen.

Bild- und Weltpunkte können in homogenen Koordinaten, die im folgenden immer mit einer Tilde gekennzeichnet sind, beschrieben werden. Auf diese Weise können affine Abbildungen durch Matrizenoperationen dargestellt werden.

Die Position des optischen Zentrums und die Orientierung der Kamera im Weltkoordinatensystem können durch sechs extrinsischen Parameter bestimmt werden. Mit drei Parametern für die Rotationswinkel werden die neun Einträge der Rotationsmatrix $\mathbf{R}$ definiert. Weitere drei Parameter definieren den Translationsvektor $\mathbf{t}$. Durch die Verwendung von homogenen Koordinaten können beide Operationen in einer 4×4 Matrix $\widetilde{\mathbf{T}}$ zusammengefaßt werden (siehe Gleichung (3.2)). In einigen Arbeiten wird das Abbildungsverhalten noch genauer beschrieben. Dort werden z.B. Parameter zur Rotation der Bildebene eingeführt [11].

Für eine zusammenfassende Darstellung der intrinsischen und extrinsischen Parameter gibt Faugeras die folgende Projektionsmatrix $\widetilde{\mathbf{P}}$ an:

$$\widetilde{\mathbf{T}} = \begin{pmatrix} \mathbf{R} & \mathbf{t} \\ 0 & 1 \end{pmatrix} \tag{3.2}$$

$$\widetilde{\mathbf{P}} = \begin{pmatrix} \alpha_u & 0 & u_0 & 0 \\ 0 & \alpha_v & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \widetilde{\mathbf{T}} \tag{3.3}$$

Die Abbildung eines Weltpunktes $\widetilde{\mathbf{M}} = (X, Y, Z, 1)^T$ in einen Bildpunkt $\widetilde{\mathbf{m}} = (u, v, 1)^T$ wird durch die Gleichung (3.4) angeben:

$$\widetilde{\mathbf{m}} = \widetilde{\mathbf{P}}\widetilde{\mathbf{M}} \tag{3.4}$$

3.2.1 Bestimmung der Abbildungsparameter

Es gibt verschiedene Standardverfahren zum Bestimmen der intrinsischen und extrinsischen Parameter. Häufig wird ein Kalibrierungsmuster verwendet. Dies kann zum Beispiel ein Würfel (siehe Abb. 17) sein. Bei den Kalibrierungsmustern ist der jeweilige dreidimensionale Aufbau bekannt. Bei dem Würfel sind die Ausmaße des Würfels, die Abmessungen der schwarzen Quadrate und die Abstände zwischen den Quadraten bekannt [13]. In einem weiteren Schritt werden Korrespondenzen zwischen den Bildpunkten $\widetilde{\mathbf{m}}$ und den bekannten 3D-Objektpunkten $\widetilde{\mathbf{M}}$ erstellt. Die Projektion eines Objektpunktes auf einen Bildpunkt wird durch Gleichung (3.4) beschrieben.

Sind mindestens sechs korrespondierende Paare von Bild- und Weltpunkten bekannt, so können die Matrixeinträge der Projektionsmatrix $\widetilde{\mathbf{P}}$ aus der Gleichung (3.4) berechnet werden. In [11] wird dazu ein geeignetes Verfahren vorgeschlagen. In diesem Verfahren entsteht ein überbestimmtes, lineares Gleichungssystem. Dieses kann mit dem Householder-Verfahren [27] gelöst werden. Das Householder-Verfahren dient zum Finden einer Ausgleichslösung von überbestimmten, linearen Gleichungssystemen.

Aus der Projektionsmatrix können wiederum die intrinsischen und extrinsischen Parameter bestimmt werden.

Obiges Verfahren hat mehrere Nachteile. Man kann die aktiven Komponenten des Binokularkopfes nicht ohne weiteres verwenden. Durch eine Veränderung der mechanischen Freiheitsgraden würden sich die extrinsischen Parameter der Projektionsmatrix $\widetilde{\mathbf{P}}$ verändern. Nach einer Aktion des Kamerasystems müßte ein weiteres Mal kalibriert werden. Diese Kalibrierung ist zeitaufwendig und würde den momentanen Handlungsablauf des Kamerakopfes unterbrechen.

In dieser Arbeit werden die intrinsischen und extrinsischen Parameter unabhängig voneinander berechnet. Die intrinsischen Parameter wurden durch eine Statistik bestimmt. Die extrinsischen Parameter werden aus den bekannten Systemparametern des Binokularkopfes berechnet.

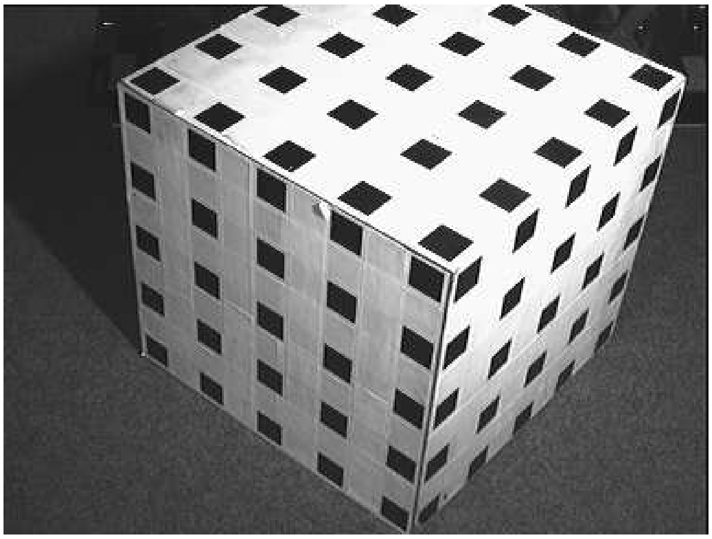


Abbildung 17: Kalibrationswürfel

3.2.2 Berechnung der extrinsischen Parameter

Durch das Anwenden der aktiven-motorischen Komponenten des Stereokamerakopfes verändern sich die externen Parameter. Die aktiven Komponenten drehen Systemeinheiten des Kamerakopfes um bekannte Achsen. Dem System sind die jeweiligen Drehwinkel bekannt. Durch das Wissen über den geometrischen Aufbau des Kamerasystems und die Drehwinkel können die extrinsischen Parameter als die entsprechende Rotationsmatrix $\widetilde{\mathbf{T}}_{l/r}$ berechnet werden [6]. Es gibt jeweils eine Matrix für die linke und für die rechte Kamera. Die Transformationsmatrix $\widetilde{\mathbf{T}}_{l/r}$ ist in homogenen Koordinaten angegeben. In ihr ist die Rotationsmatrix $\mathbf{R}_{l/r}$ und der Translationsvektor $\hat{\mathbf{t}}_{l/r}$ enthalten (3.10).

Die Gesamtrotation $\mathbf{R}_{l/r}$ kann als Verkettung von Einzelrotationen berechnet werden (3.9). Die einzelnen Rotationen sind die Rotation des Kopfes $\mathbf{R}_H$ (3.5), die Rotation der linken/rechten Kamera $\mathbf{R}_{l/rC}$ (3.6) und die Initialrotation der linken/rechten Kamera $\mathbf{C}_{l/r}$ (3.7). Die Initialrotation ist die Rotation des Kamerakoordinatensystems in das Weltkoordinatensystem nach der Initialisierung des Kamerakopfes.

Die Rotationsmatrix $\mathbf{R}_H$ ist abhängig von den Winkeln ϕ_H und χ_H. ϕ_H ist die Winkelangabe um die Neigungsachse des Kopfs. Der Winkel χ_H ist die Winkelangabe um die Schwenkachse des Kopfs. Die Belegung der Matrix $\mathbf{R}_H$ steht in der Gleichung (3.5). Sie ergibt sich aus der Verkettung von zwei Rotationen um zwei verschiedene Achsen (Neigungsachse und Schwenkachse).

$$\mathbf{R}_H = \begin{pmatrix} \cos(\chi_H) & 0 & \sin(\chi_H) \\ \sin(-\phi_H)\sin(\chi_H) & \cos(-\phi_H) & -\sin(-\phi_H)\cos(\chi_H) \\ -\cos(-\phi_H)\sin(\chi_H) & \sin(-\phi_H) & \cos(-\phi_H)\cos(\chi_H) \end{pmatrix} \tag{3.5}$$

Die Rotationsmatrix $\mathbf{R}_{l/rC}$ ist abhängig von der linken/rechten Kamera und von dem jeweiligen Drehwinkel $\chi_{l/rC}$. $\chi_{l/rC}$ ist die Winkelangabe um die Vergenzachse. Die Einträge von $\mathbf{R}_{l/rC}$ sind in der Gleichung (3.6) zu lesen.

$$\mathbf{R}_{l/rC} = \begin{pmatrix} \cos(\chi_{l/rC}) & 0 & \sin(\chi_{l/rC}) \\ 0 & 1 & 0 \\ -\sin(\chi_{l/rC}) & 0 & \cos(\chi_{l/rC}) \end{pmatrix} \tag{3.6}$$

Die Initialisierungsrotation ist davon abhängig, wie die einzelnen Kameras bei der Initialisierung zum Weltkoordinatensystem rotiert sind. Bei den durchgeführten Experimenten in dieser Arbeit wurden die Kameras immer mit dem Weltkoordinatensystem ausgerichtet, so daß für die Matrix $\mathbf{C}_{l/r}$ die Einheitsmatrix verwendet wurde (siehe Gleichung (3.7)). In Abschnitt 3.2.3 werden Fehlerquellen der extrinsischen Parameter beschrieben. Diese könnten durch eine Anpassung von $\mathbf{C}_{l/r}$ während der Aktionen des Systems verringert werden.

$$\mathbf{C}_{l/r} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{3.7}$$

Die Translation $\hat{\mathbf{t}}_{l/r}$ wird durch die Anwendung der inversen Rotation $\mathbf{R}_H$ auf den konstanten Translationsvektor $\mathbf{t}_{l/r}$ erhalten. Siehe Gleichung (3.8):

$$\hat{\mathbf{t}}_{l/r} = -\mathbf{R}_H^{-1}\mathbf{t}_{l/r} = -\mathbf{R}_H^{-1} \begin{pmatrix} \pm\frac{b}{2} \\ 0 \\ 0 \end{pmatrix} \tag{3.8}$$

Die Gesamtrotation $\mathbf{R}_{l/r}$ von den Kamerakoordinatensystemen wird durch die Gleichung (3.9) beschrieben.

$$\mathbf{R}_{l/r} = \mathbf{R}_{l/rC}\mathbf{C}_{l/r}\mathbf{R}_H \tag{3.9}$$

Die extrinsischen Parameter können durch die Matrix $\widetilde{\mathbf{T}}_{l/r}$ aus Gleichung (3.10) beschrieben werden.

$$\widetilde{\mathbf{T}}_{l/r} = \begin{pmatrix} \mathbf{R}_{l/r} & \hat{\mathbf{t}}_{l/r} \\ 0 & 1 \end{pmatrix} \tag{3.10}$$

Mit den Gleichungen ((3.5) - (3.10)) können die Transformationsmatrizen für die linke und rechte Kamera berechnet werden.

3.2.3 Fehlermöglichkeiten der extrinsischen Parameter

Zwei Fehlermöglichkeiten müssen in Betracht gezogen werden. Zum einen kann sich die Mechanik des Binokularkopfes durch mehrfache Bewegungen des Kamerakopfes verstellen. Dies hat zur Folge, daß der Kamerakopf fehlerhafte Winkelwerte dem System übermittelt. Somit sind die Abbildungsparameter mit einem Fehler behaftet. Diese Fehlerquelle wurde mit folgendem Experiment bestätigt: Der Kamerakopf fährt eine bestimmte Winkelposition an. Danach werden zufällig Bewegungen durch die Mechanik des Kamerakopfes durchgeführt. In der Abbildung 18 wird der visuelle Input der linken Kamera vor und nach den Bewegungen gezeigt. Bei 100 Bewegungen hat sich das Bild um 9 Pixel in X-Richtung und um 10 Pixel in Y-Richtung verschoben. Daraus ergibt sich ein Fehler von 0.13 Pixel pro Bewegung. Bei den durchgeführten Experimenten wurde dieser Fehler vernachlässigt. Um das System an den beschriebenen Fehler anzupassen, ist eine Rekalibration im Wahrnehmungs-Handlungs-Zyklus jedoch denkbar. Auf diese Weise könnte sich das System an die mechanischen Veränderungen adaptieren.

Eine weitere Fehlerquelle ist die Position des optischen Zentrums einer Kamera im Weltkoordinatensystem. Es wird angenommen, daß die Position des optischen Zentrums sich relativ zum Kamerakopf durch eine Drehung der Kamera nicht verändert.

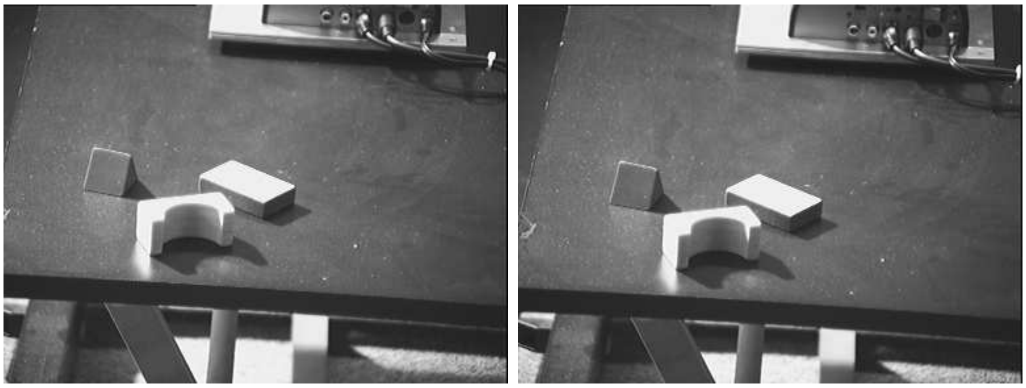


Abbildung 18: Beim linken und rechten Bild sind bei der Kamera die gleichen Winkel eingestellt. Zwischen den beiden Bildern ist die Kamera 100 Mal bewegt worden. Der Bildversatz wurde durch die Betätigung der Kameramechanik verursacht.

3.2.4 Bestimmung der intrinsischen Parameter

Die intrinsischen Parameter u_0, v_0, α_u und α_v sind abhängig vom Aufbau des CCD–Chips in der Bildebene und den Linseneinstellungen [11] (siehe Abb. 16).

Parameter	Mittelwert	Varianz	Varianz prozentual
u_0	210	24	11%
v_0	136	14	10%
α_u	721	22	3%
α_v	694	21	3%

Tabelle 2: Die Mittelwerte und Varianzen aus der Statistik von den intrinsischen Parametern (50 Datensätze), die Werteverteilung ist in Abbildung 19 und 20 dargestellt.

Sie sind unabhängig von den mechanischen Freiheitsgraden des Kamerasystems.

Für dieses Kameramodell wurde zu festen Zoom- und Fokuseinstellungen eine Statistik der Schätzung der intrinsischen Parameter entwickelt. Ein festes Paar von Zoom- und Fokuseinstellungen kann verwendet werden, da in dem interessanten Bereich von 1 - 3 Meter Entfernung von der Kamera die Tiefenschärfe groß genug ist. Es werden alle Objekte genügend scharf abgebildet. Die intrinsischen Parameter wurden mit einem Standardverfahren ermittelt (siehe 3.2.1). Als Kalibrationsobjekt wurde der Würfel (Abb. 17) verwendet.

Es wurden die intrinsischen Parameter vom TRC-Kopf mit den Einstellungen Fokus = 0.8 und Zoom = 0.9 kalibriert. Die Verteilung der Parameter ist in Abbildung 19 und 20 dargestellt. Die Statistik beinhaltet 50 Datensätze. Die sich aus der Statistik ergebenen Mittelwerte und Varianzen sind in der Tabelle 2 aufgelistet.

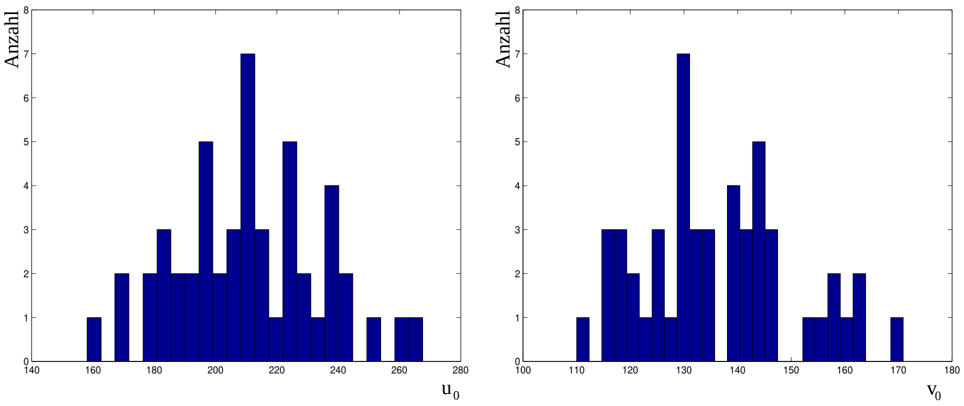


Abbildung 19: Verteilung der intrinsischen Parameter u_0 (links) und v_0(rechts) (50 Datensätze)

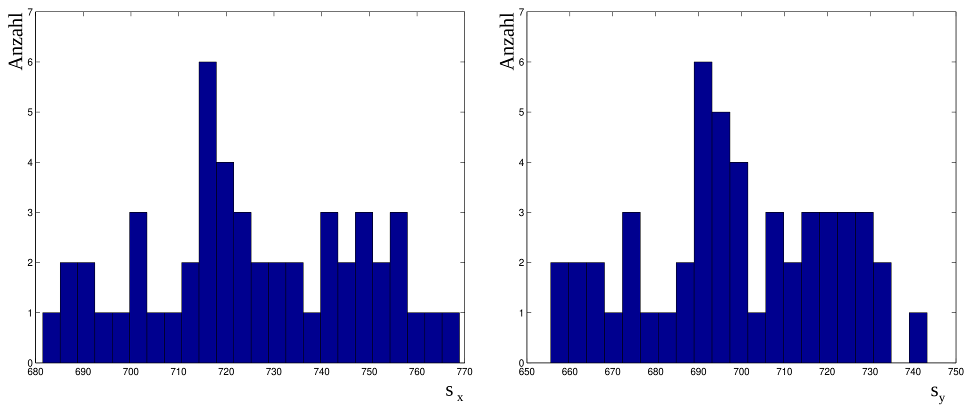


Abbildung 20: Verteilung der intrinsischen Parameter α_u (links) und α_v (rechts) (50 Datensätze)

3.2.5 3D-Punktrekonstruktion

Mit dem Wissen der intrinsischen und extrinsischen Parameter von mehreren Kameras können aus Bildinformationen 3D-Weltkoordinaten von Punkten berechnet werden. Eine wichtige Aufgabe ist hier die Lösung des Korrespondenzproblems. In unserem Fall wird darunter das Finden von Bildpunkten von zwei verschiedenen Kameras verstanden, die bezüglich ihres Wertes vom gleichen 3D-Punkt abhängig sind. Für das in dieser Arbeit beschriebene System wird in Abschnitt 4.2 für einige bestimmte Fälle Korrespondenzen ermittelt. Wenn die Korrespondenzen von zwei Bildpositionen vorhanden sind, kann mittels eines Triangulationsverfahrens (zum Beispiel [11]) die dazugehörige 3D-Weltposition berechnet werden. Dies wird in der Diplomarbeit [13] näher ausgeführt. Diese Algorithmen werden für das System, welches in dieser Arbeit beschrieben wird, verwendet.

3.3 Anmerkungen zu den Modellannahmen

Das beschriebene Modell von einem aktiven Stereokamerakopf erfüllt viele Aufgaben. Es hat aber auch seine Grenzen. Es sind viele Annahmen bezüglich des Aufbaus gemacht worden, z.B. die Lage des optischen Zentrums. Einige Annahmen sind Vereinfachungen und entsprechen nicht genau den physikalischen Gegebenheiten. Dies bedeutet, daß die Handlungen, welche mit diesem Modell durchgeführt werden, einen bestimmten Fehler besitzen. Für die Handlungen ist es daher wichtig, daß sie ihre eigene Aktionen reflektieren und eine Güte ihrer Aktion ermitteln können. Dies ermöglicht es, interne Modellannahmen anzupassen und so die Handlungen bis zu einem bestimmten Grad fehlertolerant zu machen. Das System kann sich zum Teil an die Umwelt anpassen.

Das menschliche System zeigt eine starke Adaptivität. In den frühen 50er Jahren hat Ivo Kohler [20] seine berühmt gewordenen Experimente zur Adaption der Wahrnehmung durchgeführt. Er ließ Versuchpersonen über längere Zeit Brillen tragen, die das Netzhautbild starken Veränderungen unterwarfen. Die Essenz Ivo Kohlers Versuche ist, daß sich das menschliche Sehsystem den Veränderungen durch die Brille nach einiger Zeit anzupassen vermag . Einen Überblick über die in diesem Bereich durchgeführten Experimente gibt Welch [36, 37].

In dem Modell, welches in diesem Kapitel beschrieben worden ist, können alle mechanischen Freiheitsgrade des Kamerakopfes genutzt werden. Das Modell liefert die Voraussetzung für die Berechnung der Abbildungsparameter nach Eigenbewegungen. Somit ist eine 3D-Tiefenschätzung möglich. Durch die bekannte Relationen zwischen den Koordinatensystemen und die Modellkenntnis vom Aufbau des visuellen Systems sind die Vorgaben zur Ansteuerung eines 3D-Punktes durch die Kameras gegeben.

Das Modell hat für bestimmte Parameter weitere Schwächen. Das Modell vom Kamerakopf wird in den Extrembereichen der Drehwinkel zum Teil stark fehlerhaft. Wenn der Kopf sehr weit um die Y-Achse gedreht wird, entsteht eine Drehung der einzelnen Kameras um deren optische Achse. Diese Drehung ist durch die Hardware bedingt. Durch diese Rotation entsteht ein größerer Fehler zwischen den berechneten Abbildungsparametern und den tatsächlichen Abbildungsparametern. Die Drehung um die optische Achse ist im Modell nicht berücksichtigt.

Die Grenzen des Modells bedeuten für die in dieser Arbeit realisierten Kompetenzen keine elementaren Einschränkungen.

4 Kompetenzen des Kamerasystems

Für ein autonomes System, welches unter anderem seine Umwelt explorieren soll, ist es von Vorteil, wenn dem System Kompetenzen zur Verfügung stehen. Mit diesen Kompetenzen kann sich das System selbst eine bessere Ausgangslage zum Explorieren verschaffen. Gewisse Kompetenzen gehören zum Vorwissen eines lernenden Systems. Sie müssen nicht gelernt werden. Unter diesen Kompetenzen werden in dieser Arbeit unter anderem *Aufmerksamkeit*, *mögliche* neue Objekte in der Szene Fixieren und Wiedererkennen verstanden.

Verschiedene Kompetenzen werden in Abhängigkeit der Wahrnehmung des Systems benutzt. Nach diesen Handlungen erhält das System neue Wahrnehmungen, was wieder zu weiteren Handlung führen kann. Einige Kompetenzen besitzen innerhalb eines Wahrnehmungs-Handlungs-Zykluses ein Gütemaß für die Durchführung der Handlung (z.B. Fixieren eines Objektes). Damit kann beurteilt werden, wie erfolgreich eine bestimmte Aktion war und ob diese Aktion weiter durchgeführt werden sollte.

Damit ein Ziel dieser Diplomarbeit, der taktile Kontakt zu einem Objekt, erreicht werden kann, werden verschiedene Handlungssweisen entwickelt. Das System soll eine anfangs statische Szene beobachten und auf eine Veränderung – das Hinzufügen eines Objektes – durch Fixierung und Bewegung des Roboterarmes reagieren. In diesem Kapitel werden folgende Module beschrieben: Aufmerksamkeitssteuerung, Entscheidungsfindung, ob ein neues Objekt vorhanden ist, Wiedererkennung von Objekten und Erkennung des Greifers im Bild. Die Module basieren auf dem in Kapitel 3 beschriebenen Modell eines aktiven Stereokamerakopfes.

In den folgenden Abschnitten werden mehrere Kompetenzen des visuellen Systems vorgestellt. In Abschnitt 4.1 wird erläutert, wie die Aufmerksamkeit des Systems ausgelöst wird. In Abschnitt 4.2 werden zwei Verfahren zur Detektion und Lokalisation eines Objektes beschrieben. In Abschnitt 4.4 wird erklärt, wie ein Objekt im Bild wiedergefunden werden kann. Im letzten Abschnitt 4.5 wird beschrieben, wie der Greifer visuell detektiert werden kann.

4.1 Aufmerksamkeit

Ein autonomes System, welches die Umwelt nicht willkürlich explorieren soll, muß aus seiner Umgebung *interessante Merkmale* extrahieren. Dann kann das System seine *Aufmerksamkeit* auf diese *interessanten Merkmale* richten.

Für das in dieser Arbeit entwickelte System bestehen diese *interessanten Merkmale* zunächst aus Veränderungen in der Szene, die durch das Hinzufügen von Objekten verursacht werden.

Das System beobachtet eine anfangs statische Szene und soll auf eine Veränderung – das Hinzufügen eines Objektes – reagieren. Diese Veränderung in der Szene kann auf der Zeitachse in drei Bereiche unterteilt werden. Ab dem Zeitpunkt t^0 beobachtet das System die Szene. Im ersten Abschnitt ist die Szene von t^0 bis t^1 unverändert. Zum Zeitpunkt t^1 verändert sich die Szene. Eine Person kommt z.B. in das Sichtfeld der Kameras und fügt der Szene ein Objekt zu. Anschliessend verläßt die Person das Sichtfeld der Kameras. In diesem zweiten Abschnitt verändert sich die Szene fortlaufend, z.B. durch die Bewegung einer Person. Das System wartet, bis die Szene wieder statisch wird. Dies tritt zum Zeitpunkt t^2 ein. Anschließend werden die Szeneneindrücke vom Zeitpunkt t^2 mit dem vom t^0 verglichen. Wenn ein bestimmter Unterschied erkannt wird, vermutet das System, daß ein neues Objekt vorhanden ist. Ansonsten beginnt der Algorithmus wieder bei t^0.

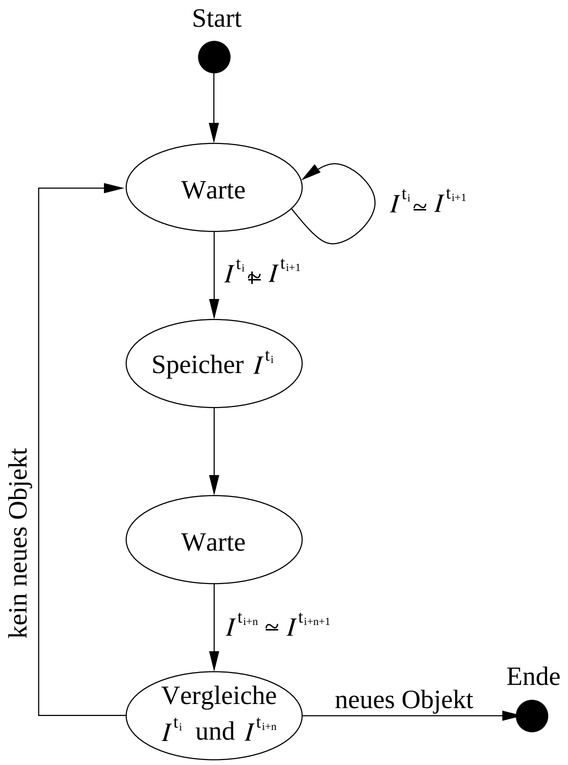


Abbildung 21: Flußdiagramm der Detektion

Algorithmus 4.1: Detektierung einer Veränderung in der Szene:
(siehe Abbildung 21-23)

Vor:
Zeit $t \in \{t_0, t_1, \ldots\}$ mit $t_i < t_{i+1}$ und $i \in \mathbb{N}$
Linkes/rechtes Bild $I_{l/r}$ zum Zeitpunkt t: $I^t_{l/r}$

1:
Warte solange, bis das linke und rechte Bild $I_{l/r}$ sich verändert.
Für die Bilder gilt dann: $I^{t_i}_{l/r} \not\simeq I^{t_{i+1}}_{l/r}$.
(siehe Abschnitt 4.1.1).

2:
Speichere $I^{t_i}_{l/r}$

3:
Warte solange, bis das linke und rechte Bild $I_{l/r}$ sich nicht mehr verändert.
Für die Bilder gilt dann: $I^{t_{i+n}}_{l/r} \simeq I^{t_{i+n+1}}_{l/r}$.

4:
Vergleiche die Bilder vor der Veränderung $I^{t_i}_{l/r}$ mit den Bildern nach der Veränderung $I^{t_{i+n}}_{l/r}$. Entscheide, ob ein Objekt hinzugefügt worden sein könnte (siehe Abschnitt 4.2), wenn ja dann gehe zu 5, ansonsten gehe zu 1.
Wenn $S(I^{t_{i+1}}_{|\Delta|}, T) > N$ gilt, wird die Veränderung auf ein neues Objekt überprüft.

5: Ende

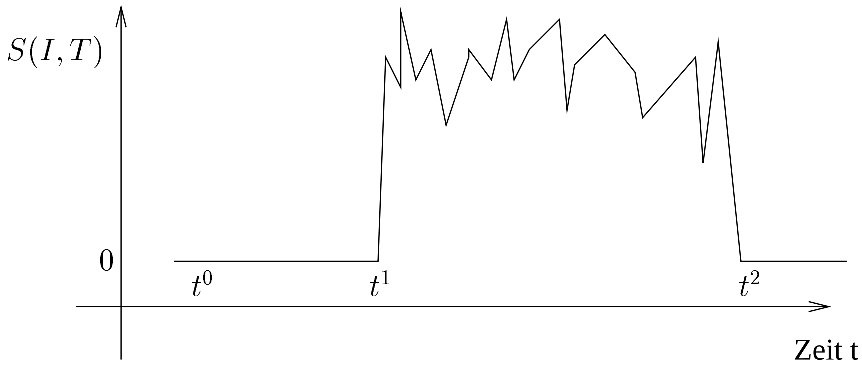


Abbildung 22: Stärke der Dynamik $S(I, T)$ (siehe Gleichung (4.13)) in der beobachteten Szene

Abbildung 23: Ablauf Detektion eines neuen Objektes

4.1.1 Veränderung in der Szene

Eine Veränderung der Szene kann durch das Betrachten der Szene zu zwei verschiedenen Zeitpunkten t_i und t_{i+1} festgestellt werden. Hierfür werden jeweils aus dem fortlaufenden Bildstrom das letzte Bild $I^{t_{i+1}}$ mit dem vorletzten Bild I^{t_i} verglichen. Dazu wird von den beiden Bildern I^{t_i} und $I^{t_{i+1}}$ die Betragsdifferenz der Grauwerte von den jeweiligen Pixeln bestimmt (siehe Gleichung (4.11) und (4.12)). Das Bild $I^{t_{i+1}}_{|\Delta|}$ wird als Betragsdifferenzbild bezeichnet:

$$I^{t_{i+1}}_{|\Delta|} = |I^{t_i} - I^{t_{i+1}}|. \tag{4.11}$$

Für alle Pixel (x, y) von dem Bild $I^{t_{i+1}}_{|\Delta|}$ gilt:

$$I^{t_{i+1}}_{|\Delta|}(x, y) = |I^{t_i}(x, y) - I^{t_{i+1}}(x, y)|. \tag{4.12}$$

Das Signal, aus dem das Betragsdifferenzbild erstellt wird, ist verrauscht. Um das Rauschen zu vermindern, wird das Betragsdifferenzbild geglättet. Zum Glätten wird ein 5×5 Binomialfilter [16] verwendet.

Um eine Veränderung vom Zeitpunkt t_i zum Zeitpunkt t_{i+1} festzustellen, werden im Betragsdifferenzbild $I^{t_{i+1}}_{|\Delta|}$ mittels der Schwellwertfunktion S alle Pixel gezählt, die über einen bestimmten Schwellwert T liegen. Die Schwellwertfunktion $S(I,T)$ (siehe Gleichung (4.13)) liefert zum Bild I die Anzahl der Bildpixel, deren Grauwert oberhalb des Schwellwertes T liegen.

$$S(I,T) = \sum_{I(x,y) \geq T} 1 \tag{4.13}$$

Somit ist $S(I^{t_{i+1}}_{|\Delta|}, T)$ ein Maß der Veränderung vom Bild I^{t_i} zum Bild $I^{t_{i+1}}$. Durch den Schwellwert T kann die Größe der minimalen Veränderung geregelt werden.

Im Algorithmus von Abschnitt 4.1 ist eine Veränderung vorhanden, wenn die Formel (4.14) für die rechte und linke Kamera gilt, ansonsten hat die Szene sich für das System nicht verändert. Die Anzahl der veränderten Pixel muß also größer als die Zahl N sein.

$$S(I^{t_{i+1}}_{|\Delta|}, T) > N \tag{4.14}$$

Dieses Verfahren ist ein low-level Verfahren. Es ist einfach und schnell. Das Kriterium funktioniert nur, wenn sich die Grauwerte genügend stark ändern. Je höher der Kontrast zwischen dem hinzugefügten Objekt und dem verdeckten Hintergrund ist, um so größer sind die Grauwerte vom Objekt im Betragsdifferenzbild. Problematisch sind Beleuchtungsveränderungen. Wenn zwischen den Zeitpunkten t^1 und t^2 eine Beleuchtungsveränderung stattfindet, z.B. Lampe, Minderung von Licht durch Wolken usw., ist eine starke Änderung im Differenzbild zu erkennen, so daß das System die Veränderung auf ein neues Objekt testet.

4.2 Objektdetektion in einer Stereoszene

Im vorherigen Abschnitt wurde ein Modul beschrieben, welches die Veränderung von Bildern ermittelt. Diese Veränderung kann in einem Bildstrom durch ein neues Objekt in der Szene oder durch andere Beleuchtungsverhältnisse verursacht werden. Das Modul, welches in diesem Abschnitt beschrieben wird, soll eine Hypothese erstellen, ob die Veränderung in der Szene durch ein neues Objekt verursacht sein kann. Wenn diese Hypothese positiv ist, dann soll das Objekt jeweils im linken und rechten Bild lokalisiert werden. Das Modul soll gegebenenfalls korrespondierende Punkte zurückgeben, so daß aus diesen mittels Triangulation eine 3D-Position des Objektes berechnet werden kann.

Im folgenden Abschnitt werden zwei verschiedene Verfahren vorgestellt. Das eine Verfahren aus Abschnitt 4.2.1 verwendet Grauwertdifferenzbilder und berechnet den Schwerpunkt und die Streuung vom Bildgebiet des vermuteten Objekts. Das andere Verfahren aus Abschnitt 4.2.3 verwendet das erste Verfahren als Vorstufe. Um das Korrespondenzproblem von zusammengehörigen Objektbildpunkten in Stereobildern genauer zu lösen, wird eine rudimentäre Objektrepräsentation aus Gaborwaveletsjets verwendet. Diese Objektrepräsentation wird in Abschnitt 4.2.2 beschrieben.

4.2.1 Grauwertbasierter Ansatz zur Objektdetektion

Dieses Verfahren berechnet analog zu Abschnitt 4.1.1 für das linke und rechte Bild jeweils das Betragsdifferenzbild $I_{|\Delta|}^{t_{i+1}}$. Wenn ein Objekt zwischen den Zeitpunkten t_i und t_{i+1} in die Szene gebracht worden ist, sollte dies durch ein zusammenhängendes Gebiet im Bild $I_{|\Delta|}^{t_{i+1}}$ festgestellt werden können.

Als Kriterium, ob ein Bild ein relativ zusammenhängendes Gebiet besitzt, wird der Schwerpunkt und die Streuung von allen Pixeln, die sich genügend stark geändert haben, berechnet. Alle Pixel mit einer großen Veränderung zeichnen sich im Bild $I_{|\Delta|}^{t_{i+1}}$ durch einen hohen Grauwert aus. Es gilt in der Regel, je größer der Grauwert, um so größer die Veränderung im Bild. Der maximale Grauwert T vom Bild $I_{|\Delta|}^{t_{i+1}}$ wird durch die Gleichung (4.15) beschrieben.

$$T = \max_{x,y}(I_{|\Delta|}^{t_{i+1}}(x,y)) \tag{4.15}$$

Der Algorithmus 4.2 stellt eine Hypothese auf, ob sich ein neues Objekt in der veränderten Szene befinden kann. Wenn diese Hypothese positiv ist, wird zusätzlich die Objektposition im Bild berechnet. Der Algorithmus trifft die Entscheidung in Abhängigkeit von mehreren Parametern: T_{min}, T_σ und f. Der Parameter T_{min} ist der Schwellwert, über dem die maximale Veränderung T liegen muß. Wenn T kleiner als der Schwellwert ist, ist die Veränderung für das System nicht groß genug und die Objektsuche wird abgebrochen. T_σ ist der obere Grenzwert für die Streuung, die sich aus der Verteilung der Grauwerte um den Schwerpunkt herum ergeben. Der Wert von f bestimmt, welche Pixelwerte in die Berechnung vom Schwerpunkt und der Streuung einfließen. Es werden alle Pixelwerte gewertet, die größer als T/f sind.

Anschließend wird die Position des Schwerpunkts $\mathbf{C}(I_{|\Delta|}^{t_{i+1}}, T/f) = \begin{pmatrix} x_C & y_C \end{pmatrix}^T$ und die dazugehörige Streuung $\sigma(I_{|\Delta|}^{t_{i+1}}, T/f)$ berechnet (siehe Gleichung (4.16), (4.17) und [4]). Der Schwerpunkt $\mathbf{C}$ wird über die Positionen aller Pixel gemittelt, deren Grauwert sich genügend stark verändert hat. Die Streuung σ berechnet sich aus dem euklidischen Abstand vom Schwerpunkt $\mathbf{C}$ und der Positionen der relevanten Pixel. Für den Algorithmus ist ein neues Objekt vorhanden, wenn entweder an der Position des Schwerpunkts die Veränderung, also die Grauwertdifferenz genügend groß ist oder die Streuung kleiner als der Grenzwert T_σ ist.

Abbildung 24: Linkes und rechtes Betragsdifferenzbild $I_{|\Delta|}^{t_{i+1}}$

Diese Berechnungen werden für das linke und rechte Bild durchgeführt. Als Standardparameter wurden in dieser Arbeit $T_{min} = 64$, $T_\sigma = 20$ und $f = 2$ verwendet.

$$\mathbf{C}(I,T) = \begin{pmatrix} x_C \\ y_C \end{pmatrix} = \frac{1}{S(I,T)} \sum_{I(x,y)\geq T} \begin{pmatrix} x \\ y \end{pmatrix} \tag{4.16}$$

$$\sigma(I,T)^2 = \frac{1}{S(I,T)-1} \sum_{I(x,y)\geq T} \left\| \mathbf{C}(I,T) - \begin{pmatrix} x \\ y \end{pmatrix} \right\|_2^2 \tag{4.17}$$

Wenn die beiden Kameras ein ungefähr gleiches Blickfeld haben, dann wird das visuelle System in der Regel im linken und rechten Betragsdifferenzbild jeweils ein Gebiet detektieren, welches das Kriterium für ein neues Objekt erfüllt. Von dieser Annahme, daß die Kameras ein ungefähr gleiches Blickfeld haben, kann ausgegangen werden, da das System die Kameras entsprechend steuern kann.

Wenn somit das System im linken und rechten Bild jeweils ein Gebiet detektiert, wird davon ausgegangen, daß die beiden Gebiete zum gleichen Objekt gehören. Das Gebiet wird jeweils durch seinen Schwerpunkt und seine Streuung beschrieben. Da die beiden Kameras das Objekt aus sehr ähnlichen Perspektiven wahrnimmt, haben die Gebiete im Betragsdifferenzbild in der Regel eine vergleichbare Gestalt. Wenn diese Annahmen gelten, dann ist das Korrespondenzproblem von gemeinsamen Objektbildpunkten in Stereobildern für die beiden Schwerpunkte somit gelöst. Aus den beiden 2D-Positionen kann mit den vorhandenen Abbildungsparametern die 3D-Position berechnet werden. Somit kennt das System von dem neuen Objekt in der Szene eine 3D-Position.

In Abbildung 25 sind mögliche Zwischenresultate aus dem Betragsdifferenzbild angegeben. Der Punkt S gibt die Position des berechneten Schwerpunkts an. Das Beispiel 1) ist ein einfaches Beispiel von einem konkaven Objekt. Es erfüllt sowohl die Bedingung 4a (Schwerpunkt) als auch 4b (Streuung) aus dem Algorithmus. In Beispiel 2) sind zwei Objekte in die Szene hinzugefügt worden.

Algorithmus 4.2: Neues Objekt

Variable: T
Konstante: T_{min}, T_σ, f
Rückgabewerte: $x_C \in \mathbb{R}, y_C \in \mathbb{R}$

1:
Berechne $I_{|\Delta|}^{t_{i+1}}$

2a:
Berechne T (siehe Gleichung (4.15))

2b:
falls $T < T_{min}$ dann Ende sonst 3a

3a:
Berechne die Position des Schwerpunkts $\mathbf{C}(I_{|\Delta|}^{t_{i+1}}, T/f) = (x_C, y_C)^T$ von der Veränderung mittels der Gleichung (4.16).

3b:
Berechne die Streuung $\sigma(I_{|\Delta|}^{t_{i+1}}, T/f)$ bzgl. des Schwerpunkts $\mathbf{C}(I_{|\Delta|}^{t_{i+1}}, T/f)$ von der Veränderung mittels der Gleichung (4.17).

4:
Wenn die Bedingung 4a oder 4b gilt, dann wird vermutet, daß an der Stelle (x_C, y_C) im Bild ein neues Objekt vorhanden ist.

Bedingung 4a: $I_{|\Delta|}^{t_{i+1}}(x_C, y_C) > T/f$
Bedingung 4b: $V(I_{|\Delta|}^{t_{i+1}}, T/f) < T_\sigma$

5:
Ende

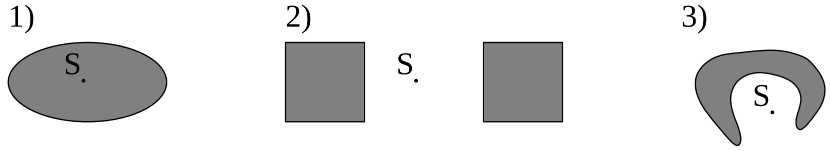


Abbildung 25: Anwenden des Algorithmus auf die Grafiken 1)-3). Die graue Fläche ist das Gebiet aus dem Betragsdifferenzbild. S ist der dazu berechnete Schwerpunkt.

Das Kriterium 4a (Schwerpunkt) wird nicht erfüllt. Das Kriterium 4b (Streuung) hängt vom Parameter T_σ ab, also welche maximale Streuung zugelassen wird. Wenn die beiden Objekte genügend dicht zusammen stehen, bewertet der Algorithmus sie als ein Objekt, ansonsten wird der Algorithmus mit dem Ergebnis, daß kein neues Objekt gefunden wurde, beendet. Da nur die Pixel im Betragsdifferenzbild einen hohen Wert besitzen, wo der Kontrast der Veränderung möglichst groß ist, wird ein konkaves Objekt nicht immer als ein konkaves Gebiet im Betragsdifferenzbild angezeigt. Es können Höhlen und Löcher entstehen. Das Ergebnis würde dem Beispiel 3) entsprechen. Dieses wird mit dem Kriterium der Streuung (Bedingung 4b) als ein neues Objekt erkannt.

Weitere Probleme können beim Vergleich des linken Bildes mit dem rechten Bild entstehen, wenn die detektierten Gebiete nicht die gleiche Form haben. In Experimenten ist es vorgekommen, daß sowohl das linke als auch das rechte Bild dem Beispiel 3) entsprochen haben. Bei beiden Gebieten sind jeweils andere Höhlen und Löcher vorhanden gewesen, so daß der berechnete Schwerpunkt nicht zum gleichen 3D-Punkt des Objektes gehörte. In diesem Fall wird die Annahme, daß beide Punkte zum gleichen 3D-Punkt gehören, verletzt. Somit wird aus den 2D-Informationen eine falsche 3D-Position berechnet. In Abschnitt 4.2.2 und 4.2.3 wird ein weiteres Kriterium vorgestellt, was bei diesem Problem der Korrespondenzfindung ansetzt. Es wird eine Repräsentation von der Bildumgebung aus dem linken Bild im rechten Bild gesucht.

4.2.2 Rudimentäre Objektrepräsentation

Das System soll ein neues Objekt detektieren und lokalisieren. Im Verfahren aus Abschnitt 4.2.1 wird implizit angenommen, daß die berechneten 2D-Positionen aus dem linken und rechten Bild aus dem gleichen 3D-Objektpunkt resultieren. Die Bildbereiche um die jeweiligen 2D-Positionen werden nicht mit dem Bildbereich im jeweils anderen Bild verglichen. Dadurch hat das oben genannte Verfahren Schwächen, da die zu berechnende 3D-Objektposition zum Teil stark fehlerhaft ist. Um diese Schwäche zu beheben, wird in diesem Abschnitt eine rudimentäre Objektrepräsentation eingeführt, die unter anderem in dem Verfahren von Abschnitt 4.2.3 verwendet wird.

Als Objektrepräsentation wird ein 2D-Gitter verwendet (siehe Abb. 26). Ein Gitter wird durch die Gitterlänge L_x und L_y beschrieben. Ein Gitter besitzt $(2\,L_x+1)(2\,L_y+1)$ Knoten. Jeder Gitterknoten repräsentiert die gleiche Art von Merkmalen. Der Merkmalsvektor ist abhängig von der Position des Knotens relativ zum Mittelpunkt (Position (0,0)) des Gitters. Als Merkmalsvektor werden *Jets* verwendet (siehe unten) [23]. Der Gitterabstand Δx und Δy von den einzelnen Knoten ist jeweils in X- und Y-Richtung konstant. Aus der Knotenposition und dem Gitterabstand berechnet sich die Bildposition des jeweiligen Jets.

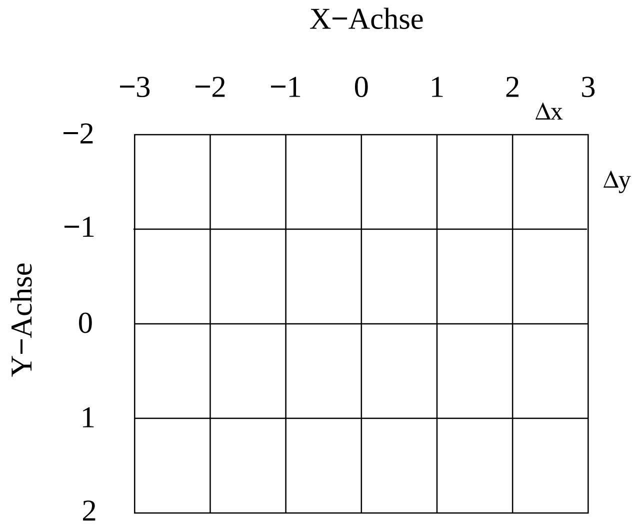


Abbildung 26: Gitter von Jets

Jets sind abgeleitet von einer Menge linarer Filteroperationen. Die Filteroperationen sind Faltungen von einer Menge von Gaborwavelets ψ mit einem Bild $I(\underline{x})$. Die Gaborwavelets $\psi_{\underline{k}}$ sind in der Wellenlänge und Orientierung mit $\underline{k}$ parametrisiert (siehe Gleichungen (4.18)-(4.21) [7]). Die $\psi_{\underline{k}}$ haben die Form von ebenen Wellen, die mit der Gaußschen Umhüllungsfunktion beschränkt sind.

$$\psi_{\underline{k}}\left(\underline{x}\right) = \frac{k^2}{\sigma^2} exp\left(-\frac{\underline{k}^2\underline{x}^2}{2\sigma^2}\right)\left[exp\left(i\,\underline{k}\,\underline{x}\right) exp\left(-\sigma^2/2\right)\right] \tag{4.18}$$

$$\underline{k}_{(d,l)} = (\phi_d, k_l) = \left(\Delta\phi\, d, k_{max}\, k_{fac}^{l}\right) \tag{4.19}$$

$$d = 0,\ldots,d_{max}, \quad l = 0,\ldots,l_{max}, \quad \Delta\phi = \frac{\pi}{d_{max}+1}, \tag{4.20}$$

$$k_{max} = \frac{\pi}{2} \quad und \quad k_{fac} = \frac{1}{\sqrt{2}} \tag{4.21}$$

$\underline{k}$ wird mit d_{max} Orientierungen und l_{max} Frequenzen abgetastet. Ein Jet $\underline{J} = (a_j)_j$ ist ein Vektor der Dimension $d_{max}\, l_{max}$ (siehe Abb. 27). Die Vektorwerte a_j ergeben sich aus dem Betrag der komplexen Filteranwort der parametrisierten Gaborwavelets auf eine Umgebung von einer Position im Bild. Ein Jet $\underline{J}$ beschreibt die Umgebung um seine Position [23].

Aus einer Faltung von einem Grauwertbild I mit einer Menge von Gaborwavelets $\psi_{\underline{k}}$ erhält man ein Jetbild I_J. In diesem Jetbild I_J wird jede Bildposition durch einen Jet $\underline{J}$ dargestellt. Der Jet $\underline{J}$ ist das Resultat der Faltung der Gaborwavelets $\psi_{\underline{k}}$ an der jeweiligen Bildposition im Bild I.

$$F_{\psi_{\underline{k}},\, d_{max},\, l_{max}} : \underbrace{\mathbb{R} \times \mathbb{N}^2}_{Bild\ I} \longrightarrow \underbrace{\mathbb{R}^{d_{max}\, l_{max}}}_{Jet} \times \mathbb{N}^2 \tag{4.22}$$

$$F_{\psi_{\underline{k}},\, d_{max},\, l_{max}} : I \longrightarrow I_J \tag{4.23}$$

Ein Gitter G wird am Knoten (x, y) mit dem Jet $\underline{J}(x, y)$ beschrieben. Das Gitter G repräsentiert in Abhängigkeit von Gitterlänge und Knotenabstand eine bestimmte Bildregion. Wenn in dieser Bildregion ein Objekt abgebildet ist, repräsentiert das Gitter dieses Objekt.

Ein wichtiger Aspekt von einer Objektrepräsentation ist ein Ähnlichkeitsmaß. Objekte, die mit dem Jet-Gitter beschrieben werden, sollten in anderen Bildbereichen wiedergefunden werden können. Die Ähnlichkeit S_G von zwei Gittern G' und G'' berechnet sich aus der Ähnlichkeit der einzelnen Jets $\underline{J}(x, y)$ (Gleichung (4.24)).

$$S_G(G', G'') = \frac{1}{(2\, L_x + 1)(2\, L_y + 1)} \sum_{(n,m)} S_J\left(\underline{J}'(n, m), \underline{J}''(n, m)\right) \tag{4.24}$$

Die Gleichung (4.25) gibt ein Ähnlichkeitsmaß S_J für die beiden Jets $\underline{J}' = ({a'}_j)_j$ und $\underline{J}'' = ({a''}_j)_j$ [23]:

$$S_J(\underline{J}', \underline{J}'') = \frac{\langle \underline{J}', \underline{J}'' \rangle}{\|\underline{J}'\| \|\underline{J}''\|} = \frac{1}{\sqrt{\sum_j a'^2_j \sum_j a''^2_j}} \sum_j {a'}_j {a''}_j \tag{4.25}$$

Daraus ergibt sich mit der Funktion P_G ((4.26) und (4.27)) ein Potential I_P vom Gitter G auf das Jetbild I_J.

$$P_G : \underbrace{\mathbb{R}^{d_{max}\, l_{max}}}_{Jet} \times \mathbb{N}^2 \longrightarrow \underbrace{\mathbb{R} \times \mathbb{N}^2}_{Bild} \tag{4.26}$$

$$P_G : I_J \longrightarrow I_P \tag{4.27}$$

$$P_{G''}(I_J)|_{u,v} = S_G(G', G'') \text{ mit } G' \in I_J \text{ am Aufpunkt } (u,v) \text{ in } I_J \tag{4.28}$$

Daraus ergibt sich die größte Ähnlichkeit A_{max} (4.29) des Gitters G'' im Bild I_P,

$$A_{max} = \max_{\underline{x}}(I_P(\underline{x})) \tag{4.29}$$

womit dann auch die Positionen $\underline{A}$ der größten Ähnlichkeiten berechnet werden können.

$$\underline{A} = \left\{ \underline{u} \in \mathbb{N}^2 | I_P(\underline{u}) = \max_{\underline{x}}(I_P(\underline{x})) \right\} \tag{4.30}$$

Wenn die Menge $\underline{A}$ genau ein Element besitzt, ist die Position der maximalen Ähnlichkeit eindeutig.

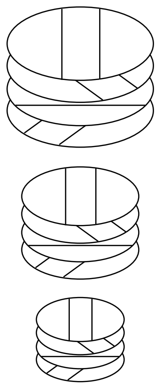


Abbildung 27: Jet mit vier Orientierungen und drei Frequenzen

4.2.3 Gaborwaveletbasierter Ansatz zur Objektdetektion

In diesem Abschnitt wird ein Verfahren mit Gaborwavelets zur Detektion von einem neuen Objekt vorgestellt, welches die rudimentäre Objektrepräsentation aus Abschnitt 4.2.2 verwendet. Es ist eine Erweiterung des Verfahrens aus Abschnitt 4.2.1. Die Erweiterung setzt bei dem schon dort angesprochenen Problem der Korrespondenzfindung an. Im Abschnitt 4.2.1 wurden die Korrespondenzen über die Schwerpunkte im Betragsdifferenzbild erstellt.

Das Verfahren aus diesem Abschnitt (siehe Algorithmus 4.3) erstellt ein Jet-Gitter (siehe Abschnitt 4.2.2) von der Umgebung des Schwerpunktes aus dem linken Bild. Dieses Gitter wird anschließend mittels der Ähnlichkeitsverteilungsfunktion P_G ((4.26)-(4.28)) im rechten Bild gesucht. Der neue korrespondierende Punkt im rechten Bild ist der Punkt mit der größten Ähnlichkeit. Die Ähnlichkeitswerte liegen alle zwischen 0 und 1, wobei 1 die maximale Ähnlichkeit ist. Im verglichenen Bild muß der korrespondierende Punkt mindestens eine Ähnlichkeit von der Größe A_{min} besitzen. Als Standardwert wird $A_{min} = 0.6$ verwendet.

Für das Jet-Gitter werden mehrere Parameter benötigt, einerseits d_{max} und l_{max} für den Jet von Gaborwavelets und andererseits $L_x, L_y, \Delta x$ und Δy für den Gitteraufbau. Die Parameter werden mit den folgenden Standardwerten aus 4.31 verwendet .

$$d_{max} = 4, l_{max} = 2, \Delta x = \Delta y = 4 \tag{4.31}$$

Um die Parameter L_x und L_y für das Gitter zu bestimmen, wird das Bild $I^{t_{i+1}}_{|\Delta|}$ weiter analysiert. Es werden die Koordinaten der Punkte $P', P'' \in \mathbb{N}^2$ im Bild bestimmt. P' und P'' definieren ein Rechteck R im Bild $I^{t_{i+1}}_{|\Delta|}$ mit $P'_x < P''_x$ $P'_y < P''_y$ und minimaler Fläche von R, so daß alle Pixel aus dem Bild $I^{t_{i+1}}_{|\Delta|}$ mit einem Grauwert größer als T/f im Rechteck enthalten sind (siehe Abb. 28). Das Rechteck R ist ein Maß für die Größe des möglichen neuen Objekts.

Die Gittergrößen L_x und L_y werden so berechnet, daß das entstehende Gitter von der Größe her das Rechteck R möglichst bedeckt, wobei L_x und L_y mindestens die Größe 1 haben und eine obere Schranke nicht überschreiten sollten. Die obere Schranke ergibt sich zum einen aus der Größe der relevanten Objekte im Bild und daß der Nutzen ab einer bestimmten Größe nicht mehr den Rechenaufwand rechtfertigt. L_x und L_y werden nach (4.32) und (4.33) berechnet.

$$1 < L_x < 11 \quad \text{und} \quad L_x = \frac{(P'' - P')|_x}{2\Delta x} - \frac{1}{2} \tag{4.32}$$

$$1 < L_y < 11 \quad \text{und} \quad L_y = \frac{(P'' - P')|_y}{2\Delta y} - \frac{1}{2} \tag{4.33}$$

Für den Algorithmus liefert die Bildserie aus Abbildung 23 die Eingabewerte. Abbildung 24 ist das sich daraus ergebende Betragsdifferenzbild. Aus dem linken Ursprungsbild wird das Jet-Gitter G_l gebildet. In Abbildung 29 sind zwei Ähnlichkeitsverteilungen I_P abgebildet. Das linke Bild ist die Ähnlichkeit von G_l mit dem linken Bild. Das rechte Bild ist die Ähnlichkeit von G_l mit dem rechten Bild.

Viele Objekte, die dem ersten Verfahren Probleme bereiten, werden mit dem zweiten Verfahren gut lokalisiert. Die Jet-Repräsentation hat weitere Vorteile.

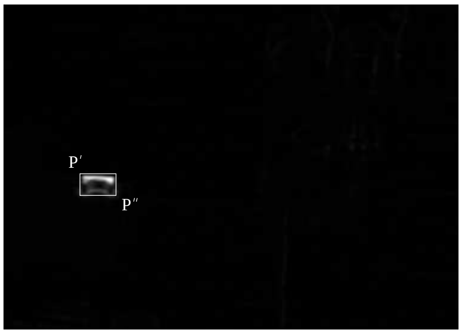


Abbildung 28: Begrenzungsrechteck R

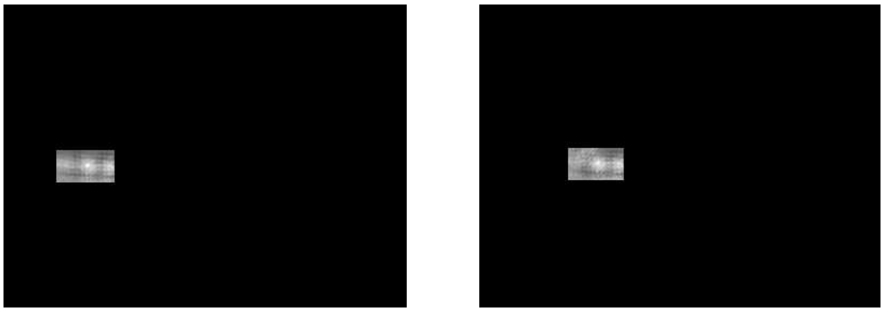


Abbildung 29: Linkes und rechtes Bild I_P der Ähnlichkeitsverteilung

Sie ist robuster gegen Helligkeitsschwankungen als das erste Verfahren. Die Jet-Gitter bilden eine gute Ausgangslage für den weiteren Handlungsablauf. Nachdem das Objekt detektiert und lokalisert ist, fixiert das System die 3D-Position vom Objekt. Danach kann es mit Hilfe der Jet-Gitter wieder lokalisiert werden. Hierzu weiteres in Abschnitt 4.3.

Der vorgestellte Algorithmus kann in der Ausführung beschleunigt werden, in dem der Suchraum eingeschränkt wird. Einerseits kann im rechten Bild davon ausgegangen werden, daß das interessante Objekt innerhalb des Rechtecks R_r liegt. Das Rechteck R_r schließt alle Pixel ein, die sich im Betragsdifferenzbild durch große Werte ausgezeichnet haben, somit braucht die Ähnlichkeitsfunktion nur auf dieses Gebiet angewendet werden. Andererseits ist die Ähnlichkeitsfunktion eine glatte Funktion. Es ist somit sinnvoll, anfangs nicht alle Pixelpositionen zu vergleichen, sondern in Abhängigkeit von Δx und Δy das Bild mit einer Schrittweite zu vergleichen. Es wird dann im unterabgetasteten Bild das Maximum über die Ähnlichkeit gebildet. Anschließend wird in einer Umgebung um die Position des Maximums pixelweise die Ähnlichkeit gebildet.

Algorithmus 4.3
mit dem gaborwaveletbasierten Ansatz zur Objektdetektion:

Eingabe I_l und I_r
Variable: T
Konstante: A_{min}

1:
Wende den Algorithmus 4.2 auf das linke und rechte Bild I_l und I_r an. Wenn für beide Bilder die Kriterien aus dem Algorithmus erfüllt sind, gehe zu 2), ansonsten Ende.

2:
Berechne eine Jet-Gitter Repräsentation G_l aus dem linken Bild I_l an der Position des in 1) gefundenen Schwerpunktes (x_{lC}, y_{lC}).

3:
Erstelle aus dem rechten Bild I_r eine Jetbild I_J ((4.22) und (4.23)).

4:
Erstelle das Potential I_P vom Jetbild I_J bezüglich der Repräsentation G_l mit der Funktion P_{G_l} aus (4.26)-(4.28)

5:
Suche im Potential I_P den maximalen Ähnlichkeitswert A_{max} und dessen Position $\underline{A}$ ((4.29)-(4.30)).

6:
Wenn A_{max} größer als A_{min} ist, berechne aus den 2D-Positionen (x_{lC}, y_{lC}) und (x_a, y_a) die 3D-Position, ansonsten Ende

7:
Ende

4.3 Fixieren

Ein wichtiges Grundmodul ist das Fixieren. Wenn man einen Gegenstand oder einen Ort näher betrachten möchte, ist es von Vorteil, wenn dieser mit einer hohen Bildauflösung betrachtet werden kann. Beim Menschen ist im Bildzentrum die Auflösung um ein vielfaches höher als am Bildrand [24]. Um bei einer Kamera die Auflösung zu erhöhen, kann der Zoom verwendet werden. Damit das betrachtete Objekt beim heranzoomen im Bildbereich bleibt, sollte das Objekt vorher vom System fixiert worden sein.

4.3.1 Fixieren eines 3D-Punktes

Für das Fixieren des Kamerasystems werden nur die externen Parameter des Kameramodells benötigt [6]. Die Kameras sollen den Punkt $\mathbf{Q} = (Q_x, Q_y, Q_z)^T$ im Weltkoordinatensystem $\mathbf{O}$ (siehe Abschnitt 3.1.1) fixieren. Hierfür werden zuerst die Drehwinkel χ_H und ϕ_H für den Kamerakopf mit (4.34) berechnet (siehe Abb. 30). Nach der Berechnung müssen die Winkel bezüglich der vorgegebenen oberen/unteren Ansteuerungswerte durch das Kamerasystem überprüft werden. Falls die Winkel χ_H und ϕ_H außerhalb der jeweiligen oberen/unteren Grenzwerte liegen, nehmen die Winkel den jeweiligen Grenzwert an. In bestimmten Fällen kann ein Punkt noch mittels geeigneter Drehwinkel um die Vergenzachse fixiert werden.

Um den Drehwinkel $\chi_{l/rC}$ um die Vergenzachse der linken/rechten Kamera berechnen zu können, muß der Punkt $\mathbf{Q}$ mit (4.35) in das jeweilige Kamerakoordinatensystem transformiert werden. Dann kann der Drehwinkel $\chi_{l/rC}$ mit (4.36) berechnet werden. Auch hierbei sind die jeweiligen Grenzwerte zu berücksichtigen. Wenn alle vier Drehwinkel berechnet sind, kann der Kamerakopf entsprechend gedreht werden. Die Formel (4.34) ergibt sich aus der Winkelbeziehung gemäß Abbildung 30 und [4]. Die Vorzeichen ergeben sich aus der Ausrichtung des Koordinatensystems und der jeweiligen Drehrichtung des Kamerakopfes. Die Formel (4.36) leitet sich entsprechend her.

$$\chi_H = -\arctan(Q_x, Q_z) \quad \text{und} \quad \phi_H = -\arctan\left(\frac{Q_y}{\sqrt{Q_x^2 + Q_z^2}}\right) \tag{4.34}$$

$$\mathbf{Q}_l = \begin{pmatrix} x_l \\ y_l \\ z_l \end{pmatrix} = \mathbf{C}\mathbf{R}_H\mathbf{Q} - \mathbf{t}_l \quad \text{und} \quad \mathbf{Q}_r = \begin{pmatrix} x_r \\ y_r \\ z_r \end{pmatrix} = \mathbf{C}\mathbf{R}_H\mathbf{Q} - \mathbf{t}_r \tag{4.35}$$

$$\chi_{lC} = \arctan(x_l, z_l) \quad \text{und} \quad \chi_{rC} = \arctan(x_r, z_r) \tag{4.36}$$

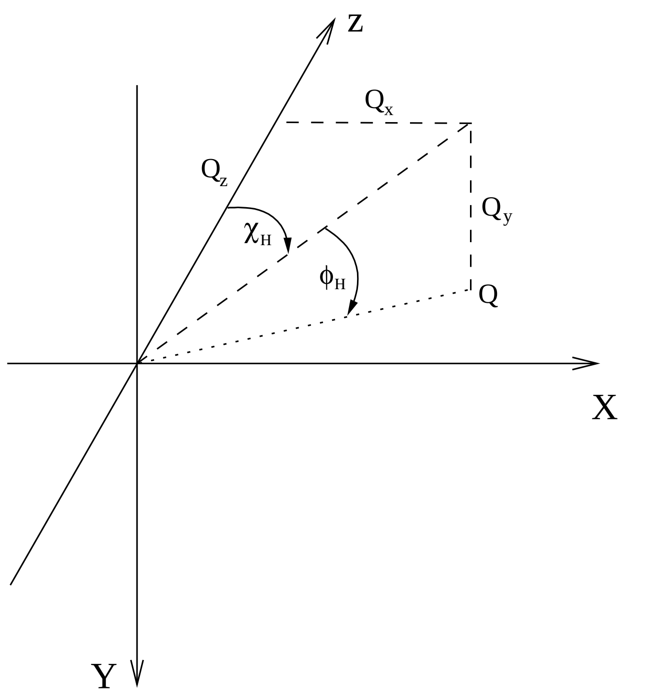


Abbildung 30: Grafische Darstellung der Winkelberechnung für den Punkt **Q**

4.3.2 Fixieren eines Objektes

Um ein Objekt zu fixieren, muß dem System die 3D-Position **Q** im Weltkoordinatensystem bekannt sein. Die 3D-Position eines neuen Objekts kann mit einem Verfahren aus Abschnitt 4.2 bestimmt werden. Anschließend wird diese Position mit Hilfe des Verfahrens aus 4.3.1 fixiert. Dieser Handlungsablauf ist überwiegend erfolgreich, wobei das System den Erfolg der Fixierung nicht beurteilen kann, da keine Kontrollmöglichkeit gegeben ist.

Wenn davon ausgegangen werden kann, daß der Blickwinkel auf das Objekt vor und nach der Fixierung sich nur wenig geändert hat, bietet die Objektrepräsentation mit dem Jet-Gitter eine Kontrollmöglichkeit. Hierfür muß vor der Fixierung ein Jet-Gitter G_l und G_r um die Objektposition $\mathbf{Q}_{l/r} \in \mathbb{N}^2$ im linken und rechten Bild erstellt werden. Die Objektposition ist durch die Verfahren aus Abschnitt 4.2 bekannt. Dieses Jet-Gitter wird dann nach der Fixierung in der jeweiligen Bildtransformierten gesucht. Das Wiedererkennen von Objekten in Bildern wird in Abschnitt 4.4 beschrieben. Die Wiedererkennung liefert dem System zwei Aussagen: Zum einen die neue Objektposition $\mathbf{Q'}_{l/r} \in \mathbb{N}^2$ im Bild und zum anderen den dazugehörigen Ähnlichkeitswert $A_{l/r}$.

Mit dem Ähnlichkeitswert kann eine Aussage getroffen werden, ob die Fixierung erfolgreich war oder ob durch vorherige Fehler das Objekt nicht gefunden werden konnte. Mit der 2D-Position $\mathbf{Q'}_{l/r}$ kann unter der Annahme, daß die Fixierung erfolgreich war, eine Aussage über die Güte der Fixierung gemacht werden. Hierfür wird der 3D-Objektpunkt $\mathbf{Q}$ mit der Projektionsmatrix aus (3.3) und (3.10) ins Bild projektiert. Die Projektion von $\mathbf{Q}$ ist $\mathbf{Q}''_{l/r} \in \mathbb{N}^2$ (4.37).

$$\tilde{\mathbf{Q}}''_{l/r} = \tilde{\mathbf{P}}\tilde{\mathbf{Q}} \tag{4.37}$$

Der Wert q_Δ aus (4.38) ist die euklidische Norm der Differenz von der angenommenen Position und der gefundenen Position in Pixel.

$$q_\Delta = \|\mathbf{Q}''_{l/r} - \mathbf{Q}'_{l/r}\|_2 \tag{4.38}$$

Der Wert q_Δ sollte möglichst klein werden. In Abhängigkeit von q_Δ kann eine weitere Fixierung durchgeführt werden, bis p_Δ kleiner als ein Schwellwert q_S ist. Der Schwellwert sollte nicht zu klein sein, da Modellungenauigkeiten vorhanden sind und eine genauere Fixierung um wenige Pixel durch die Ungenauigkeit der Motoren nicht möglich ist. Als Standardwert für q_S wurde 8 gewählt.

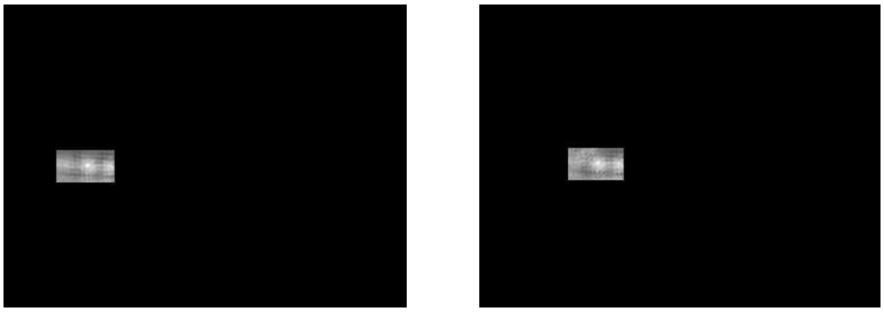


Abbildung 31: Das detektierte Objekt

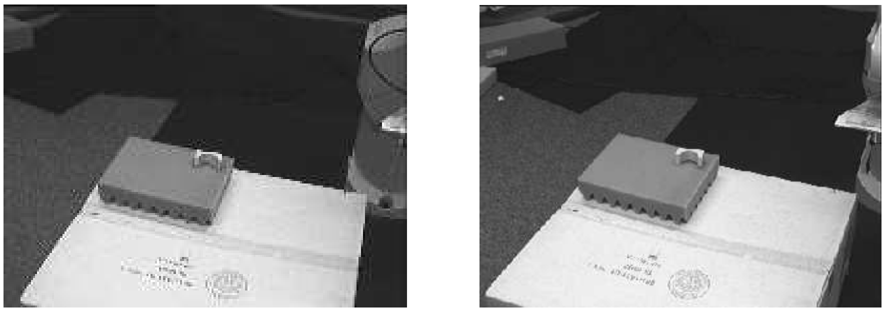


Abbildung 32: Linkes und rechtes Bild nach der Fixierung

In der Abbildung 32 wurde das Objekt, welches in 4.2 verwendet und in den Bildern 31 detektiert wurde, fixiert.

4.4 Wiedererkennung des Objektes

Um ein Objekt im Bild wiederzuerkennen, wurde ein Verfahren entwickelt, welches auf dem Jet-Gitter aus Abschnitt 4.2.2 basiert. Für die Wiedererkennung wird für das linke und rechte Bild jeweils eine vorhandene rudimentäre Objektrepräsentation $G_{l/r}$ benötigt. Diese Objektrepräsentation kann in einem vorherigen Handlungsschritt erstellt worden sein.

Auf das Bildsignal $I_{l/r}$ wird zuerst mit der Funktion $F_{\psi_{\underline{k}}}$ ein Jetbild $I_{J,\,l/r}$ erstellt. Danach wird mit der Ähnlichkeitsfunktion $P_{G_{l/r}}$ eine Ähnlichkeitsverteilung $I_{A,\,l/r}$ generiert. Aus diesen beiden Bildern wird mit (4.29) und (4.30) der maximale Ähnlichkeitswert $A_{max,\,l/r}$ und die Position $\underline{A}_{\,l/r}$ bestimmt. Wenn der Ähnlichkeitswert $A_{max,\,l/r}$ größer als der Schwellwert A_{min} ist, gilt das Objekt als wiedererkannt. Dann kann aus den korrespondierenden Punkten $\underline{A}_{\,l}$ und $\underline{A}_{\,r}$ die neue 3D-Objektposition $\mathbf{Q}'$ berechnet werden.

Wenn die 3D-Objektposition $\mathbf{Q}$ ungefähr bekannt ist, kann der Suchbereich im Bild gezielt eingeschränkt werden. Der Punkt $\mathbf{Q}$ kann mit der Projektionsmatrix $\tilde{\mathbf{P}}$ in die Bildebene projeziert werden (4.37). Die projezierten Bildpunkte sind $\mathbf{Q}''_{l/r}$. Die Größe des Suchbereiches sollte abhängig sein von der Größe des zu suchenden Objektes. Die Größe des zu suchenden Objektes wurde mit den Parametern L_x und L_y vom Jetgitter versucht anzupassen. Weitere wichtige Parameter für die Ausdehnung des Jetgitters sind die Parameter Δx und Δy. Somit können diese vier Parameter zum Erstellen des Suchbereiches verwendet werden. Die Ausdehnung des Suchbereiches wird mit dem Aufpunkt $\mathbf{Q}''_{l/r}$ und $\mathbf{d}$ beschrieben. Die linke untere Ecke des Rechtecks ergibt sich aus $\mathbf{Q}''_{l/r} - \mathbf{d}$. Die rechte obere Ecke des Rechtecks ist $\mathbf{Q}''_{l/r} + \mathbf{d}$. Der Vektor $\mathbf{d}$ wird wie folgt beschrieben:

$$\mathbf{d} = \begin{pmatrix} d_x \\ d_y \end{pmatrix} + d_m \begin{pmatrix} L_x\,\Delta x \\ L_y\,\Delta y \end{pmatrix} \tag{4.39}$$

Als Standardparameter werden $d_x = d_y = 40$ und $d_m = 4$ verwendet.

Abbildung 33: Linkes und rechtes Bild $I_{A,\,l/r}$ der Ähnlichkeitsverteilung mit eingeschränktem Suchbereich

4.5 Visuelle Erkennung des Greifers

Die Hardware des Gesamtsystems besteht aus einem aktiven Binokularkopf und einem Roboterarm mit einem Greifer als Manipulator. Der Aufbau und die entwickelten Module des Kamerasystems wurden in Kapitel 3 und 4 beschrieben. Die Interaktionen des Systems mit dem Roboterarm folgen im Kapitel 5. Dort wird auch die Transformation des Roboterkoordinatensystems in das Weltkoordinatensystem beschrieben.

Das Gesamtsystem kennt jeweils die Position des Roboterarms und des Greifers. Die Genauigkeit dieser Positionsschätzungen ist von der Güte der durchgeführten Kalibration der beiden Koordinatensysteme abhängig. Um eine Güte hierfür zu ermitteln, ist es von Vorteil, wenn das System die Position des Greifers visuell bestimmen kann. Die visuell bestimmte Position muß nicht mit der "gewußten" Position übereinstimmen. Wenn die Greiferposition sich visuell ermitteln läßt, ergeben sich weitere Handlungsmöglichkeiten für das Gesamtsystem. Zwei Möglichkeiten: Das visuell gesteuerte Heranführen eines Greifers zu einem Objekt und eine Kalibration mittels der visuellen Erkennung eines Greifers werden in Kapitel 5 beschrieben. Dieser Abschnitt beschäftigt sich mit der visuellen Erkennung des Greifers.

Um den Greifer visuell zu erkennen, wird das Verfahren zum Wiedererkennen von Objekten aus Abschnitt 4.4 verwendet. Um dieses Modul verwenden zu können, muß dem System bereits eine rudimentäre Objektrepräsentation vom Greifer bekannt sein. Dazu werden von verschiedenen Posen des Greifers eine Menge $\mathcal{O}$ von 2D-Objektrepräsentationen (Abschnitt 4.2.2) erstellt (siehe unten). Jede einzelne Objektrepräsentation aus $\mathcal{O}$ wird mit dem Wiedererkennungsverfahren 4.4 im aktuellen Bild I gesucht. Auf diese Weise erhält das System eine Menge $\mathcal{A}$ von Relationen zwischen Ähnlichkeitswert A^i_{max} und zugehöriger Position $\underline{A}^i$ bezüglich der i-ten Objektrepräsentationen aus $\mathcal{O}$ (4.40).

$$\mathcal{A} = \{(A^i_{max}, \underline{A}^i)| \quad 0 < i \leq |\mathcal{O}|\} \tag{4.40}$$

Die Objektrepräsentation mit dem höchsten Ähnlichkeitswert A^i_{max} gilt als erkannt. Für die weitere 3D-Positionsberechnung wird die zum Ähnlichkeitswert A^i_{max} gehörende 2D-Position $\underline{A}^i$ verwendet.

Wenn dem System eine Transformation der Koordinatensysteme Roboter-Kamera von einer bestimmten Güte bekannt ist, kann die Zuverlässigkeit der Erkennung des Greifers weiter optimiert werden. Aus den Bildinformationen des Greifers läßt sich die 3D-Position $\mathbf{B}$ berechnen. Gemäß der Systeminformationen (Transformation Koordinatensystem Roboter-Kamera) befindet sich der Greifer an der Position $\mathbf{Q}$. Wenn diese Transformation eine bestimmte Güte hat, sollte der Abstand **Q** zu **B** unterhalb einer Schranke D sein (4.41).

$$\|\mathbf{Q} - \mathbf{B}\|_2 < D \tag{4.41}$$

Durch den Grenzwert D stimmen die visuell geschätzten Greiferpositionen eher mit den tatsächlichen Positionen überein. Auf diese Weise werden vor allem Ausreißer mit falschen Schätzungen für weitere Berechnungen nicht verwendet.

Um die Menge $\mathcal{O}$ von Objektrepräsentationen zu erhalten, ist ein gesonderter Arbeitsschritt notwendig. Hierbei fährt der Greifer mehrere 3D-Positionen an. Das Kamerasystem fixiert jeweils den Greifer und speichert die entsprechenden Bilder von den verschiedenen Posen (siehe Abb. 35). Die Abbildung 34 zeigt einen vergrößerten Bildausschnitt vom Greifer. Der Greifer ist mit einem schwarzem Punkt in der Mitte markiert. Die Position dieser Marke wird von allen Bildern jeweils manuell bestimmt. Diese Marke wird für den Aufpunkt der rudimentären Objektrepräsentation verwendet. Damit alle Objektrepräsentationen für die Wiedererkennung genutzt werden können, muß der Aufpunkt zum gleichen 3D-Punkt des Greifers gehören.

Abbildung 34: Greifer

Wenn die Positionen der Aufpunkte bekannt sind, können die rudimentären Objektrepräsentationen (siehe Abb. 4.2.2) von den verschiedenen Posen erstellt werden. Da der Aufbau und das Aussehen des Greifers bekannt sind, können die Parameter für die Jets besser angepaßt werden. Folgende Parameterwerte werden verwendet:

$$d_{max} = 4, l_{max} = 2, \Delta x = \Delta y = 4, L_{x/y} = 5 \tag{4.42}$$

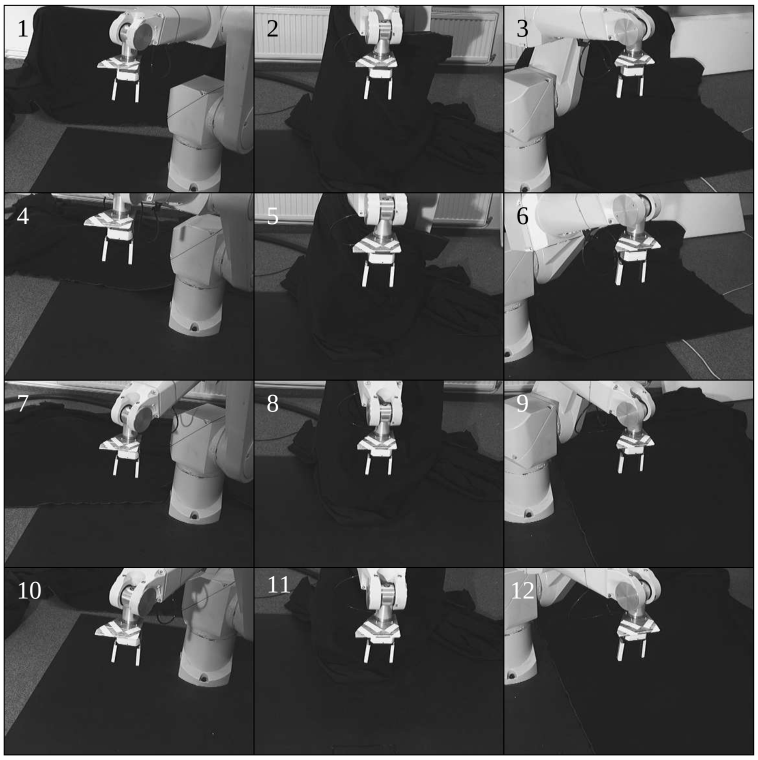


Abbildung 35: Bilder von den verschieden Posen, der linken Kamera

5 Integration von Kompetenzen zu zwei komplexeren Kompetenzen: Aufmerksamkeit und Kalibration im PAC

In den beiden vorherigen Kapiteln wurde ein Stereokamerakopfsystem mit den dazu entwickelten Modulen vorgestellt. In diesem Kapitel werden die aktorischen Module eines Roboterarms mit einem Zweifinger-Greifer mit dem Kamerasystem in ein Gesamtsystem integriert. Der Roboterarm wird in Abschnitt 5.1 näher beschrieben.

Viele dieser Möglichkeiten sind auf eine geeignete Transformation der Koordinatensysteme Roboter und Kamerakopf ineinander angewiesen. Eine Möglichkeit, um eine Approximation einer solchen Transformation zu berechnen, wird in Abschnitt 5.1.1 vorgestellt.

In den Abschnitten 5.2 und 5.3 werden zwei neue komplexe Kompetenzen des Gesamtsystems beschrieben, in denen die Module des Kamerasystems mit denen des Roboterarms integriert werden. In 5.2 wird vorgestellt, wie durch die Interaktion von Kamera und Roboter ein neues Objekt in der Szene wahrgenommen wird und der Greifer durch die Wahrnehmung der Kameras zum Objekt geführt wird. In 5.3 wird ein Verfahren zur iterativen Approximation der Transformationsparameter des Roboterkoordinatensystems in das Kamerakopfkoordinatensystem vorgestellt. Ein Vorteil dieses Verfahrens besteht darin, daß es während anderer Handlungssequenzen durchgeführt werden kann.

5.1 Roboterarm RX90

Der Roboter RX90 der Firma Stäubli besitzt sechs Drehgelenke, mit denen der Manipulator fast beliebig positioniert und orientiert werden kann [30, 31]. Zur Positionierung des Manipulators müssen lediglich die gewünschte Orientierung und die Koordinaten an die steuernde Peripherie übermittelt werden.

Zur Positionierung gibt es zwei Koordinatensysteme: Das Roboterkoordinatensystem und das Werkzeugkoordinantensystem. Beide Koordinatensysteme sind in der Abbildung 36 dargestellt. Ein Raumpunkt wird immer durch sechs Parameter X, Y, Z, y, p, r dargestellt. Die Werte X, Y, Z werden in mm angegeben. Sie geben die Entfernung zwischen Mittelpunkt des Werkzeugflansches und dem Ursprung des Roboterkoordinatensystems an. Diese Parameter stellen die Roboterposition dar. Die Werte y, p, r werden in Grad ausgedrückt und geben die Orientierung des Werkzeugkoordinatensystems an. Das werkzeugbezogene Bezugssystem besitzt die gleiche Orientierung wie das Robotersystem. Der

Parameter y repräsentiert eine Drehung um die Z-Achse des Werkzeugssystems. Der Wert p definiert die Rotation um die Y-Achse des Werkzeugssystems, die im Anschluß an die erste Drehung erfolgt. Der Parameter r beschreibt die dritte Rotation, welche um die Z-Achse des Werkzeugssystems erfolgt. Der Roboter hat eine Wiederholgenauigkeit von ± 0.02 mm.

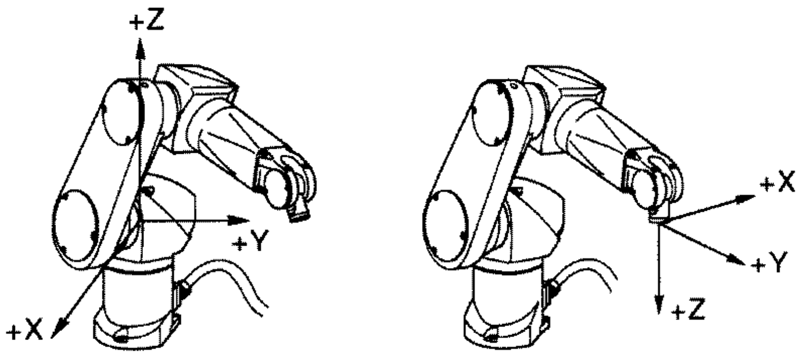


Abbildung 36: Koordinatensysteme des Roboters RX90: Links ist das Roboterkoordinantensystem und rechts das Werkzeugkoordinatensystem.

5.1.1 Koordination des Roboterkoordinatensystems mit dem Weltkoordinatensystem

Damit das Kamerasystem mit dem Roboterarm interagieren kann, sollte die Position und die Orientierung des Roboterkoordinatensystems im Weltkoordinatensystem bekannt sein. Dann können alle Positionsangaben, die aus dem visuellen System gewonnen werden, leicht auf das Roboterkoordinatensystem übertragen werden.

Ein Verfahren zur Kalibration der Kamera- und des Roboterkoordinatensystems wurde in [9] entwickelt und benutzt. Bei diesem Verfahren werden eine Menge von 2D-3D–Korrespondenzen erstellt. Aus dieser Menge können mit einem Verfahren aus Abschnitt 3.2 die Abbildungsparameter berechnet werden. Aus den Abbildungsparametern können widerum die externen Parameter berechnet werden. Jetzt ist die Transformation vom Roboterkoordinatensystem zum linken/rechten Kamerakoordinatensystem bekannt. Da die Transformation vom linken/rechten Kamerakoordinatensystem zum Weltkoordinatensystem berechnet werden kann (siehe Abschnitt 3.2.2), kann dadurch die Transformation vom Roboterkoordinatensystem zum Weltkoordinatensystem berechnet werden.

5.2 Körperlicher Kontakt durch Integration verschiedener Kompetenzen im PAC

Um mit dem Greifer einen körperlichen Kontakt zu einem neuen Objekt in der Szene herzustellen, werden viele Kompetenzen miteinander verknüpft. Ein Ablaufdiagramm ist in Abbildung 37 dargestellt.

Als erstes erkennt das System mit seiner Aufmerksamkeitssteuerung (Abschnitt 4.1), ob sich die Szene verändert. Wenn das System eine Veränderung festgestellt hat, wird mit einem Verfahren eine Vermutung aufgestellt, ob ein neues Objekt in der Szene enthalten ist (Abschnitt 4.2). Wenn diese Vermutung positiv ausfällt, wird das Objekt von den Kameras fixiert (Abschnitt 4.3). Hierbei wird eine 2D-Objektrepräsentation von Gaborjets vom Objekt erstellt. Diese wird unter anderem für die Objektwiedererkennung (Abschnitt 4.4) nach der Fixierung benötigt. Das Objekt wird im Bild wiedererkannt, um die 3D-Position aus diesem Blickwinkel zu berechnen.

Diese berechnete 3D-Position ist durch die insgesamt ungenaue, absolute Tiefenberechnung nur eine Schätzung. Sie bietet aber eine gute Ausgangslage für eine erste Bewegung des Robotergreifers. Nach dieser Annäherung des Greifers zum Objekt kann der Greifer visuell erkannt werden (Abschnitt 4.5). Aus den Daten der visuellen Wiedererkennung läßt sich eine 3D-Position berechnen. Nun kann eine relative Tiefenschätzung zwischen Objekt und Greifer erfolgen. Dieser relative Abstandsvektor zwischen Objekt und Greifer wird für die nächste Roboterbewegung verwendet. Dies führt zu einer weiteren Annäherung und kann solange wiederholt werden, bis ein Kontakt zwischen Greifer und Objekt entstanden ist. Dieser Kontakt könnte visuell oder taktil überprüft werden. Der Erfolg einer Berührung wird in diesem Handlungsmodul nicht überprüft. Sie könnte durch andere Handlungsmodule erfolgen. Der Erfolg der Berührung könnte z.B. mittels taktiler Sensoren (siehe Abb. 8) überprüft werden.

Eine weitere Möglichkeit zur näheren Untersuchung des Kontakts ermöglicht die Verwendung des Zooms der Kameras. Da die Kameras das Objekt fixiert haben, ist im gezoomten Bild sowohl das Objekt als auch der Greifer. Da im herangezoomten Bild ein Teil der Szene stark vergrößert ist, werden Bildmerkmale besser abgebildet, z.B. können Kanten oder Ecken besser erkannt werden. Es können also Bildverarbeitungsverfahren zum Teil besser angewendet werden als auf ein Bild der gesamten Szene.

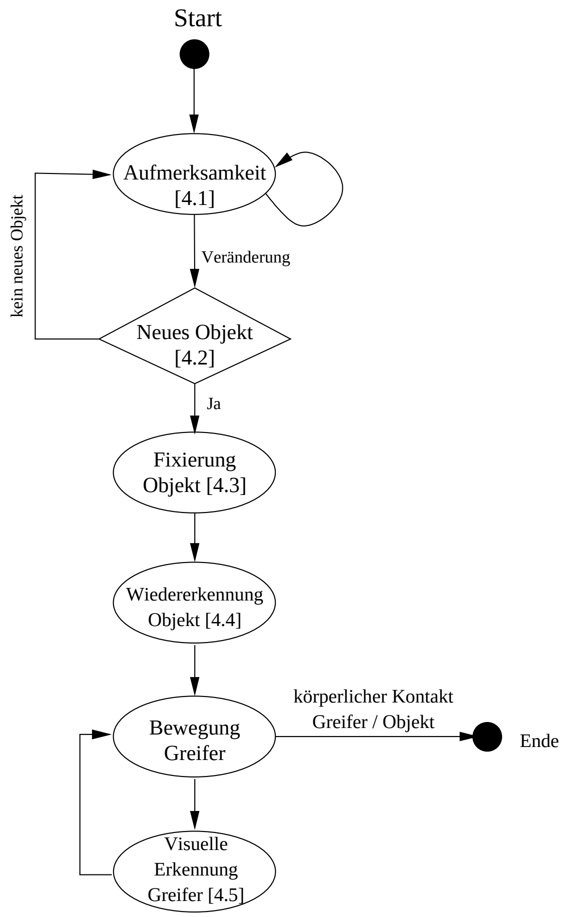


Abbildung 37: Ablaufdiagramm zum Erlangen des körperlichen Kontakts. In den eckigen Klammern stehen jeweils die Abschnittsnummern, in denen die aufgeführten Module beschrieben worden sind.

5.2.1 Experiment

Im folgenden wird der Verlauf und die Ergebnisse von einem Experiment exemplarisch beschrieben. Die Abbildungen 38 (linke Kamera) und 39 (rechte Kamera) wurden im Verlauf des Experiments mit dem Kamerakopf aufgenommen. Während des Experimentes wurden mehrere 3D-Positionen berechnet. Diese werden jeweils in das entsprechende Teilbild projeziert und dort mit einem schwarzen Kreuz markiert.

Als Ausgangslage für das Experiment betrachtet der Kamerakopf die Szene (Teilbild a)) und wartet auf eine Veränderung. Eine Veränderung der Szene wird durch eine Person ausgelöst, die in die Szene geht und ein Objekt in die Szene legt. Danach verläßt die Person die Szene. Während dieser Aktion stellt das System eine Veränderung in der Szene fest. Nachdem die Szene sich nicht mehr

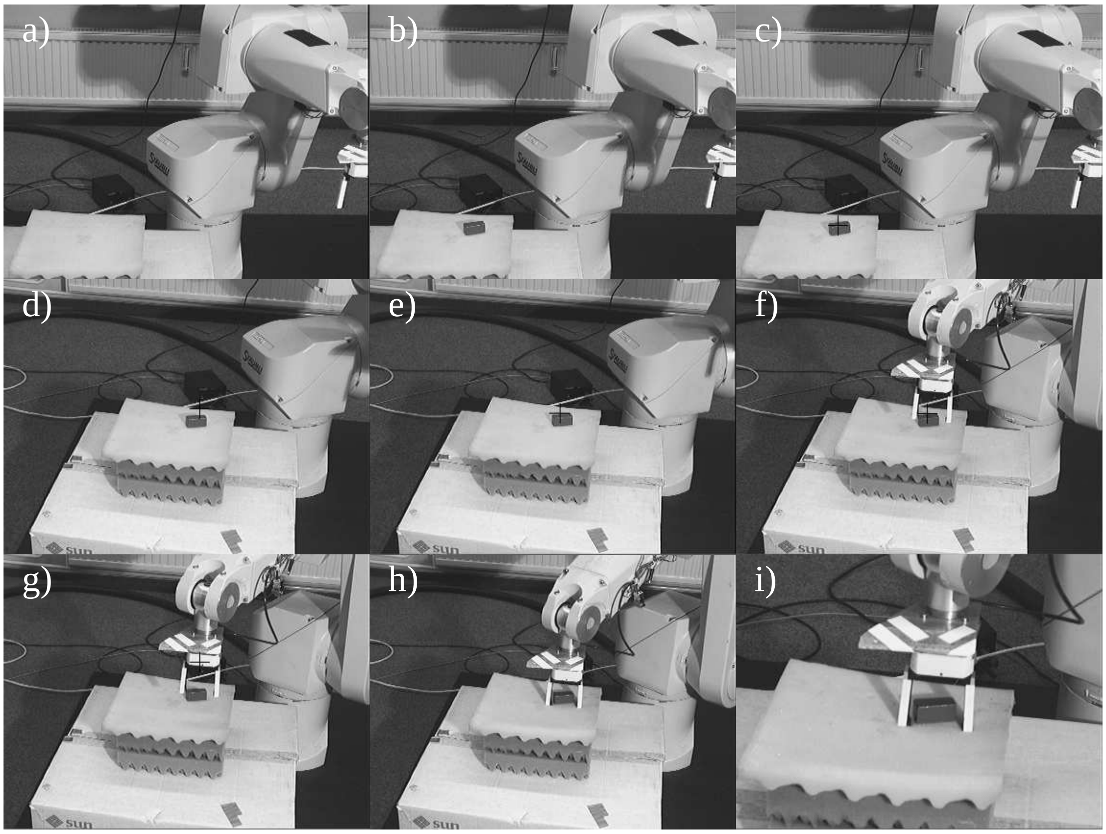


Abbildung 38: Bilder der linken Kamera vom Experiment aus Abschnitt 5.2

verändert (Teilbild b)), lokalisiert das System ein neues Objekt in der Szene mit der Position $\mathbf{Q} = (-189, 678, 1249)^T$. Im Teilbild c) ist die 3D-Objektposition $\mathbf{Q}$ ins Bild projeziert worden. Die Projektion stimmt mit dem Objekt gut überein. Im nächsten Schritt wird das Objekt fixiert (Teilbild d)). Die Makierung für die geschätzte Objektposition $\mathbf{Q}$ stimmt jetzt nicht mehr mit dem Objekt überein. Nach der Fixierung wird das Objekt im Bild wiedererkannt, um die Position neu zu bestimmen. Die neue 3D-Positionsschätzung ist $\mathbf{Q} = (-203, 710, 1267)^T$ (Teilbild e)). Die Projektion von $\mathbf{Q}$ stimmt jetzt mit dem Objekt besser überein.

Der Greifer wird als erste Annäherung zur Position $\mathbf{G} = \mathbf{Q}$ bewegt (Teilbild f)). Nach der Roboteraktion wird der Greifer im Bild visuell detektiert (Teilbild f)). Die 3D-Positionschätzung des Greifers nach den visuellen Daten ist $\mathbf{G}_v = (-185, 578, 1144)^T$. Aus den Vektoren $\mathbf{Q}$ und $\mathbf{G}_v$ kann ein Abstandsvektor $\mathbf{D} = \mathbf{Q} - \mathbf{G}_v$ berechnet werden, so daß der Greifer zu einer neuen Position $\mathbf{G} + \mathbf{D} = (-221, 842, 1390)^T$ bewegt werden kann (Teilbild h)). Wie im gezoomten Teilbild i) gesehen werden kann, hat der Greifer erfolgreich die Position des Objektes angefahren.

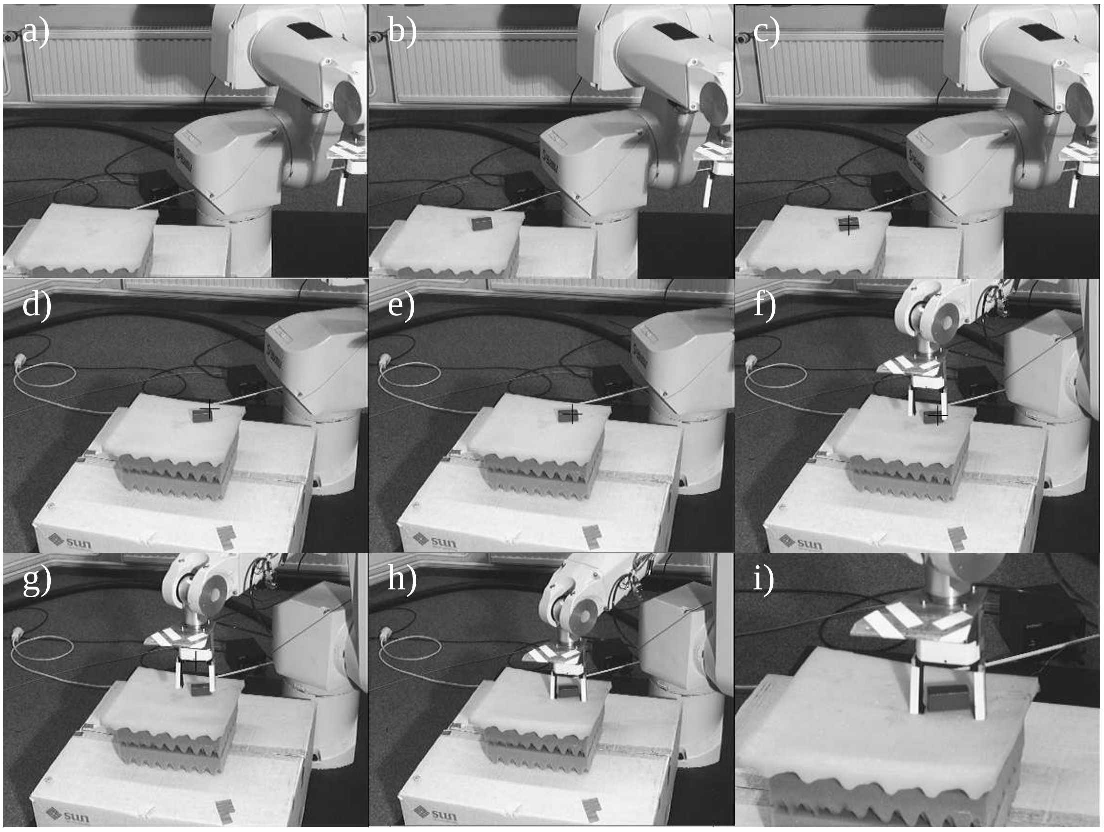


Abbildung 39: Bilder der rechten Kamera vom Experiment aus Abschnitt 5.2

5.3 Kalibration im PAC

Für die Interaktion des Roboters mit dem Kamerasystem ist die Transformation der beiden Koordinatensysteme (Roboter und Welt) ineinander wichtig. Um eine Approximation der Transformationsparameter zu erhalten, wurde in Abschnitt 5.1.1 bereits ein Verfahren vorgestellt. Der wesentliche Nachteil des Verfahrens besteht darin, daß es nur unter *sehr künstlichen* Bedingungen funktioniert und daß es nicht während anderer Handlungen oder Handlungssequenzen verwendet werden kann.

In diesem Abschnitt wird hierfür ein weiteres Verfahren vorgestellt. Die wesentlichen Merkmale des Verfahren sind: Visuelle Erkennung des Greifers (Abschnitt 4.5), Verwendung des Verfahrens innerhalb eines *normalen Betriebes* (z.B. keine Abdunkelung), Gütekontrolle der Transformation und eine iterative Approximation der Transformationsparameter.

Die Kalibration der Transformationsparameter kann als eigenstängige Handlung ablaufen oder innerhalb einer anderen Handlung eingebettet sein. Für die Kalibration ist es nur entscheidend, daß der Roboterarm und somit der Greifer möglichst viel bewegt wird. Innerhalb eines Experiments hat der Greifer Zufallspositionen (Positionen im Roboterkoordinatensystem) angefahren (siehe Abschnitt 5.3.4). Die Transformationsparameter müssen am Anfang für den Algorithmus nicht bekannt oder korrekt sein. Nach dem ersten Iterationsschritt besitzt das System eine erste Approximation von den Parametern. Wenn eine Schätzung der Parameter bereits vorhanden ist, berechnet das System automatisch einen Qualitätswert der Transformationsdaten (Abschnitt 5.3.2). In dessen Abhängigkeit werden die neuen Transformationsdaten für den nächsten Iterationsschritt berechnet. Im folgenden wird zunächst ein einzelner Iterationsschritt in 5.3.1 erläutert, dann die Rückkopplung der Parameter in 5.3.2 und anschließend der gesamte Algorithmus in 5.3.3. Zum Abschluß werden in 5.3.4 experimentelle Ergebnisse vorgestellt.

5.3.1 Ein Iterationsschritt

Während eines Iterationsschritts muß der Blickwinkel des Kamerakopfes unverändert bleiben. Der Ablauf in einem Iterationsschritt ist in Abbildung 41 dargestellt. Innerhalb eines Iterationsschritts wird eine Liste L von 2D-3D–Korrespondenzen aufgebaut. Nach einer Bewegung des Roboterarms versucht das System den Greifer visuell mit dem Verfahren aus Abschnitt 4.5 zu erkennen (siehe Abb. 40). Wenn die Erkennung erfolgreich war, wird der Liste ein Tupel 2D-3D Koordinaten hinzugefügt. Anschließend wartet das System auf die nächste Bewegung des Greifers.

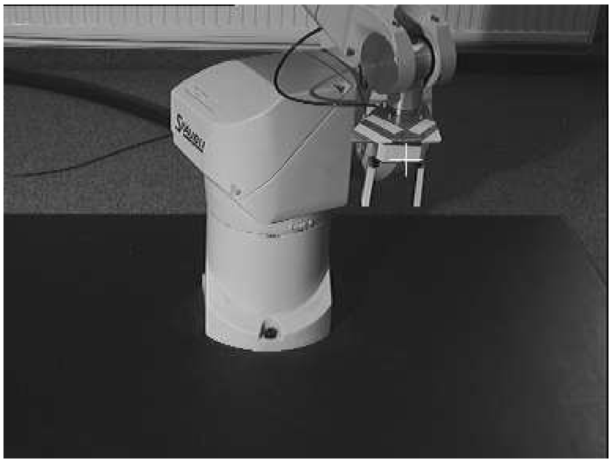


Abbildung 40: Beispielbild, welches innerhalb einer Iteration aufgenommen worden ist. Der Greifer wurde erfolgreich erkannt.

Wenn die Liste genügend groß ist, kann aus der Liste ein Gleichungssystem erstellt werden. Das Erstellen und Lösen des Gleichungssystems entspricht dem

Gleichungssystem aus Abschnitt 3.2.1. Die Lösung des Gleichungssystems liefern die Abbildungsparameter der Kameras mit dem Roboterkoordinatensystem als Bezugssystem. Aus diesen Abbildungsparametern können wiederum die externen Abbildungsparameter berechnet werden. Die externen Abbildungsparameter sind eine Transformation $\tilde{\mathbf{T}}^i_{R \to K,l/r}$ des Roboterkoordinatensystems in das linke/rechte Kamerakoordinatensytem, wobei i die bisherige Iterationszahl angibt.

Da die Transformation $\tilde{\mathbf{T}}_{l/r}$ (3.10) des Kamerakoordinatensystems in das Weltkoordiantensystem bekannt ist (Abschnitt 3.2.2), kann die Transformation $\tilde{\mathbf{T}}^i_{R \to W}$ des Roboterkoordinatensystems in das Weltkoordinatensystem berechnet werden. Bei der Berechnung von $\tilde{\mathbf{T}}^i_{R \to W}$ wird zuerst gemäß (5.43) $\tilde{\mathbf{T}}^i_{R \to W,l}$ und $\tilde{\mathbf{T}}^i_{R \to W,r}$ berechnet.

$$\tilde{\mathbf{T}}^i_{R \to W,l} = \tilde{\mathbf{T}}_l \tilde{\mathbf{T}}^i_{R \to K,l} \qquad \tilde{\mathbf{T}}^i_{R \to W,r} = \tilde{\mathbf{T}}_r \tilde{\mathbf{T}}^i_{R \to K,r} \tag{5.43}$$

Algebraisch betrachtet müßten $\tilde{\mathbf{T}}^i_{R \to W,l}$ und $\tilde{\mathbf{T}}^i_{R \to W,r}$ unabhängig von der linken oder rechten Kamera die gleiche Transformation beschreiben. Es müßte (5.44) gelten.

$$\tilde{\mathbf{T}}^i_{R \to W} = \tilde{\mathbf{T}}^i_{R \to W,l} = \tilde{\mathbf{T}}^i_{R \to W,r} \tag{5.44}$$

Durch Ungenauigkeiten der Meßwerte und der Systemparameter gilt die Gleichheit in (5.44) nicht exakt. Um $\tilde{\mathbf{T}}^i_{R \to W}$ für das System zu berechnen, werden jeweils die Transformationsparameter (3 Translationsparameter und 3 Rotationsparameter) von $\tilde{\mathbf{T}}^i_{R \to W,l}$ und $\tilde{\mathbf{T}}^i_{R \to W,r}$ gemittelt (5.45).

$$\tilde{\mathbf{T}}^i_{R \to W} = \frac{\tilde{\mathbf{T}}^i_{R \to W,l} + \tilde{\mathbf{T}}^i_{R \to W,r}}{2} \tag{5.45}$$

Wenn vor einem Iterationsschritt dem System eine Approximation der Transformationsparameter bekannt ist, kann aus der 2D-3D-Korrespondenzliste L mit (5.46) ein Fehlerwert F_i berechnet werden. Ein 3D-Punkt $\mathbf{M} \in L$ kann mit der Gleichung (3.4) ins Bild projeziert werden. Der Punkt $\mathbf{M}$ wird auf den Bildpunkt $\mathbf{m}_p$ abgebildet. Als nächstes wird der zum 3D-Punkt zugehörige Bildpunkt $\mathbf{m}$ mit dem projezierten Bildpunkt $\mathbf{m}_p$ in der euklidischen Norm verglichen. Als Fehlerwert wird die euklidische Norm der Differenz der beiden Bildpunkte genommen. Diese Berechnung wird für alle Punkte in der Liste L durchgeführt.

$$F_i = \frac{1}{|L|} \sum_{\mathbf{M} \in L} \|\mathbf{m}_p - \mathbf{m}\| \tag{5.46}$$

Aus der Verteilung der Fehler der linken und rechten Kamera wird ein Mittelwert F_i berechnet. In den Grafiken von Abschnitt 5.3.4 wird der Verlauf der

Mittelwerte und der Varianzen für mehrere Iterationsschritte von einem Experiment gezeigt. Der Mittelwert F_i des Fehlers wird im nächsten Abschnitt zur Approximation der Abbildungsparameter verwendet.

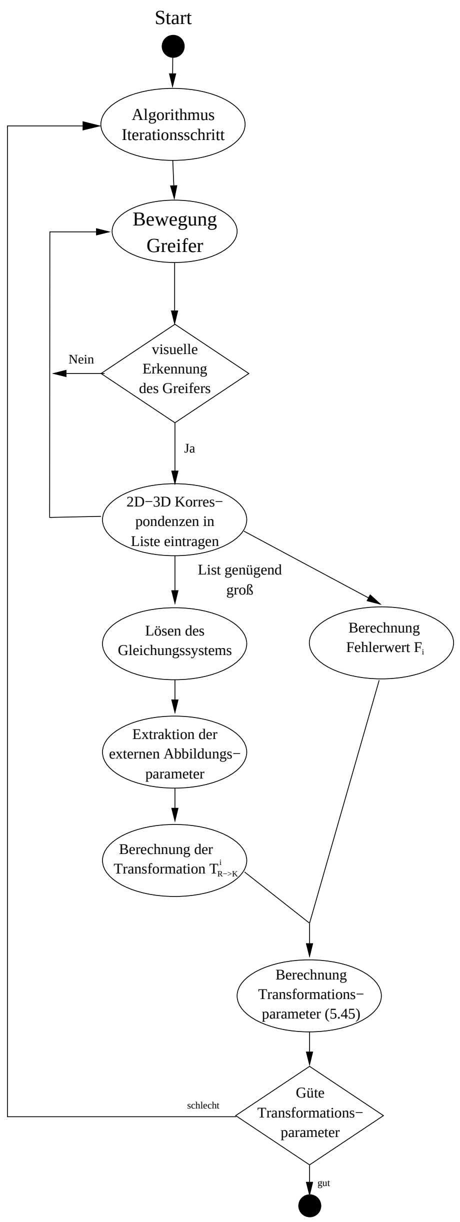


Abbildung 41: Ablauf in einem Iterationsschritt

5.3.2 Rückkopplung der Transformationsparameter

Aus einem Iterationsschritt erhält das System einerseits mit F_i einen Gütewert der bisherigen Transformationsparameter $\tilde{\mathbf{T}}^{alt}_{R\to W}$ und andererseits neue Transformationsparameter $\tilde{\mathbf{T}}^{i}_{R\to W}$. Mit der Gleichung (5.48) werden die einzelnen Transformationsparameter (drei Rotationsparameter und drei Translationsparameter) von $\tilde{\mathbf{T}}^{neu}_{R\to W}$ neu berechnet. In dieser Gleichung wird der Fehlerwert F_i mit der positiven Konstanten c mit (5.47) auf den Wert E_i normiert.

$$E_i = \min(\frac{F_i}{c}, 1) \geq 0 \quad \text{mit} \quad F_i \geq 0 \tag{5.47}$$

$$T^{neu} = (1 - E_i)T^{alt} + E_i T^i \tag{5.48}$$

Falls dem System die bisherigen Transformationsparameter $\tilde{\mathbf{T}}^{alt}_{R\to W}$ nicht bekannt sind (z.B. im ersten Iterationsschritt), kann kein Fehlerwert F_i berechnet werden. Für diesen Fall wird $F_i = c$ gesetzt. Im Experiment von 5.3.4 wurde $c = 150$ Pixel gewählt.

Bei einem hohen Fehlerwert werden durch (5.48) die im Iterationsschritt gewonnenen Transformationsparameter stark gewichtet. Wenn der Fehlerwert wiederum niederig ist, sind die vorhandenen Transformationsparameter T^{alt} sehr gut und der Anteil der innerhalb des Iterationsschritts berechneteten Parameter T^i in die neuen Parameter T^{neu} ist gering.

5.3.3 Gesamtalgorithmus

Der Gesamtalgorithmus zur Kalibrierung der Transformation im Wahrnehmungs-Handlungs-Zyklus besteht in der Hintereinanderausführung der einzelnen Iterationsschritte. Nach jedem Iterationsschritt wird eine neue Approximation der Transformationsparameter berechnet. Die Berechnung ist von der Güte der bisherigen Transformation abhängig. Prinzipiell müßte der Algorithmus zur Approximation nie terminieren. Allerdings ist die Durchführung so lange wichtig, so lange die Güte eine bestimmte Qualität nicht erreicht hat. Die benötigte Qualität ist wiederum abhängig von der geplanten Aktion. Die verschiedenen Prioritäten dieses Algorithmuses sind im Gesamtzusammenhang in Tabelle 1 in Abschnitt 2.1.1 bereits angesprochen.

5.3.4 Experimentauswertung

In diesem Abschnitt werden die Ergebnisse eines Experiments vom obigen Algorithmus vorgestellt. Innerhalb des Experiments fährt der Roboter je Iterationsschritt soviele zufällige Positionen an, daß die Liste L genügend Einträge

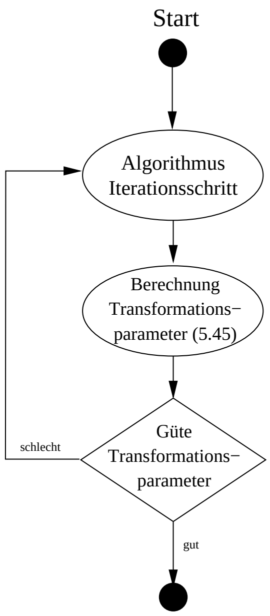


Abbildung 42: Gesamtablauf der Iteration

zum Lösen eines überbestimmten Gleichungssystems gemäß Abschnitt 3.2.1 hat. Die angefahrenen Positionen können sich auch außerhalb des Bildbereiches der Kamera befinden.

Vor dem Experiment wurde eine Bestimmung der Transformationsparameter mit dem Verfahren aus [9] durchgeführt. Diese Daten werden in den Tabellen 3 mit $T_{Orig.}$ als *originale* Transformationsparameter bezeichnet. Der Translationsvektor wurde im Experiment um 35 cm auf der X-Achse verschoben. Diese veränderten Daten wurden dem System als T_{Start} verfälschtes Vorwissen mitgegeben. Diese Veränderung hätte z.B. durch ein Verschieben der Roboter- und Kameraeinheit passieren können. Das Experiment soll zeigen, inwieweit das System den Fehler feststellt und diesen Fehler korrigieren kann.

In der Tabelle 3 sind die Veränderungen der Transformationsparameter und Rotationsparameter im Laufe der Iterations aufgelistet. Vor dem Iterationsschritt i sind dem System die Parameter $T^{i-1,neu}$ bekannt. Innerhalb eines Schrittes werden die Transformationsparameter T^i berechnet. Gleichzeitig wird die Güte von den Parametern $T^{i-1,neu}$ berechnet. Der normierte Fehlerwert E_i wird gemäß (5.48) ermittelt. Aus $T^{i-1,neu}$, T^i und E_i werden mit (5.47) die neuen Transformationsparameter $T^{i,neu}$ berechnet.

In den Abbildungen 43 und 44 sind jeweils für die linke und rechte Kamera einmal der Mittelwert F_i und die Varianz der Fehlerfunktion von F_i aufgetragen. In der Abbildung 45 ist der Wert E_i aufgetragen.

Im ersten Iterationsschritt kann in der Abbildung 43 gesehen werden, daß der

i	T	X-Achse	Y-Achse	Z-Achse	Rotat. Achse Z	Rotat. Achse Y	Rotat. Achse Z	E_i
	$T^{Orig.}$	-39.59	737.06	1829.56	3.12	1.61	-1.58	
0	T^{Start}	310.40	737.06	1829.56	3.12	1.61	-1.58	
	T^{0}	47.28	845.16	1776	-3.12	1.57	-1.56	1.0
1	$T^{1,neu}$	47.28	845.16	1776	3.12	1.61	-1.58	
	T^{1}	218.31	579.79	1684.41	-3.04	1.65	-1.51	0.40
2	$T^{2,neu}$	115.80	738.84	1739.31	3.17	1.63	-1.55	
	T^{2}	-106.90	613.20	1541.92	3.08	1.62	-1.59	0.50
3	$T^{3,neu}$	5.17	676.43	1641.26	-3.15	1.62	-1.57	
	T^{3}	-12.03	951.73	2023.32	3.13	1.58	-1.56	0.23
4	$T^{4,neu}$	1.13	741.10	1731.01	3.12	1.61	-1.57	
	T^{4}	5.92	593.12	1879.66	3.13	1.69	-1.57	0.15
5	$T^{5,neu}$	1.86	718.53	1753.68	3.13	1.63	-1.57	
	T^{5}	105.81	603.27	2013.22	-3.11	1.72	-1.55	0.23
6	$T^{6,neu}$	25.40	692.43	1812.46	3.14	1.65	-1.56	
	T^{6}	-86.91	621.37	1975.16	3.10	1.71	-1.59	0.30
7	$T^{7,neu}$	-8.73	670.83	1861.9	3.12	1.67	-1.57	
	T^{7}	-228.32	807.02	1918.02	3.04	1.61	-1.62	0.32
8	$T^{8,neu}$	-79.01	714.42	1879.86	3.09	1.65	-1.59	
	T^{8}	-128.70	757.28	1798.2	3.07	1.61	-1.60	0.22
9	$T^{9,neu}$	-89.96	723.86	1861.87	3.09	1.64	-1.59	

Tabelle 3: Auflistung der Translationsparameter (in Millimeter) und Rotationsparameter (in Bogenmaß) in den verschiedenen Iterationsschritten

Fehler im Bild, also die gesehene Position und die berechnete Position des Greifers in 2D, sehr groß ist. Der Fehler im Bild beträgt über 160 Pixel. Somit ist der normierte Fehlerwert $E_0 = 1$. Die Daten T^{Start} werden in diesem Iterationsschritt durch die gewonnenen Daten T^0 ersetzt. Im nächsten Schritt nimmt der Fehler stark ab. Nach dem zweiten Iterationsschritt steigt einerseits der Fehlermittelwert leicht und andererseits die Varianz stark an. Während sich in den nächsten Iterationsschritten der Fehlermittelwert zwischen einem Wert von 30 und 50 Pixel einpegelt, fällt die Varianz stetig auf ein Minimum ab. Das System hat also den Fehler erkannt und korrigiert.

Die Regel zur Neuberechnung der Parameter (5.48) sollte ein Minimum in einem sechs dimensionalen Suchraum finden. Durch die starke Abnahme der Varianz kann davon ausgegangen werden, daß mindestens ein lokales Minimum gefunden worden ist.

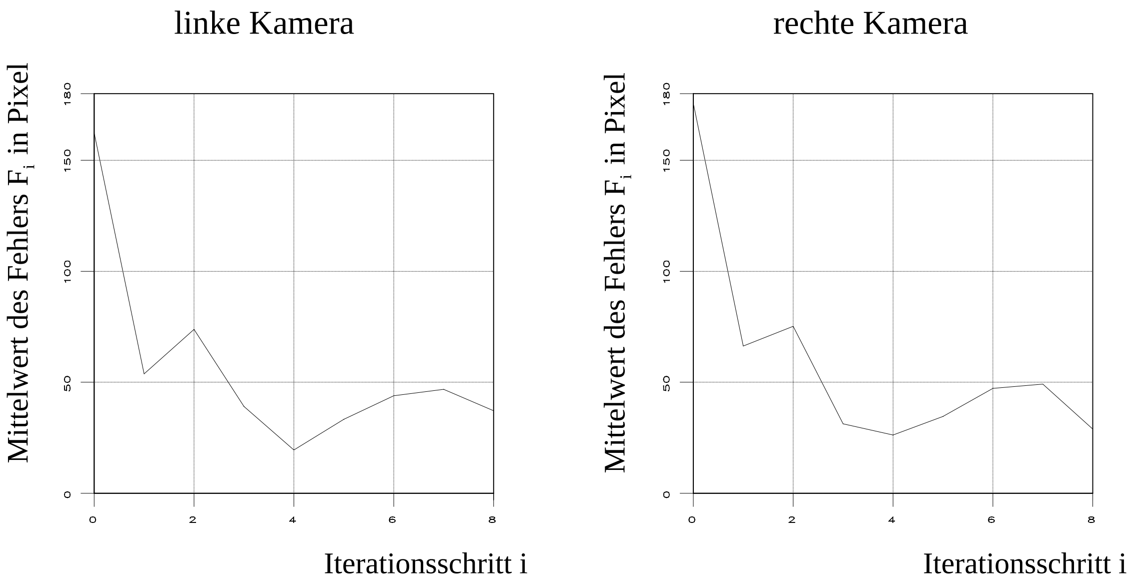


Abbildung 43: Mittelwert des Fehlers F_i je Iterationsschritt von der linken/rechten Kamera

Innerhalb eines Iterationsschrittes wurden mehrere Bilder aufgenommen, die der Abbildung 46 entsprechen. Die erkannte Bildposition des Greifers ist im Bild mit einem weißen Kreuz markiert. In der Abbildung 47 sind alle Bildausschnitte vom Greifer aus dem achten Iterationsschritt, bei denen der Greifer im Bild erkannt wurde, enthalten. Wie in Abbildung 46 sind auch hier die mit dem Algorithmus aus Abschnitt 4.5 erkannten 2D-Positionen vom Greifer jeweils mit einem weißen Kreuz markiert. Auf dem Greifer kann in der Mitte ein schwarzer Punkt erkannt werden. Dieser Punkt wurde in Abschnitt 4.5 als Referenzposition des Greifers genommen. Je dichter der Mittelpunkt des weißen Kreuzes dem

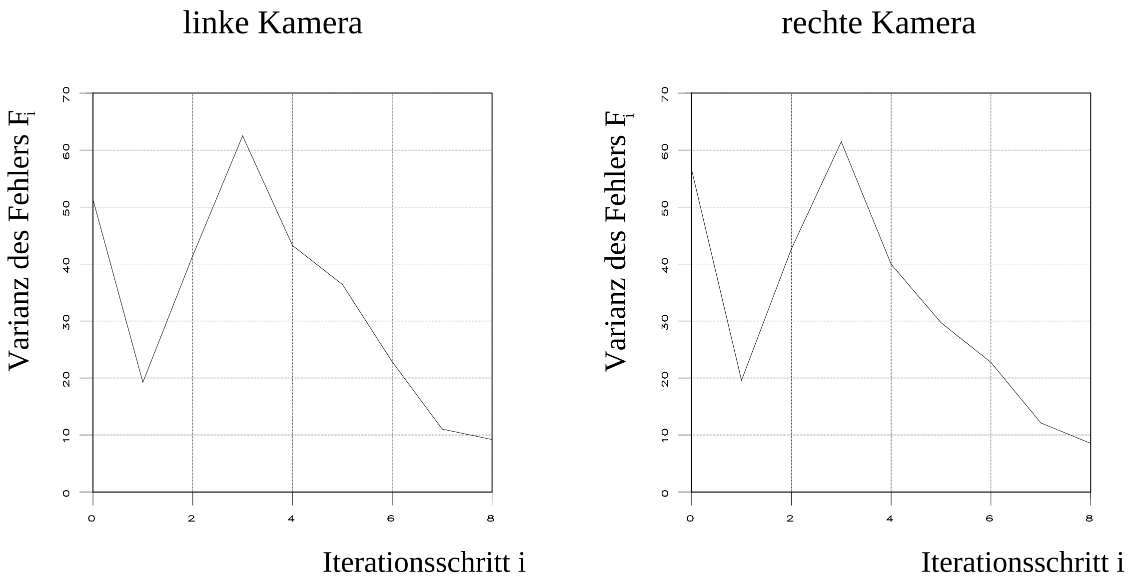


Abbildung 44: Varianz des Fehlers F_i je Iterationsschritt von der linken/rechten Kamera

schwarzem Punkt ist, um so besser hat das System die 2D-Position des Greifers erkannt.

Die Berechnung der 3D-Position des Greifers basiert auf der 2D-Position des Greifers im Bild. Hierbei zeigt sich nun, obwohl die 2D-Position eine Ungenauigkeit von einigen Pixeln haben kann, daß die Iteration insgesamt erfolgreich abläuft und ein Minimum findet. Eine höhere Genauigkeit im Dreidimensionalen könnte durch eine höhere Genauigkeit der visuellen Erkennung im Zweidimensionalen erlangt werden. Eine höhere Genauigkeit im Zweidimensionalen kann durch eine Erhöhung der Anzahl der Bildposen in Abschnitt 4.5 erlangt werden. Es kann natürlich auch ein anderes Verfahren zur Positionsbestimmung des Greifers verwendet werden, welches eventuell andere Merkmale des Greifers extrahiert, als das Verfahren aus Abschnitt 4.5. Der Algorithmus aus Abschnitt 5.3 und die damit entwickelte Software ist so modular entwickelt, daß ein Hinzufügen von anderen Verfahren möglich ist.

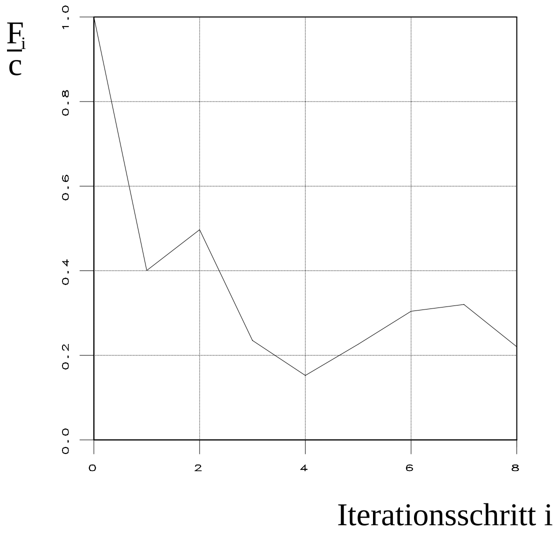


Abbildung 45: Verlauf des normierte Fehlers $E_i = \frac{F_i}{c}$

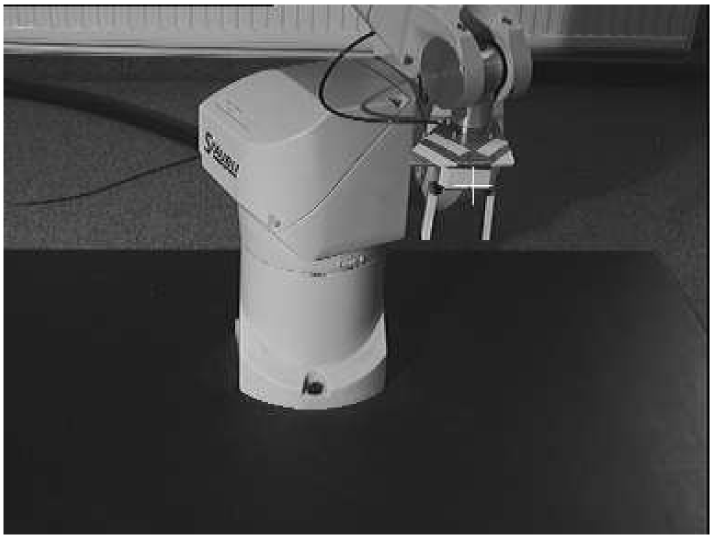


Abbildung 46: Beispielbild, welches innerhalb einer Iteration aufgenommen worden ist. Auf diesen Bilddaten wurden der Algorithmus angewendet.

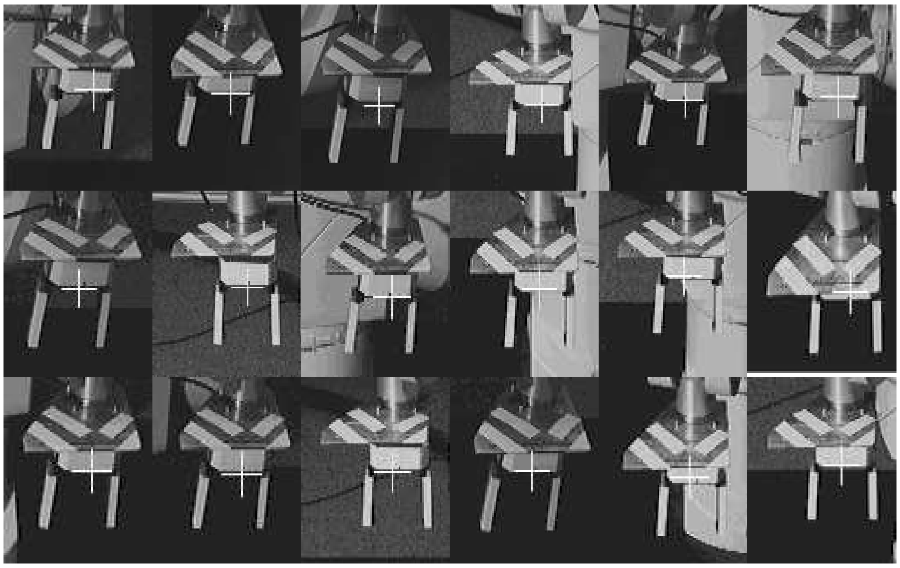


Abbildung 47: Bildausschnitte aus dem achten Iterationsschritt, bei denen der Greifer im Bild erkannt wurde. Die erkannte 2D-Position des Greifers ist jeweils mit einem weißen Kreuz markiert. Die Position sollte der Position des schwarzen Punktes in der Mitte des Greifers entsprechen.

6 Ausblick

In dieser Arbeit wurden zwei komplexe Kompetenzen eines visuell basierten Robotersystems implementiert: Die Herstellung eines körperlichen Kontaktes zu einem neuen Objekt und die Kalibration der Transformationsparameter von Welt– und Robotersystem während Wahrnehmungs-Handlungs-Zyklen ohne Hinzunahme spezieller Kalibrationsmerkmale.

Es wurde gezeigt, daß einfache Kompetenzen erfolgreich zu komplexeren Kompetenzen verbunden werden konnten und eine autonome Berührung neuer Objekte unter Berücksichtigung von gewissen Randbedingungen (z.B. Sicherheit von Personen) erreicht wurde. Weiterhin wurde gezeigt, daß das System die Fähigkeit hat, die Qualität der Roboter-Kamera-Koordination zu bestimmen und einen starken Fehler zu verbessern.

Bei der Erstellung der Diplomarbeit hat sich die Kalibration der Kameras vor allem in Bezug auf verschiedene Linseneinstellungen als ein großes Problem erwiesen. Sicherlich könnte das entwickelte Konzept auf mehrere feste Linseneinstellungen angewandt werden, wobei damit nicht das grundsätzliche Problem gelöst wäre. Weiterhin könnte sich auch die Ungenauigkeit der Ansteuerung der Linsen als Problem erweisen [33]. Ein Ziel sollte es sein, ein Verfahren zu entwickeln, um einen kalibrierten Zustand der Kameras bei variabler Linseneinstellung zu erhalten. Dann könnte das betrachtete Objekt in Abhängigkeit von seiner Größe herangezoomt und mittels bekannter Bildverarbeitungsmethoden gut untersucht werden.

Weiterhin wäre es interessant, die Kalibration der Bezugsysteme Roboter und Kamera noch mehr in den *normalen* Handlungsablauf einzubetten und die Iteration zu beschleunigen. Dabei wäre es zum Beipiel von Vorteil, wenn der Greifer besser verfolgt und dieser verschiedene Orientierungen annehmen könnte. Ein Trackingverfahren, welches auch mit allgemeineren Posen umgehen kann, ist in [29] beschrieben. Ein zusätzliches Ziel sollte es sein, die Genauigkeit der Objektlokalisation zu erhöhen. Durch eine höhere Genauigkeit würden die darauf basierenden Approximationen verbessert werden.

Eine weitere interessante Fortsetzung dieser Arbeit ist die Benutzung eines taktilen Sensors [32], mit dem eine erfolgreiche Berührung des Objektes durch einen vom Sehsystem unabhängigen Sensor festgestellt werden kann. Die Integration der taktilen und visuellen Objektinformation dürfte eine noch vollständigere Objektrepräsentation möglich machen und führt in den Bereich der Sensor Fusion [35].

Ein zentraler Aspekt dieser Arbeit ist die Integration von einfachen Modulen zu komplexeren Kompetenzen, die in enger Verknüpfung von Wahrnehmung und

Handlung miteinander interagieren. Die hier beschriebenen Kompetenzen sind wichtig für ein visuell basiertes Robotersystem: Es sollte sowohl eine stabile Kalibration während des normalen Wahrnehmungs–Handlungs–Zyklus gesichert als auch ein sinnvolles Anfangsverhalten definert sein.

Ein wichtiges Vorhaben weiterer Arbeiten könnte die Entwicklung noch komplexerer Handlungsstrukturen sein. Diese basieren vermutlich noch weniger auf vordefinierten Strukturen, sondern sollten sich im Laufe der Erfahrung entsprechend der dem System gesetzten Ziele entwickeln. Die Bereitstellung mächtiger Basismodule (wie hier beschrieben) ist vermutlich auch für noch komplexere Handlungsstrukturen wesentlich.

Literatur

[1] Marcus Ackermann. Akkumulieren von Objektrepräsentationen im Wahrnehmungs-Handlungs-Zyklus. Diplomarbeit, Christian-Albrechts-Universität zu Kiel, 2000.

[2] M. A. Arbib, P. Erdi und J. Szentagothai. *Structure, Function and Dynamics : An Integrated Approach to Neural Organization.* The MIT Press, 1997.

[3] T.G.R. Bower. The object in the world of the infant. *Scientific American*, 225:30–38, 1971.

[4] Ilja N. Bronstein, Günter Grosche und Eberhard Zeidler. *Teubner – Taschenbuch der Mathematik.* B. G. Teubner, Stuttgart, 1996.

[5] R.A. Brooks. Intelligence without reason. *International Joint Conference on Artificial Intelligence*, pages 569–595, 1991.

[6] J. Bruske, M. Hansen, L. Riehn und G. Sommer. Dynamic Cell Structures for Calibration-free Adaptive Saccade Control of a four-Degrees-of-Freedom binocular Head. Technical report, Christian-Albrechts-Universität zu Kiel, Institut für Informatik und praktische Mathematik, 1996.

[7] J.D. Daugman. Complete discrete 2-d Gabor transforms by neural networks for image analysis and compression. *IEEE Trans. Acoustics, Speech and Signal Processing*, 36:1169–1179, 1988.

[8] Kord Ehmcke. Eine Strategie zur Haptischen Exploration von Objekten. Diplomarbeit, Christian-Albrechts-Universität zu Kiel, In Arbeit.

[9] Kord Ehmcke und Oliver Granert. Lernen von Objektrepräsentationen im Action-Perception-Cycle. Fortgeschritten Praktikum Computer Vision II, 2000.

[10] Jörg Ernst. Kalibrierung der internen Parameter einer aktiven Kamera. Diplomarbeit, Christian-Albrechts-Universität zu Kiel, 1996.

[11] Oliver Faugeras. *Three-Dimensional Computer Vision.* The MIT Press, 1993.

[12] Gsta H. Granlund. The complexity of vision. *Signal Processing*, 1999.

[13] Marco Hahn. Semiglobale Verfahren zur Generierung von Eckpunkthypothesen in 2D und 3D. Diplomarbeit, Christian-Albrechts-Universität zu Kiel, 1999.

[14] Fred Hamker. *Visuelle Aufmerksamkeit und Lebenslanges Lernen im Wahrnehmungs-Handlungs-Zyklus.* Dissertation, Technische Universität Ilmenau, 1998.

[15] Michael Hansen. *Stereosehen: Ein verhaltensbasierter Zugang unter Echtzeitbedingung.* Dissertation, Christian-Albrechts-Universität zu Kiel, 1998.

[16] Bernd Jähne. *Digitale Bildverarbeitung.* Springer-Verlag, Berlin, 1997.

[17] Philip J. Kellman und Martha E. Arterberry. *The Cradle of Knowledge: Development of Perception in Infancy.* The MIT Press, 1998.

[18] KiViGraP. (Homepage of the Kieler Vision and Grasping Project). http://www.ks.informatik.uni-kiel.de/∼ kivi/kivi.html.

[19] J.J Koenderink. Wechsler's vision: An essay review of computational vision by Harry Wechsler. *Ecological Psychology*, 4:121–128, 1992.

[20] Ivo Kohler. Über Aufbau und Wandlungen der Wahrnehmungswelt. In *Sitzungsberichte*, volume 227. Österreichische Akademie der Wissenschaften, 1952.

[21] Norbert Krüger, Eric Maël, Mike Pagel und Christoph von der Malsburg. Autonomous Learning of Object Representations Utilizing Self-Controlled Movements. In *Proceedings of Neural Networks in Applications.* Magdeburg, 1998.

[22] Norbert Krüger, Daniel Wendorff und Gerald Sommer. Work in progress on vision-based robotic system: Visual-haptic attention and accumulation of object representations. In *Proceedings of the Workshop on Dynamische Perception.* Workshop Dynamische Perzeption, Ulm, 2000.

[23] M. Lades, J.C. Vorbrüggen, J. Buhlmann, J. Lange, C. von der Mahlsburg, R.P. Würtz und W. Konen. Distortion invariant object recognition in the dynamik link architecture. *IEEE Transactions on Computers*, 42(3):300–311, 1993.

[24] Hanspeter A. Mallot, editor. *Sehen und Verarbeiten visueller Informationen.* vieweg, 1998.

[25] Harry Paul, editor. *Lexikon der Optik*, volume 1,2. Spektrum Akademischer Verlag, 1999.

[26] J. Piaget. *La contruction du reél chez l'enfant.* Delachaux & Niestlé, 1937.

[27] William H. Press, Brian P. Flannery, Saul A. Teukolsky und William T. Vetterling. *Numerical Recipes.* Cambridge University Press, 1986.

[28] H. Rauth. Frühe Kindheit. *Entwicklungspsychologie*, pages 167–248, 1995.

[29] B. Rosenhahn, N. Krüger, T. Rabsch und G. Sommer. Automatic tracking with a novel pose estimation algorithm. *Robot Vision 2001*, 2001.

[30] STÄUBLI S.A. *Roboter RX 90 - CS7 Bedienungsanleitung*, 1993.

[31] STÄUBLI S.A. *Steuerschrank CS7 Bedienungsanleitung*, 1993.

[32] Peer Schmidt. Entwicklung und Aufbau von taktiler Sensorik für eine Roboterhand. *Institut für Neuroinformatik Bochum (Internal Report)*, 2000.

[33] Josef Siemes. Entfernungsbestimmungen aus Akkomodation mit einem aktiven Kamerasystem. Diplomarbeit, Christian-Albrechts-Universität zu Kiel, 1998.

[34] G. Sommer. Algebraic aspects of designing behavior based systems. In G. Sommer und J.J. Koenderink, editors, *Algebraic Frames for the Perception and Action Cycle*, volume 1315 of *Lecture Notes in Computer Science*, pages 1–28. Proc. Int. Workshop AFPAC'97, Kiel, Springer–Verlag, Heidelberg, 1997.

[35] R.J. Walker und C.J Harris. A multi-sensor fusion system for a laboratory based autonomous vehicle. In *1st Int. Workshop on Intelligent Autonomous Vehicles*. Southampton, 1993.

[36] Robert B. Welch. *Perceptual Modification.* ACADEMIC PRESS, 1978.

[37] Robert B. Welch. Adaption of space perception. In *Handbook of perception and human performance*, volume 1. A Wiley-Interscience Publication, 1986.

[38] Daniel Wendorff. Plattformunabhängige Ansteuerung eines Binokularkopfes am Beispiel von Iris- und Fokussteuerung. Studienarbeit, Christian-Albrechts-Universität zu Kiel, 1999.

Danksagung

Mein besonderer Dank gilt Dr. Norbert Krüger für die hervorragende Betreuung und für seine Idee, das Softwarepaket KiViGraP zu erstellen. Er hat mit KiVi viele Studenten zusammengeführt. Allen KiVianeren danke ich für die gute Zusammenarbeit.

Dem Institut für Informatik und Praktische Mathematik im allgemeinen und dem Lehrstuhl für Kognitive Systeme und Prof. Dr. Gerald Sommer im speziellen danke ich für ein angenehmes Studienumfeld. Den Korrekturlesern Oliver Granert, Sandra Pursian, Alexander Slomian, meinen Eltern Christa und Erwin Wendorff und vor allem Bodo Rosenhahn danke ich für die vielen Hinweise auf Fehler und die zahlreichen Anregungen. Meinen Eltern gilt mein besondere Dank. Ohne sie wäre mein Studium und diese Arbeit nicht in dieser Form möglich gewesen.

Erklärung

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

Kiel, den 25. April 2001