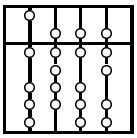
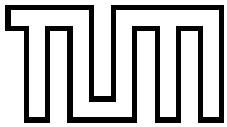


Institut für Informatik der
Technischen Universität München



Interpretation von Videobildfolgen zur Beobachtung artikularer Bewegung von Personen anhand eines generischen 3D Objektmodells

Dissertation

Christof Ridder

Institut für Informatik
der Technischen Universität München
Lehrstuhl Univ.-Prof. Dr. B. Radig

Interpretation von Videobildfolgen zur Beobachtung artikularer Bewegung von Personen anhand eines generischen 3D Objektmodells

Christof Ridder

Vollständiger Abdruck der von der Fakultät für Informatik der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitzender: Univ.-Prof. (komm.) Dr. G. Specht
Prüfer der Dissertation:
1. Univ.-Prof. Dr. B. Radig
2. Hon.-Prof. Dr. G. Hirzinger

Die Dissertation wurde am 13. 12. 1999 bei der Technischen Universität München eingereicht und durch die Fakultät für Informatik am 8. 3. 2000 angenommen.

für Birgit

Danksagung

Die hier vorzustellenden Entwicklungen entstanden während meiner Tätigkeit am Bayerischen Forschungszentrum für Wissensbasierte Systeme (FORWISS) in München. In der Forschungsgruppe Kognitive Systeme habe ich das notwendige Umfeld zur Realisierung vorgefunden. Hierfür möchte ich dem Leiter der Forschungsgruppe und meinem Doktorvater Herrn Prof. Dr. Bernd Radig danken. Ebenso gilt mein Dank Herrn Prof. Dr. Gerd Hirzinger vom DLR-Oberpfaffenhofen für die Übernahme des Zweitgutachtens.

Ich bedanke mich außerdem bei allen Kollegen aus der Forschungsgruppe und dem Lehrstuhl von Prof. Radig, die mich in dieser Zeit unterstützt haben. Stellvertretend möchte ich mich bei Olaf Munkelt und Christoph Zierl für so manch kritischen Ratschlag, anregende Diskussionen und Korrekturen an dem vorliegenden Werk bedanken.

Viele Einzelheiten des Konzeptes wären ohne die studentischen Hilfskräfte in dem von mir geleiteten STABIL^{++}–Team nicht realisiert worden. Mein Dank an das Team gilt daher neben Sónia Antunes, Thomas Diederich, Felix Ko und Ingmar Rauschert insbesondere David Hansel, der mich immer wieder durch unermüdlichen Einsatz und unerschöpflichen Ideenreichtum auf's Neue überraschte.

Eine lange Zeit meiner Tätigkeit am FORWISS teilte ich mit Walter Hafner ein Büro. Da blieb es nicht aus, daß ich ihn mit etlichen Fragen löcherte und – er blieb mir nie eine Antwort schuldig. Auch nach seinem Ausscheiden vom FORWISS ließ meine "Fragerei" nicht nach und per Telefon bekam ich noch weitere Tips und Ratschläge. Hierfür und für die Durchsicht meiner Entwürfe möchte ich mich bedanken.

In der letzten Phase des Zusammenschreibens habe ich mich nicht mehr sehr intensiv um das Tagesgeschäft kümmern können und vieles ist an meinen Kollegen hängen geblieben. Im Besonderen möchte ich mich hierfür bei Christoph Hansen entschuldigen und bedanken. Neben der Übernahme des "Telefondienstes" stand er mir auch immer wieder als Testperson für die Experimente zur Verfügung.

Schließlich möchte ich mich noch bei meiner Familie und meinen Freunden dafür entschuldigen, daß ich mich in der heißen Endphase bei fast keinem mehr habe blicken lassen und so manchen Kontakt sträflichst vernachlässigt habe. Dies wird wieder anders werden!

München, im Dezember 1999

Kurzfassung

Die videobasierte Objektlokalisation ist aus Anwendungen zur Vermessung in der industriellen Inspektion, der Navigation von autonomen Systemen und bei Greifvorgängen in der Robotik bekannt. Hierbei stützen sich die Systeme meist auf starre Objektmodelle, die aus CAD-Daten generiert werden. Im Gegensatz hierzu sind auch nicht-starre Objekte und deren Bewegung mit Bildinterpretationssystemen zu beobachten, wobei die Objekte entsprechend ihrer Beweglichkeit noch weiter klassifiziert werden. Eine solche Klasse sind die artikularen Objekte, die sich aus mehreren einzelnen Objektmodellteilen zusammensetzen, die durch Gelenke miteinander verbunden sind.

Es wird daher ein dreischichtiges generisches Modell zur Repräsentation von artikularen Objekten vorgestellt. Über eine hierarchische, innere Struktur des Modells kann ein beliebiger innerer Zusammenhang zwischen Objektmodellteilen beschrieben werden. Mit einer geometrischen Struktur ist die 3D Lage der einzelnen Objektmodellteile zueinander bestimmt, so daß sich die Kinematik der Objekte beschreiben läßt. In einer äußeren Struktur wird schließlich die Erscheinung der einzelnen Objektmodellteile über 3D Volumenkörper und über, im Videobild zu detektierende, Merkmale repräsentiert.

Im Unterschied zu anderen Ansätzen wird kein Bewegungsmodell vorausgesetzt, um die zu bestimmende Konfiguration der Objekte vorherzusagen. Daher wird bei der Interpretation kein 2D/2D Vergleich von einem projizierten 3D Modell mit extrahierten Bildmerkmalen, sondern die Interpretation im 3D Raum vorgenommen. Hierzu wird ein Interpretationsbaum verwendet, dessen Aufbau durch die innere Objektmodellstruktur bestimmt ist. Restriktionen, die sich aus der geometrischen und äußeren Modellstruktur ergeben, begrenzen die Suche im Baum. Werden Objekte und deren Bewegung nach einer initialen Detektion verfolgt, wird die Suche darüber hinaus durch 3D Suchräume eingeschränkt, die aus der erfaßten Bewegung der einzelnen Objektmodellteile heraus prädiziert werden.

Das Objektmodell ist derart gestaltet, daß die für die Interpretation notwendigen 3D Positionen der Merkmale sowohl über einen monokularen Ansatz geschätzt, als auch durch die Verwendung von mehreren Ansichten mit einem Stereoansatz vermessen werden können. Die bei der Interpretation zu verwendenden Kameras werden entsprechend der Sichtbarkeit der 3D Suchräume ausgewählt, wobei die Kameras zu einem 3D Bezugssystem kalibriert sind. Hierüber und durch die konsequente 3D Modellierung ist implizit eine Verfolgung des Objektes über mehrere Kameras realisiert, sowie bei aktiven Kamerasystemen eine Optimierung der Sichtbereiche durch gezielte Positionierung der Kameras möglich.

Die Beobachtung artikularer Bewegung wird meist mit der Beobachtung des menschlichen Körpers gleichgesetzt, dementsprechend werden auch hier Anwendungen auf Personen gezeigt. Hierzu realisiert das modellbasierte Bildinterpretationssystem STABIL^{++} das dargestellte Konzept. Die Flexibilität der Modellierung erlaubt es, neben einer 3D Personendetektion und -verfolgung für Anwendungen in der Sicherheitstechnik, Anwendungen zur 3D Bewegungserfassung für die Analyse von Bewegungsabläufen unter z.B. ergonomischen Gesichtspunkten zu realisieren.

Inhaltsverzeichnis

1 Einleitung

1.1 Zielsetzung und Motivation

Die Entwicklungen von intelligenten Systemen zum maschinellen Erfassen, Verstehen und Interpretieren der Umwelt schreiten immer weiter voran. Dies ist hauptsächlich durch den schnellen Fortschritt der Rechnertechnologie bedingt. Insbesondere bei Systemen, die zur Erfassung der Umwelt Videokameras als Sensoren einsetzen, konnten in den vergangenen Jahren rapide Entwicklungen verzeichnet werden. Neben der Weiterentwicklung der Verfahren zur Verarbeitung von digitalen Bildern war dies auch durch die Weiterentwicklung der Rechnerkomponenten begünstigt. Zum einen basiert dieser Trend auf der allgemeinen Steigerung der Leistung von RechnersystemenXS, zum anderen jedoch auf der Tatsache, daß die Anzahl der zur Verfügung stehenden Digitalisierungskarten zur Erfassung analoger Videosignale für Standardrechner beträchtlich angestiegen ist. Daher kommen die videobasierten Systeme in zunehmendem Maße ohne Spezialkomponenten aus.

Diese Systeme, die auch als *Bildinterpretationssysteme* bezeichnet werden, kommen z.B. bei der visuellen Inspektion zur Qualitätskontrolle, der Luftbildauswertung oder der kameragestützten Navigation von Robotern zur Anwendung. Alle diese Systeme haben gemeinsam, daß zuvor bestimmt sein muß, nach welchen Informationen in den digitalen Videobildern gesucht wird; hierzu sind Merkmale zu definieren. Werden bei der Interpretation komplexere Objekte erwartet, so können diese nur als eine Gruppierung von Merkmalen im Videobild gefunden werden. Es wird dann von *Objekterkennung* gesprochen.[1] Es soll hier jedoch im weiteren von *Objektdetektion* gesprochen werden, da bei der Interpretation zunächst nur das Vorhandensein eines Objektes festgestellt wird. Erst in einem weiteren Schritt kann dann ein detektiertes Objekt identifiziert werden, so daß dieser Schritt als Objekterkennung bezeichnet werden kann.

Mit einem strukturierten Zusammenschluß von Merkmalen wird für eine spezifische Objektart bestimmt, wie die Objekte in den zur Interpretation zur Verfügung stehenden Videobildern dargestellt werden. Diese Gruppierungen von Merkmalen können als Modell der Objekte bezeichnet werden. Man spricht daher von einem *Objektmodell*, das Objekte einer Objektart repräsentiert. Systeme, die sich zur Interpretation auf Objektmodelle stützen, werden als modellbasierte Bildinterpretationssysteme bezeichnet. Betrachtet man mit dem System nicht nur Bilder von unabhängigen Zeitpunkten, sondern Bildfolgen, so kann man darüberhinaus die Bewegung von Objekten erfassen. Man erhält Systeme zur Verfolgung und zur Bewegungserfassung der Objekte.

Ein weites Anwendungsgebiet für Videokameras ist in zunehmendem Maße die Beobachtung von Personen, deren Bewegungen und Handlungen. Es werden zum einen vermehrt Videokameras in sicherheitsrelevanten Bereichen zur Beobachtung und Überwachung eingesetzt. Zum anderen werden Videokameras zur Analyse von Bewegungen unter ergonomischen Gesichtspunkten verwendet. Das gleiche gilt für medizinische Beobachtung von Bewegung des

[1] engl. *object recognition*.

Abbildung 1.1: *Anwendung zur Personendetektion und -verfolgung: Bei der Beobachtung des Raumes mit zwei Kameras sind drei Personen detektiert und lokalisiert worden. Hierbei wird als Merkmal die "hautfarbene" Ellipse des Gesichts verwendet. Anhand der ermittelten 3D Positionen wird der Weg der beobachteten Personen sichtbar gemacht.*

menschlichen Körpers. Alle Anwendungen der Videotechnik haben gemeinsam, daß bisher eine manuelle Auswertung der Videoaufnahmen vorgenommen werden muß. So muß in Sicherheitssystemen das Wachpersonal im Videobild erkennen, ob eine Person einen sicherheitsrelevanten Bereich betreten hat. Es werden auch Kameras manuell über Steuerpulte hinter Personen her geschwenkt, so daß die Handlung lückenlos aufgezeichnet werden kann. In der Ergonomie und in der Medizin werden die Bewegungsabläufe visuell mit Bewegungsmustern verglichen oder es werden Modelle manuell überlagert, so daß Aussagen über die Haltung der beobachteten Person gemacht werden können.

Durch die Verwendung von Personenmodellen zur modellbasierten Bildinterpretation können die dargestellten Anwendungen (teilweise) automatisiert werden. Dies wird erreicht, indem Personen in Videobildern detektiert, lokalisiert und deren Bewegung erfaßt werden. In sicherheitsrelevanten Anwendungen kann somit automatisch Alarm ausgelöst werden, wenn eine Person einen zu überwachenden Bereich betritt. Weiterhin sind Systeme denkbar, mit denen die detektierten Personen über mehrere Kameras hinweg verfolgt werden. Ebenso kann anhand der Lokalisation der detektierten Person eine in der Sicherheitstechnik übliche Schwenk- / Neigekamera so positioniert werden, daß die Person im Videobild immer optimal abgebildet wird.

In Anwendungen zur Analyse der Bewegung des menschlichen Körpers ist ein erster Schritt die Erfassung der Bewegung. Beispielsweise könnte ein Ziel der automatischen Auswertung der Videoaufnahmen die Bestimmung der 3D Gelenkpositionen oder der Gelenkwinkel sein. Für eine aufgenommene Bildfolge können die Trajektorien der Gelenke oder die Gelenkwinkelverläufe ermittelt werden, die anschließend unter medizinischen oder ergonomischen Gesichtspunkten analysiert werden.

Das hier vorzustellende Konzept ist mit dem Ziel entwickelt worden, beide der aufgezeigten Anwendungsbereiche abdecken zu können. Hierzu wurde ein generisches Objektmodell entwickelt, das entsprechend der Anwendung konfiguriert werden kann. Die Möglichkeit der Konfiguration erlaubt es zudem, die Modellierung auch auf weitere Objekte anzuwenden. Ein weiteres Ziel war es, ein Interpretationssystem zu schaffen, das aufgrund der heterogenen Anwendungsgebiete nicht auf eine bestimmte Systemkonfiguration beschränkt ist. Daher muß die Anzahl der zur Interpretation zu verwendenden Kameras beliebig sein. Es wird neben dem Objektmodell in dem vorzustellenden Konzept ein Szenenmodell verwendet. Das Szenenmodell beinhaltet, neben einer hinreichenden Beschreibung der Umgebung, die Definitionen der zur Interpretation zur Verfügung stehenden Kameras. Mit dem Objektmodell ist somit bestimmt, *wonach* im Interpretationsprozeß zu suchen ist und mit dem Szenenmodell, *wo* und *womit* die

Abbildung 1.2: *Anwendung zur Bewegungserfassung: Beobachtung eines Einstiegsvorgangs in einen PKW unter ergonomischen Gesichtspunkten. Hierzu werden die Gelenke der zu beobachtenden Person anhand farbiger Markierungen in mind. zwei Ansichten extrahiert, lokalisiert und die Gelenkwinkel bestimmt. Der erfaßte Bewegungsablauf läßt sich als animierte Computergraphik simulieren.*

Interpretation durchgeführt wird.

Basierend auf diesem Konzept ist das modellbasierte Bildinterpretationssystem STABIL^{++}[2] zur Anwendung auf die Beobachtung von Bewegung des menschlichen Körpers realisiert worden. Das System ist erfolgreich für Anwendungen im Sicherheitsbereich auf der weltgrößten Sicherheitsmesse *SECURITY* in Essen im Oktober 1998 gezeigt worden. Eine beispielhafte Anwendung ist in Abb. 1.2 für die Beobachtung eines Raumes mit zwei Kameras gezeigt. Eine Anwendung zur Erfassung der Bewegung unter ergonomischen Gesichtspunkten wurde auf dem *Symposium Zulieferer Innovativ* der Bayerischen Innovations- und Kooperationsinitiative Automobilzulieferindustrie in Ingolstadt im Juni 1999 präsentiert. Dort wurden die Bewegungen beim Einstieg in einen PKW erfaßt, vgl. Abb. 1.2.

[2] STABIL^{++} ist als Weiterentwicklung aus dem Projekt STABIL am Bayerischen Forschungszentrum für Wissensbasierte Systeme (FORWISS), München realisiert worden, vgl. [MK96, MRHH98, RMR+99]. STABIL^{++} ist ein Akronym der englischen Bezeichnung "**S**ystem for **T**racking **a**rticulated O**b**jects using **I**mage-based 3D **L**ocalization implemented in C++".

1.2 Kapitelübersicht

Zur Darstellung des Konzeptes wird in dem folgenden *Kapitel 1.3* zunächst ein Überblick über zugrundeliegende und verwandte Arbeiten gegeben. Daran schließt sich in *Kapitel 1.4* ein Überblick über die Systemstruktur von STABIL^{++} und eine Abgrenzung zu anderen Arbeiten an. Das Konzept der Modellierung wird in *Kapitel 2* dargestellt. Hierbei ist das Kapitel in Abschnitte zum Szenenmodell (*2.2*), zum Objektmodell (*2.3*), zu den Merkmalen des Objektmodells (*2.4*) und zum Kameramodell (*2.5*) unterteilt.

Im *Kapitel 3* wird der auf der Modellierung basierende Interpretationsprozeß beschrieben. Im ersten Teil wird dort der Ablauf des Prozesses erläutert, s. *Kapitel 3.1*. Die weiteren Abschnitte *Kapitel 3.2 – 3.8* gliedern sich entsprechend der Reihenfolge des dargestellten Prozeßablaufes. An die Erläuterung der Experimente mit einem Modell/Modell-Vergleich in *Kapitel 4.1* schließen sich die Darstellungen von Anwendungen zur Objektdetektion und -verfolgung in *Kapitel 4.2* und von Anwendungen zur Bewegungserfassung in *Kapitel 4.3* an.

Auf eine Zusammenfassung und einen Ausblick im *Kapitel 5* folgen im Anhang weitere Erläuterungen. Dies sind im *Anhang A* eine Beschreibung der Bestimmung von Gelenkwinkeln, gefolgt von einer Darstellung von Transformationen im 3D Raum in *Anhang B.* Daran schließt sich im *Anhang C* eine Erläuterung der verwendeten Kamerakalibrierung und im *Anhang D* ergänzende Beispiele zum Interpretationsprozeß an.

Zur Orientierung sind abschließend, nach dem Anhang noch *Verzeichnisse* der Abbildungen, Tabellen und Algorithmen angegeben. Weiterhin ist das *Symbolverzeichnis* zu beachten, in dem alle die mathematischen Symbole aufgelistet sind, die bei der dargestellten Modellierung Verwendung finden und zur Erläuterung des Interpretationsprozesses verwendet werden.

1.3 Wissenschaftlicher Kontext

In diesem Abschnitt wird ein Überblick über Arbeiten gegeben, die aus der Literatur bekannt sind und die den gleichen Kontext wie das vorzustellende Konzept zur modellbasierten Beobachtung artikularer Bewegung haben. Es wird daher zunächst allgemein auf die modellbasierte Objekterkennung / Objektdetektion eingegangen, die sich meist auf starre Objekte bezieht. Daran schließt sich eine Betrachtung von nicht-starren Objekten und deren Bewegung an, wobei besonderes Augenmerk auf artikulare Objekte gelegt wird.

Nachdem die Beobachtung artikularer Bewegung meist mit der Beobachtung des menschlichen Körpers gleichgesetzt wird, folgt ein Überblick zu Arbeiten zur Bewegungserfassung bei Personen, Personendetektion und -verfolgung, sowie abschließend ein Ausblick auf Arbeiten zur Analyse und Erkennung von Bewegung des menschlichen Körpers.

Das Ziel ist hierbei mehr eine Auflistung von verschiedenen Ansätzen und eine Bestimmung von Begriffen, als eine ausführliche Diskussion verschiedener Ansätze.

1.3.1 Modellbasierte Objektdetektion

Ein videobasierter Interpretationsprozeß dient zur Interpretation einer Szene, die von Kameras beobachtet wird. Das Ziel ist dabei, Objekte zu detektieren und zu lokalisieren. Dem Interpretationsprozeß muß hierzu in geeigneter Weise Wissen über die Ausprägung der Objekte zur Verfügung gestellt werden. Bei der Verwendung von CCD-Kameras als Sensoren kann man auch davon sprechen, wie die Objekte “aussehen”. Es reicht jedoch nicht aus, verbal zu beschreiben, welche Form, Größe und Farbe die Objekte haben, sondern es muß eine explizite Modellvorstellung verwendet werden.

Zur Detektion der Objekte, also um das Vorhandensein eines bestimmten Objektes feststellen zu können, muß in einem entsprechenden Objektmodell eine Anzahl von Merkmalen festgelegt sein, die die Erscheinung des Objektes geeignet beschreiben. In diesem Zusammenhang wird auch von Objekterkennung gesprochen, denn ein zunächst unbekanntes Objekt wird bei der Detektion anhand seiner, im Videobild extrahierten, Merkmale einem bestimmten Objektmodell zugeordnet. Diese “Erkennung” bedeutet jedoch nicht, daß das Objekt identifiziert wird.[3] Vielmehr wird das bisher unbekannte Objekt klassifiziert, in dem es einem Objektmodell zugeordnet wird. Verschiedene Objektmodelle können somit als Modellklassen bezeichnet werden, wobei für ein detektiertes Objekt eine Instanz einer Modellklasse instantiiert wird. Man spricht daher nach einer initialen Detektion eines Objektes von einer *Objektmodellinstanz*. Eine Objektmodellinstanz zeichnet sich dann dadurch aus, daß zusätzlich zu den Merkmalen, die allen Instanzen einer Modellklasse zu eigen sind, diesem noch weitere Attribute zugeordnet werden, die es eindeutig charakterisieren. Diese weiteren Attribute sind z.B. aktuelle Positionsdaten, die durch die Lokalisation bestimmt werden.

Die hier eingeführte Unterscheidung zwischen Erkennung und Detektion ist in der Literatur leider nicht deutlich zu finden. So wird in der üblicherweise englischsprachigen Literatur der Prozeß des Auffindens eines Objektes basierend auf Modellinformation als *model-based object recognition* bezeichnet, vgl. hierzu auch den Überblick zu dieser Thematik von Pope [Pop94]. Grundlage der modellbasierten Ansätze ist zwangsläufig eine geeignete Repräsentation der Objekte. In [Pop94] werden daher die Repräsentationen entsprechend der zwei verschiedenen An-

[3] Bei Systemen zur videobasierten Zutrittskontolle mit Gesichtserkennung werden jedoch, über eine Detektion hinaus, Personen identifiziert, so daß dort von einer Objekterkennung gesprochen werden kann.

sichten unterschieden, aus denen ein Objekt beschrieben werden kann. Dies sind:

- *viewer-centred representation* und
- *object-centred representation*

Die erste Modellrepräsentation beschreibt das Objekt aus der Sicht des Betrachters und somit aus Sicht der Kamera. Diese Repräsentation wird auch als *multi-view* Repräsentation bezeichnet, denn es wird für die Modellierung das Objekt entsprechend unterschiedlicher 3D Lagen aus verschiedenen Ansichten betrachtet. Man erhält somit viele charakteristische 2D Abbildungen des Objektes, die auch als Aspekte bezeichnet werden. Für eine weitere Diskussion des Aspektbegriffes sei auf [Mun96] verwiesen.

Bei der zweiten Modellrepräsentation wird das Objekt aus sich selbst heraus oder auf sich selbst bezogen beschrieben. Es wird bei dieser Repräsentation für ein Objekt ein Koordinatensystem definiert, auf das sich dann beschreibende Elemente beziehen, die seine Erscheinung modellieren. Durch eine geometrisch eindeutige Darstellung kann somit die Erscheinung aus verschiedenen Ansichten rekonstruiert werden. Man spricht davon, daß das Modell, basierend auf den Abbildungseigenschaften der Kameras, in die Bilder projiziert wird.[4]

Bei der eigentlichen Interpretation müssen Merkmale, die im Bild extrahiert werden, mit Merkmalen des Modells verglichen werden; man spricht hierbei von einem *Struktur-Vergleichsverfahren*, auch besser unter dem englischen Fachbegriff *matching* bekannt. Allgemein wird der notwendige Vergleich bei der ersten Repräsentation als weniger aufwendig betrachtet. Dies liegt darin begründet, daß für den Vergleich nicht zusätzlich das 3D Modell in das Bild projiziert werden muß, denn es liegen schon 2D Ansichten vor. Es kann somit direkt ein 2D/2D Vergleich von 2D Merkmalen im Bild vorgenommen werden, hierbei sind sog. *Bildmerkmale* mit projizierten *Modellmerkmalen* zu vergleichen.

Bei der zweiten Art der Projektion spricht man von einem 3D/3D oder 3D/2D Vergleich. Der 3D/3D Vergleich setzt voraus, daß aus den Sensordaten 3D Informationen gewonnen werden können. Dies kann z.B. durch Kreuzpeilung von Sendern erreicht werden, die an einem Objekt befestigt sind, vgl. [N$^+$94, D$^+$97]. Bei der Verwendung von mehreren Kameras als Sensoren können über Methoden der Stereobildverarbeitung ebenfalls 3D Lageinformationen von Merkmalen gewonnen werden, vgl. [XZ96]. Man spricht dann von 3D *Szenenmerkmalen*, die mit 3D Modellmerkmalen zu vergleichen sind. Wird hingegen ein 3D/2D Vergleich angestrebt, so muß das 3D Modell zuvor in die zur Interpretation zu verwendenden Kamerabilder projiziert werden. Hierfür muß jedoch eine Lage des 3D Objektes zur Kamera angenommen werden. Zur exakten Lokalisation der Objekte, z.B. für kameragestützte Greifaufgaben in der Robotik, wird daher zunächst eine angenäherte Lage der Objekte angenommen, die 3D Objekte werden projiziert und unter Beachtung von Zuordnungstoleranzen wird eine gültige Zuordnung von 2D Bildmerkmalen zu projizierten 3D Modellmerkmalen vorgenommen. Die zunächst angenommene Lage kann dann in weiteren Schritten iterativ verfeinert werden, wobei Wissen aus dem Modell und den extrahierten Bildmerkmalen genutzt wird, vgl. z.B. [Wun98]. Das Ziel ist dabei, die Lage des Objektes so zu bestimmen, daß die Fehler zwischen dem projizierten Modell des Objektes und dem Abbild des Objektes minimal werden. Vgl. hierzu auch die Fehlerminimierung[5] in der Arbeit von Lanser [Lan98].

Viele Systeme setzen bei der beschriebenen Vorgehensweise eine initiale Lage für einen 3D/2D Vergleich voraus. Bei der Interpretation von Bildfolgen mit genügend großer Bildwiederholrate kann die jeweils letzte bestimmte Objektlage als angenommene Lage für die folgende

[4] Zur Beschreibung der Abbildungseigenschaften werden Kamera- / Projektionsmodelle verwendet. In STABIL^{++} kommt das Lochkameramodell mit radialer Verzerrung zur Anwendung, vgl. Kap. 2.5.4.

[5] Methode der kleinsten Quadrate; (engl.) *least-squares*.

Interpretation verwendet werden. Ist für ein Objekt ein Bewegungsmodell vorhanden, so kann hieraus die Ausgangslage für die jeweils folgende Interpretation prädiziert werden. Eine weitere Möglichkeit ist die Bestimmung der ungefähren Objektlage durch andere Sensoren, z.B. Laserabtaster, wenn multisensorielle Systeme eingesetzt werden.

Eine weitere Einteilung von Objektmodellen und Objektrepräsentationen haben Ponce et al. in [PHZ96] vorgenommen. Dies sind:

- die geometrische und topologische Repräsentation und
- die erscheinungsbasierte Repräsentation

Hierbei ist für den geometrischen / topologischen Ansatz ein Objektmodell als ein diskreter Satz von geometrischen Merkmalen wie Volumen- oder Oberflächenelementen zu definieren. Bei der Interpretation werden explizite Merkmale verglichen. Bei dem erscheinungsbasierten Ansatz ist die Modellierung durch eine Menge von Bildern gegeben, so daß man von Mustern spricht. Es werden hier zur Interpretation keine expliziten Merkmale im Bild extrahiert. Vielmehr wird bei diesem Ansatz nach Ähnlichkeiten zwischen Modelldatensatz und dem Bild gesucht, dies kann auf der Basis von Grauwerten und Farben im Bild oder anhand von geometrischen 2D Formen vorgenommen werden.

Diese Einteilung der Ansätze kann mit der o.a. Einteilung aus [Pop94] direkt verglichen werden. Die geometrischen Repräsentationen sind auf das Objekt bezogen, wogegen die erscheinungsbasierten Ansätze meist blickrichtungabhängig sind. Jedoch muß ein multi-view Ansatz nicht zwangsläufig ein erscheinungsbasierter Ansatz sein, denn es können auch bei diesen explizite Merkmale verwendet werden. Bei einer erscheinungsbasierten Repräsentation wird das Modell empirisch durch Trainingsphasen ermittelt, so daß von einer Klassifikation des Bildes gesprochen werden kann.

Bei den geometrischen Ansätzen sind die Modelle hingegen analytisch zu erstellen, d.h. die Objekte werden vermessen oder aus CAD-Daten bestimmt, vgl. z.B. [Mun96, WH96]. Die sich hier aus der starren Anordnung von Modellelementen und Merkmalen ergebenden geometrischen und topologischen Restriktionen können genutzt werden, um inkonsistente Zuordnungen beim Struktur-Vergleichsverfahren zu vermeiden.[6] Ein gängiges Verfahren zum Auffinden von gültigen Zuordnungen von Merkmalen ist die Suche in einem sog. *Interpretationsbaum*. Mit diesem werden für die Merkmale eines Modells nacheinander Zuordnungen vorgenommen, die anhand der Restriktionen auf ihre Gültigkeit überprüft werden, vgl. hierzu die Arbeiten von Grimson [GLP84], [GLP87]. Die Restriktionen, die sich aus dem geometrischen Modell ergeben, begrenzen den aufgespannten Suchraum des Baumes [LZ96].

Bei der Verwendung von geometrischen Modellen zur Objektdetektion und -lokalisation besteht bei einer annähernd bekannten Lage oder einer aus der Bewegung des Objektes heraus prädizierten Lage die Möglichkeit, für die Merkmale Suchräume vorherzusagen.[7] Bei der Verwendung von 3D Modellen können insbesondere bei der Verfolgung von Objekten in Bildfolgen 3D Suchräume vorhergesagt werden, die dann in die Bilder der zur Interpretation zu verwendenden Kameras zu projizieren sind. Bei dem Struktur-Vergleichsverfahren müssen dann nur noch die Bildmerkmale berücksichtigt werden, die in den Bildbereichen liegen, die für ein korrespondierendes Modellmerkmal vorhergesagt worden sind. Die Suchräume für die Merkmale

[6] Die Verwendung von Restiktionen wird mit dem englischen Fachbegriff als *constrained search* bezeichnet.

[7] Die Verwendung des Begriffs “Suchraum” kann hier zu Verwirrung führen: Mit der Größe des aufgespannten Suchraumes eines Interpretationsbaumes wird ein Maß für seine Komplexität angegeben. Dies ist nicht mit den 3D Suchräumen zu verwechseln, mit denen die Lage von Objektmerkmalen vorhergesagt wird.

des Objektes können daher ebenfalls als Restriktion verwendet werden, mit denen falsche Zuordnungen vermieden werden. Dies führt zu einer weiteren Reduzierung der Komplexität des Struktur-Vergleichsverfahrens.

Weiterführende Überblicke zu Arbeiten der modellbasierten Objektdetektion / Objekterkennung sind in [Mun96], [Wun98] und [Lan98] zu finden.

1.3.2 Artikulare Objekte

In den im vorhergehenden Abschnitt aufgeführten Arbeiten zur modellbasierten Objektdetektion werden allgemein in sich starre Objekte betrachtet, deren Modelle sich z.B. aus CAD-Daten generieren lassen. Eine weitere Klasse von Objekten sind die nicht starren Objekte.[8] Aggawarl et al. geben in [ACLS94] und [ACLS98] einen ausführlichen Überblick über Forschungsarbeiten zur videobasierten Analyse von Bewegung nicht starrer Objekte. Es wird dort, basierend auf [KGTH94], eine weitere Einteilung nach Bewegungstypen vorgenommen. Generell wird zunächst zwischen rigider und nicht rigider Bewegung unterschieden. Eine rigide Bewegung kann von einem starren Objekt ausgeführt werden und ist daher lediglich eine Translation des Objektes oder eine Rotation. Kann das Objekt selbst seine Form verändern, dann wird von nicht rigider Bewegung gesprochen; bei dieser wird weiterhin zwischen einer Klasse genereller, unbeschränkter Bewegung und einer Klasse beschränkter Bewegung unterschieden. Diese beiden Klassen sind noch weiter unterteilt:

- beschränkte Bewegung:
  - artikulare Bewegung (*articulated*)[9]
  - quasi rigide Bewegung (*quasi-rigid*)[10]
  - isometrische Bewegung (*isometric*)[11]
  - homothetische Bewegung (*homothetic*)[12]
  - sich anpassende Bewegung (*comformal*)[13]
- generelle, unbeschränkte Bewegung:
  - elastische Bewegung (*elastic*)[14]
  - Bewegung von Flüssigkeiten (fluid)[15]

Aufgrund der vorzustellenden Konzepte wird hier nur die artikulare Bewegung genauer betrachtet: Artikulare Bewegung ist die Bewegung, die von Objekten ausgeführt werden kann, die sich aus einzelnen starren Teilen zusammensetzen. Die einzelnen Objektteile sind hierbei durch Gelenke verbunden. Jedes einzelne Teil kann für sich nur eine rigide Bewegung ausführen, so daß sich die artikulare Bewegung aus den rigiden Bewegungen aller Objektteile ergibt.

Ein artikulares Objekt kann verschiedene Konfigurationen / Haltungen annehmen, die sich zum einen durch unterschiedliche Lagen der einzelnen Objektteile zueinander und zum anderen

[8] Die engl. Fachbegriffe sind: *rigid* und *non-rigid objects*.
[9] Stückweise rigide Bewegung von einzelnen Objektteilen.
[10] Verformungen, die in der zu beobachtenden Zeitspanne sehr klein sind.
[11] Bewegungen entlang einer Oberfläche, meist reine Längenausdehnung von Objekten.
[12] Bewegungen mit gleichmäßiger Ausdehung von Objekten entlang einer Oberfläche.
[13] Bewegung, die z.B. einer gekrümmten Oberfläche folgt.
[14] Bewegung, deren einzige Beschränkung durch Kontinuität und Gleichmäßigkeit gegeben ist.
[15] Bewegung, die nicht kontinuierlich sein muß, jedoch topologische Veränderungen und Turbulenzen aufweist.

durch die Lage des Gesamtobjektes zum Raum ergeben. Eine Lage kann hierbei durch Translation und Rotation beschrieben werden, so daß im Modell die Position und die Orientierung der Objektmodellteile zueinander erfaßt werden können. Es bietet sich daher an, neben einem Koordinatensystem für das Gesamtobjektmodell, für jedes Objektmodellteil ein lokales Koordinatensystem zu definieren. Die zu beobachtende artikulare Bewegung setzt sich somit aus einer Abfolge von Gelenkkonfigurationen zusammen.

Sind aufgrund der Struktur des zu modellierenden artikularen Objektes in den Gelenken nur Rotationen möglich, so kann man von einer Skelettstruktur des Objektes sprechen. Aufgrund der Forderung, daß die einzelnen Objektteile starr sind, werden mit der Skelettstruktur feste Abstände zwischen den Ursprüngen der lokalen Koordinatensysteme der Objektmodellteile vorausgesetzt. Dies kann als Restriktion zur Bestimmung der Konfiguration von artikularen Objekten genutzt werden. In den Arbeiten von Hel-Or und Wermann [HOW94, HOW96] zur Lokalisation von artikularen Objekten wird diese Restriktion auf die Beobachtung einer Schreibtischlampe mit Schwenkarm angewendet. Zunächst werden die 3D Positionen einzelner Punkte, die mit Punktmerkmalen des Objektes korrespondieren, durch einen Stereoaufbau gewonnen. Bei der Suche nach gültigen Zuordnungen von Modellmerkmalen zu den ermittelten Szenenmerkmalen werden die festen Abstände zwischen den einzelnen Modellteilen ausgenutzt. Mit diesen Zuordnungen sind die Positionen der einzelnen Modellteile bestimmt. Zum Auffinden der exakten Konfiguration der Gelenke im Schwenkarm der Lampe wird unter Verwendung eines Kalmanfilters die Lage der Modellteile zueinander geschätzt, wobei auch hier die festen Translationen zwischen den Modellteilen ausgenutzt werden.

Ein weiteres Beispiel zur Detektion und Lokalisation von artikularen Objekten ist von Hauck et al. in [HLZ97] gegeben. Dort wird eine hierarchische Ordnung der Objektmodellteile vorausgesetzt. Hierauf aufbauend wird zunächst das Basismodellteil der Hierarchie lokalisiert. Es werden hierzu die Verfahren zur Detektion und Lokalisation starrer Objektmodellteile verwendet, wie sie in [Lan98] vorgestellt sind. Ausgehend von der Lage des ersten Modellteiles werden dann die weiteren Modellteile nacheinander, entsprechend der Hierarchie, lokalisiert. Hierbei werden kinematische Restriktionen, die sich aus der Struktur des modellierten Objektes ergeben, für die Gelenke zwischen den Objektmodellteilen ausgenutzt. Als Beispielobjekte sind ein Türrahmen mit zu öffnendem Türblatt, ein Karteischrank mit herausziehbarer Schublade und eine Unterrichtstafel mit klappbaren Seitenelementen angegeben.

Die kurz vorgestellten Beispiele beziehen sich beide auf technische Objekte, so daß sich die Objektmodelle und Objektmodellteile z.B. aus CAD-Daten generieren lassen. Im Gegensatz hierzu wird jedoch meist die Beobachtung artikularer Objekte und deren Bewegung mit der Beobachtung von Personen gleichgesetzt. Es wird daher im folgenden Abschnitt besonders auf Systeme zur Beobachtung des menschlichen Körpers eingegangen.

1.3.3 Beobachtung von Bewegung des menschlichen Körpers

Die Beobachtungen der Bewegung des menschlichen Körpers mit videobasierten Systemen haben unterschiedliche Zielsetzungen. Die Anwendungen reichen von einer Bestimmung der Position und der Verfolgung von einzelnen oder mehreren Personen bis zur hochgenauen Erfassung von Bewegungsabläufen einzelner Gliedmaßen, vgl. auch Tab. 1.1. Ebenso kann die Bewegung von Personen erfaßt werden, um diese für eine Mensch-Maschine-Kommunikation zu deuten. Weiterhin wird Bewegung erfaßt, um virtuelle “Personen” zu animieren. Zur Verwendung von virtuellen “Personen” hat Badler in [Bad97] einen weitreichenden Überblick über die Anwendungen, die Technologien und die Forschung gegeben. Zu den unterschiedlichen Ansätzen und Arbeiten zur Personendetektion und -verfolgung, sowie der Erfassung der Bewegung von Per-

	Anwendungsgebiet	Anwendung
1	VR (*virtual reality*)	interaktive virtuelle Welten Computerspiele virtuelle Studios Animation von Charakteren Telefonkonferenzen
2	Überwachungssysteme	Zutrittskontrolle Parkplatzüberwachung Überwachung von Supermärkten, Kaufhäusern Überwachung von Geldautomaten Verkehrsüberwachung
3	Intelligente Benutzer-schnittstellen	Erkennung von Zeichensprache Steuerung von Maschinen über Gesten Handzeichenerkennung bei starker Lärmbelastung
4	Bewegungsanalyse	Inhaltsbasierte Suche nach Bewegungen in Bildfolgen Optimierung von Bewegungsabläufen bei Sportlern Unterstützung bei Choreographie von Tanz und Ballett Klinische Studien bei Orthopädiepatienten Ergonomische Analyse von Bewegungsabläufen
5	Modellbasierte Kodierung	Kompression von Videobilddaten

Tabelle 1.1: *Anwendungen von Systemen zur Beobachtung der Bewegung von Personen; angelehnt an Auflistung Applications of "Looking at People" in [Gav99].*

sonen sind in [Gav99], [AC99], [Wac97] und [Kin94] ausführliche Überblicke gegeben. Hierbei sind jeweils unterschiedliche Kriterien zur Einordnung der betrachteten Arbeiten angesetzt worden.

Kinzel teilt die Ansätze in Kategorien ein, [Kin94]: In der ersten Kategorie sind die Arbeiten verzeichnet, die sich auf Lichtpunkte stützen. Diese Ansätze sind besser unter dem Begriff *moving light display* (MDL) bekannt. Die MDLs gehen auf die Arbeiten von Johansson [Joh73] zurück, in denen gezeigt wird, daß allein durch die Beobachtung der Gelenkpunkte, die z.B. durch kleine Lichtquellen markiert sind, die Bewegung von Personen analysiert werden kann; vgl. zu MDLs auch noch [BL97]. Mit den weiteren Kategorien wird eine Einteilung bezüglich der Verwendung einer binären Vorder- / Hintergrundtrennung, der Bewegungsfreiheit der Kameras und der verwendeten Merkmale vorgenommen. Weiterhin wird betrachtet, ob explizite Bewegungsmodelle vorausgesetzt werden und inwieweit die Ansätze eine Echtzeitfähigkeit aufweisen.

In der Arbeit von Wachter [Wac97] wird zunächst eine Einteilung nach den zu detektierenden Objekten vorgenommen. Hierbei wird zwischen "Fußgänger", "Person"[16] und einzelnen Gliedmaßen, wie Arme und Beine oder der Detektion von Händen unterschieden. Für die dargestellten Ansätze ist angegeben, ob ein 2D oder 3D Modell verwendet wird und wieviele Freiheitsgrade in den Modellen realisiert sind. Weiterhin ist für die Systeme die Anzahl der zur Interpretation zu verwendenden Kameras aufgezeigt. Zusätzlich sind Angaben zur Art der verwendeten Bildmerkmale gemacht und ob diese direkt oder indirekt genutzt werden.

Im Überblick von Gavrila in [Gav99] wird zunächst zwischen 2D- und 3D-Ansätzen unter-

[16] Zur Unterscheidung: ein "Fußgänger" vollführt eine Gehbewegung.

	Anwendung	a priori Modell der Erscheinung	Modellstruktur / Merkmale
A	Bewegungserfassung einzelner Körperteile	nicht verwendet	Strichmännchen
			2D Konturen
		verwendet	Strichmännchen
			2D Konturen
			Volumenkörper

	Anwendung	Anzahl der verwendeten Kameras / Ansichten	Merkmale
B	Verfolgung von Personenbewegung (keine Identifikation einzelner Körperteile)	einzelne Ansicht	Punkte
			2D *Cluster*
		mehrere Perspektiven	Punkte
			blobs
			Volumenkörper

	Anwendung	Erkennungsmethode / Suchraum	Merkmale
C	Erkennung und Deutung der Bewegung des menschlichen Körpers aus Bildfolgen	Mustervergleich (*template matching*)	Punkte
			Maschengitter
		Zustandsräumen (z.B. *hidden Markov model*)	Punkte
			Linien
			blobs

Tabelle 1.2: *Gliederung der Ansätze zur Beobachtung von Bewegung des menschlichen Körpers; zusammengestellt aus Übersicht in [AC99].*

schieden. Die 2D-Ansätze sind weiterhin in Arbeiten mit und ohne explizitem Modell, das die Erscheinung im Bild beschreibt, unterteilt.[17] Bei den 3D-Ansätzen geht Gavrila auf die unterschiedliche 3D Modellierung des menschlichen Körpers in den verschiedenen Arbeiten, auf die Bestimmung der 3D Haltung und Verfolgung, sowie auf das Problem der Korrespondenzfindung der Merkmale ein.

Eine weitreichendere Einteilung von Arbeiten zur Beobachtung der Bewegung des menschlichen Körpers ist von Aggawarl und Cai in [AC99] gegeben, die hier näher vorgestellt werden soll. So wird dort zunächst eine Gliederung nach drei grundlegenden Anwendungsrichtungen vorgenommen; die Ansätze dieser drei Gebiete sind noch weiter unterteilt. Faßt man die Unterteilungen zusammen, so erhält man eine Gliederung für Ansätze zur Beobachtung der Bewegung des menschlichen Körpers entsprechend Tab. 1.2.[18]

Vergleicht man die Unterteilung A – C in Tab. 1.2 mit der Auflistung der Anwendungsgebiete 1 – 5 in Tab. 1.1, so kann man folgende Zuordnung vornehmen: Die Anwendungen aus Anwendungsgebiet 1 (virtual reality) nutzen Bewegungsdaten zur Animation. Hierzu sind reale, natürliche Bewegungen zu erfassen, so daß zunächst eine Korrespondenz zu den Ansätzen in Gruppe A hergestellt werden kann. Wird darüberhinaus auch eine Interaktion mit den virtuellen Systemen realisiert, so stützen sich diese auf die Erkennung von Bewegungen aus Kategorie C.

[17] engl. *explicit shape model*.

[18] *blobs* sind in der Form undefinierte, jedoch zusammenhängende Bereiche, denen ein Attribut zugewiesen ist.

In den Anwendungen zur Überwachung (2) werden meist eine oder mehrere Personen verfolgt, ohne daß die einzelnen Körperteile identifiziert werden müssen. Man kann hier von einer Personendetektion, -verfolgung und -lokalisation sprechen. Diese Anwendungen entsprechen dem Punkt B in Tab. 1.2. Aber auch hier kann es zur Anwendung von Ansätzen aus C kommen, falls den detektierten Personenbewegungen eine Handlung zugeordnet werden soll.

Intelligente Benutzerschnittstellen (3) sind Anwendungen, bei denen generell eine Erkennung von Bewegungen und Handlungen notwendig ist, so daß diese den Ansätzen aus C zuzuordnen sind. Bei der Bewegungsanalyse (4) muß wieder unterschieden werden, ob hierbei nur eine Bewegung z.B. durch Gelenkwinkelverläufe oder Trajektorien erfaßt werden soll (A), oder ob die erfaßten Bewegungen auch Bewegungsmustern zuzuordnen sind (C). Hier wird jedoch auch deutlich, daß die Erkennung und Deutung von Bewegung sich sehr wohl auf einer primären Bewegungserfassung abstützen kann.

Das letzte aufgeführte Anwendungsgebiet 5 ist die modellbasierte Kodierung. Das Ziel ist hierbei, Bilddaten nicht als einzelne Bildpunkte, sondern den Bildinhalt beschreibend zu kodieren, um dadurch bei der Übertragung mit einer geringeren Bandbreite auszukommen. Es kann hierzu die Bewegung einer Person durch ihre Bewegungsdaten kodiert werden, aus denen dann basierend auf einem Personenmodell die Bilder rekonstruiert werden können. Dieser Ansatz korrespondiert zu der Gruppe A in Tab. 1.2. Weiterhin können jedoch bei einer begrenzten Anzahl von zulässigen Bewegungen und Handlungen diese auf der Sendeseite erkannt und auf der Empfangsseite anhand einer parametrierten Beschreibung der Handlung wieder rekonstruiert werden.

Es soll im weiteren nicht auf alle einzelnen in [AC99] vorgestellten und dort den einzelnen Untergruppen der Anwendungsbereiche A – C aufgelisteten Ansätze eingegangen werden. Vielmehr wird auf einige wichtige und auch auf neuere Ansätze und Anwendungen verwiesen. Hierzu wird die Dreiteilung aus Tab. 1.2 übernommen: A – Bewegungserfassung, B – Personendetektion und -verfolgung und C – Bewegungsanalyse und -erkennung.

Bewegungserfassung

Systeme und Ansätze zur Bewegungserfassung arbeiten, insbesondere wenn eine geometrische Rekonstruktion der Bewegung ermöglicht werden soll, mit 2D oder 3D Modellen. Eine der ersten Arbeiten, in der die Verwendung von sog. Strichmännchen (*stick figures*) für die Beschreibung des inneren Zusammenhangs von Modellteilen erläutert wird, ist die Arbeit von Marr und Nishihara [MN78]. Dort wird ebenfalls die Modellierung der einzelnen Modellteile durch 3D Volumenprimitive dargestellt. Bis heute bauen die meisten Systeme auf diesen Ansätzen auf.

Eine weitere grundlegende und viel zitierte Arbeit zur modellbasierten Bewegungserfassung von Menschen stammt von O'Rourke und Badler, [OB80]. Jedoch arbeitete das vorgestellte System mit simulierten Eingabedaten. Um in der geschichtlichen Reihenfolge zu bleiben, sind als nächstes die Arbeiten von Hogg [Hog83] und Akita [Aki84] zu erwähnen. Während Akita einzelne Aufnahmen von gymnastischen Bewegungen analysiert, beobachtet Hogg Bildsequenzen von Fußgängern. Es ist dort das System "WALKER" zur modellbasierten Erfassung der Bewegung von Fußgängern vorgestellt worden, vgl. auch [Hog87]. Auf diesen Arbeiten bauen die Ansätze von Rohr [Roh93, Roh97] auf, in denen ebenfalls eine "gehende" Person detektiert wird. Hierzu wird zum einen vorausgesetzt, daß die Person sich parallel zur Bildebene bewegt und zum anderen, daß die Bewegung einem bestimmten Bewegungsmodell entspricht. Es sind die in [Mur67] vorgestellten Bewegungsdaten für Schulter-, Ellbogen-, Hüft- und Kniegelenk für den gehenden Menschen verwendet worden.

Bei diesen Ansätzen wird für das Modell, entsprechend des Bewegungsmodells, für den

aktuellen Zeitpunkt eine Haltung angenommen. Das so konfigurierte 3D Modell wird in die Bilder projiziert, so daß man 2D Merkmale erhält, die mit Bildmerkmalen verglichen werden können. Dieser 2D/2D Vergleich von projizierten Modellkanten und Bildkanten wird auch in [Wac97, WN97] verwendet. Jedoch wird dort kein expliziter Bewegungsablauf vorausgesetzt, vielmehr wird die Modellkonfiguration, in Form der Freiheitsgrade in den Gelenken, vor der Projektion durch ein iteriertes erweitertes Kalmanfilter geschätzt. Offen bleibt jedoch, wie die initialen Werte für die Filter gesetzt werden.

Entgegen einer Projektion des Modells mit anschließendem Kantenvergleich wird in [LY95] und [HHD98] zunächst der Umriß / die Silhouette der Person im Bild durch eine Vordergrund / Hintergrundtrennung und durch Kantenfilter extrahiert. Beide Ansätze arbeiten mit einem 2D Modell, um aus dem Umriß die Lage einzelner Körperteile zu bestimmen. Ebenfalls über die Umrißkanten arbeitet der Ansatz von [GD95]. Jedoch wird hier ein 3D Modell und entsprechend ein Ansatz mit mehreren Kameras verwendet. Ebenfalls auf Konturen basierend und unter Verwendung von mehreren Kameras werden in den neueren Ansätzen [DF99] und [IOTS99] die Haltungen von Personen bestimmt. Bei dem zuletzt genannten Ansatz sind die drei Kameras jeweils um 90° versetzt anzuordnen, so daß mit einer Vorder-, Seiten- und Draufsicht gearbeitet werden muß.

Um jedoch eine genaue Rekonstruktion von Körperhaltungen aus der Beobachtung mit videobasierten Systemen zu erhalten, können zusätzliche, künstliche Merkmale an den Objekten hilfreich sein. In [JW97] wird z.B. ein 3D Modell mit elf Modellteilen verwendet, wobei 24 Marken verwendet werden, um die Lage der Modellteile aus einem Mehrfachstereoansatz zu rekonstruieren.

Die notwendige Feinheit / Granularität eines Objektmodells für den menschlichen Körper ist zunächst durch die Anwendung bestimmt. Soll für eine Bewegungserkennung nur die Haltung des Oberkörpers, des Kopfes und der Arme zueinander bestimmt werden, so reicht ein Modell aus, das nur aus vier Modellteilen besteht. Vergleiche die Anwendung des PFINDER-Systems des MIT[19] in einem T'ai Chi Trainingssystem, [Bec96]. Dort ist es nur notwendig, den Kopf und die Hände anhand der Hautfarbe zu detektieren.

Soll jedoch eine naturgetreue Animation von Körperbewegungen möglich sein, so müssen wesentlich mehr einzelne Körperteile modelliert werden. Es reicht dann z.B. nicht mehr aus, den Oberkörper mit nur einem Volumenkörper zu modellieren. Dies gilt auch für die Modellierung von Menschen zur Beobachtung von Haltungen und Bewegungen unter ergonomischen Gesichtspunkten. So wird in Menschmodellen, die zur realistischen Vorhersage von Körperhaltungen in CAD-Systemen verwendet werden, wie z.B. RAMSIS[20], die Wirbelsäule nicht nur durch zwei, sondern durch sieben Modellteile modelliert, vgl. [Sei94, Geu94]. In [Sei94] wird jedoch darauf hingewiesen, daß die Anzahl der Gelenke die kinematische Flexibilität des Menschmodells festlege, jedoch die Realisierung aller Gelenke nicht immer notwendig sei, da mit Ersatzkörperelementen mechanisch stark gekoppelte Körperelemente zusammengefaßt würden. Sollen die Haltungen für ein solches CAD-Modell mit videobasierten Systemen ermittelt werden, so ist es notwendig, weitere Gelenke zusammenzufassen. Dies liegt darin begründet, daß jedes einzelne modellierte Modellteil anhand seiner Merkmale im Videobild noch zu detektieren sein muß. Man spricht ansonsten von einem Segmentierungsproblem, wenn die Bildmerkmale im Bild nicht mehr zu vereinzeln sind.

[19] PFINDER = *person finder*, System zur Verfolgung von Personen und zur Interpretation ihrer Handlungen, Massachusetts Institute of Technology, Media Laboratory, [WADP97]. Das im PFINDER-System verwendete Personenmodell ist auf den Unterkörper erweiterbar. Weiterhin ist mit der, als SFINDER bezeichnten Weiterentwicklung, ein Stereoansatz zur Betimmung der Merkmalspositionen bekannt.

[20] RAMSIS = **R**echnergestütztes **A**nthroplogisch-**M**athematisches **S**ystem zur **I**nsassen **S**imulation.

Neben der Granularität des artikularen Modells muß auch noch die Größe und die Ausdehnung der einzelnen Modellteile parametriert werden. Soll die Bewegung einer individuellen Person beobachtet werden, so kann das zu verwendende, individuelle Modell zuvor angepaßt werden, indem die Person vermessen wird. Daß dies auch direkt aus Videosequenzen erreicht werden soll, wird in [PFD99] gezeigt. Dort geht die Modellierung über die einfache Ausdehnung der Volumenkörper hinaus, denn es werden im Modell auch Muskeln, Fett und Haut berücksichtigt. Soll die Bewegung einer beliebigen Person beobachtet werden, die zuvor nicht bekannt ist, so muß ein Objektmodell mit mittlerer Größe angenommen werden. Diese Daten beruhen dann auf anthropometrischen Datenerhebungen, vgl. auch die Angaben in [MK96] zu den Proportionen des menschlichen Körpers.

Personendetektion und -verfolgung

Bei den Anwendungen zur Personendetektion und -verfolgung ist es das Ziel, ein zunächst unbekanntes Objekt als Person zu identifizieren[21] und zu lokalisieren. Die exakte Haltung einer Person soll hier nicht bestimmt werden. Wird für diese Ansätze ein Personenmodell verwendet, so muß dieses daher nicht annähernd so detailliert aufgebaut sein, wie bei Anwendungen zur Bewegungsanalyse.

In Systemen zur Personenverfolgung werden z.B. über eine Bildfolge hinweg die Lokalisationen einer der detektierten Personen zu einer Trajektorie zusammengefaßt. Werden mehrere Personen gleichzeitig erfaßt, so muß zur Unterscheidung der Objekte das Korrespondenzproblem der Zuordnung zwischen Bildmerkmalen eines individuellen Objektes in aufeinanderfolgenden Bildern gelöst werden, damit man einzelne Trajektorien erhält. Die Trajektorien können im einfachen Fall 2D Linien in der Bildebene sein, [BT97b]. Werden die Personen jedoch im 3D Raum lokalisiert, so erhält man 3D Trajektorien, durch die ein Bezug zum observierten Raum hergestellt werden kann.

Die Einteilung in Tab. 1.2 bezüglich der Anzahl der Kameras kann sich bei mehreren Perspektiven zum einen auf Stereoansätze zur Bestimmung von 3D Positionen, zum anderen auf eine Erweiterung des Observierungsbereiches durch die Verwendung mehrerer, sich im Sichtbereich nur knapp überlappender, Kameras beziehen. Zudem läßt sich diese Einteilung um aktive Kamerasysteme erweitern. Mit diesen kann eine Person dann aktiv verfolgt werden. Hierbei sind Schwenk- / Neigekameras und Kameras auf / in mobilen Systemen zu unterscheiden. Die Beobachtung von Personen mit mobilen, d.h. sich bewegenden Kameras ist z.B. bei Fahrerassistenzsystemen zu berücksichtigen. Ansätze zur Detektion von Fußgängern mit Kameras aus PKWs heraus sind in [Kin94] und [GP99] gegeben.

Ein Beispiel für ein System mit einer Schwenk-/Neige-/Zoom-Kamera ist in [YTBH98] dargestellt. Dort wird jedoch nicht jede mögliche Kameraposition angefahren, sondern nur eine ausgewählte Anzahl. Für diese Positionen ist jeweils ein Bild des Hintergrundes in einem Referenzdatensatz abgelegt. Die Detektion des Objektes erfolgt hier durch eine einfache Vorder- / Hintergrundtrennung mittels Bildvergleich. Das einzelne, zu verfolgende Objekt ist daher allein durch die extrahierte 2D Vordergrundregion bestimmt.

Weiterhin sind Kombinationen von feststehenden und aktiven Kameras bekannt. So wird in [PBA98] eine feststehende Kamera zur Vorder- / Hintergrundtrennung verwendet, um die Objekte grob zu lokalisieren. Anschließend wird, basierend auf dem mit der Übersichtskamera bestimmten Sichtstrahl, mit einem aktiven Stereokamerakopf die Person fokussiert und die 3D Position ermittelt. Aufgrund eines Grauwertvergleiches werden die beiden Kameras des Kopfes

[21] Mit Identifikation ist hier die Erkennung der Zugehörigkeit zu einer Objektklasse “Person” gemeint, jedoch nicht eine Personenidentifikation, wie in Zutrittskontrollsystemen.

ständig auf das Objekt fokussiert, um so hinter dem zu verfolgenden Objekt her zu schwenken. Es wird damit eine 3D Trajektorie bestimmt.

Weitere Anwendungen von speziellen Kameras sind mit den omnidirektionalen Kameras bekannt. Bei diesen ist ein 180° Rundumblick durch die Beobachtung eines konvexen Spiegels gegeben. In [OYYT98] wird ein System vorgestellt, das zur Verfolgung von Personen eine solche Kamera verwendet. Hierzu ist ein omnidirektionales Bild als Referenzbild gespeichert worden. Über Differenzbildtechnik werden die Regionen, in denen Objekte aus dem Vordergrund abgebildet sind, ermittelt. Diese Systeme ermöglichen ohne mechanische Schwenk- / Neigeeinrichtung und der damit verbundenen Trägheit eine Verfolgung im ganzen Raum. Durch die Verwendung nur einer CCD-Aufnahmeeinheit für den kompletten Raum ist jedoch die Auflösung und somit die Qualität der entzerrten, perspektivischen Aufnahmen der verfolgten Person geringer. Eine Kombination von omnidirektionaler Kamera und einer weiteren Schwenk- / Neigekamera zum Verfolgen von Personen wird daher in [CHG98] dargestellt. Zum Verfolgen werden mit einem Farbklassifikator hautfarbene Regionen gesucht, die als Position des Kopfes interpretiert werden. Durch die Kombination des Rundumblicks und einer weiteren Kamera kann das System Verdeckungen aufgrund anderer Objekte oder anderer Hindernisse handhaben.

Ebenfalls mit dem Merkmal der Hautfarbe arbeitet der in [DGW+98] dargestellte Ansatz, um zunächst hautfarbene Regionen im Bild zu finden. Mit Methoden zur Suche von Mustern werden aus diesen Bildregionen die Regionen ausgewählt, in denen Gesichter abgebildet sind. Zusätzlich werden mit einem Stereoansatz Tiefeninformationen aus den Bildpaaren gewonnen. Hierüber werden nur die als Köpfe detektierten Bereiche weiter betrachtet, die in einem festgelegten Abstand vor dem System stehen. Das System ist für eine spielerische Mensch-Maschine-Interaktion entwickelt. Hierbei wird das Gesicht einer Person ausgewählt und verfolgt. Eines der aufgenommenen Kamerabilder wird auf einem Monitor ausgegeben, wobei jedoch die dem verfolgten Kopf zugehörige Bildregion verzerrt dargestellt ist; es wird daher von einem virtuellen Spiegel gesprochen.

Der Ansatz von Cai und Aggawarl [CA96] zur Verfolgung von Personen basiert wiederum auf einer Vorder- / Hintergrundtrennung durch Vergleich mit einem Referenzbild. Es wird jedoch zusätzlich ein einfaches 2D Personenmodell verwendet, um andere sich bewegende Objekte von Personen zu unterscheiden. Weiterhin wird angenommen, daß bei einer aufrecht gehenden Person der Kopf als eine Ellipse mit einem Achsenverhältnis von 1:1,5 im oberen Teil der Vordergrundbildregion abgebildet ist. Unterhalb dieser elliptischen Region wird eine rechteckige Region erwartet, die den Oberkörper repräsentiert. Die Größen der Regionen für Kopf und Oberkörper müssen hierbei in einem bestimmten Verhältnis stehen, damit die Gesamtregion als Abbild einer Person bezeichnet wird. Entlang der senkrechten Mittelachse der 2D Region werden Punkte ausgewählt, deren 2D Positionen in einem Vektor der geometrischen Merkmale des detektierten Objektes zusammengefaßt werden. Zusätzlich werden noch die Grauwerte der Punkte in einem Vektor der visuellen Merkmale zusammengefaßt.

Zur Verfolgung der Personen werden die Korrespondenzen für die entsprechenden Regionen in zwei aufeinanderfolgenden Bildern einer Sequenz über einen Vergleich der Merkmalsvektoren vorgenommen. Hierzu werden die Summen der Mahalanobis-Abstände ermittelt und die Zuordnung entsprechend der Bayes'schen Entscheidungstheorie getroffen. Für eine Erweiterung des Systems auf mehrere Kameras wird eine 3D Kalibrierung des Gesamtsystems vorausgesetzt, so daß zu den einzelnen Merkmalspunkten, die auf der Mittelachse der detektierten Personenregion liegen, Sichtstrahlen in das Bild einer weiteren Kamera projiziert werden können. Für diese Strahlen werden, ähnlich wie bei der Verfolgung eines Objektes innerhalb der Bildsequenz einer Kamera, die Mahalanobis-Abstände zu den Merkmalsvektoren der extrahierten Regionen aus der weiteren Kamera ermittelt und bewertet. Ist das Objekt in beiden Kameras

sichtbar, so ergibt sich für die Korrespondenz ein entsprechend geringer Abstand des normalisierten Mahalanobis-Abstandes. Die Erweiterung des Ansatzes um eine aktive Umschaltung auf weitere Kameras bei der Verfolgung ist in [CA98] beschrieben. Dort wird zur Verfolgung jeweils die Kamera ausgewählt, die entweder mit dem Sichtbereich direkt an den Sichtbereich der aktuellen Kamera angrenzt, für die Objektverfolgung die beste Konfidenz aufweist oder bei der zu erwarten ist, daß die zu verfolgende Person über die größte Anzahl von aufeinanderfolgender Bilder sichtbar bleibt.

In [NKI98] ist das Personenmodell noch weiter vereinfacht: Es wird für den kompletten Körper nur nach einer Ellipse gesucht. Hierbei werden zudem Annahmen über die Bewegung der zu verfolgenden Personen getroffen. Es wird vorausgesetzt, daß die Person zwischen zwei Aufnahmen ihre Position mit einer exakt festgelegten Geschwindigkeit verändert hat. Auf dieser Grundlage werden für eine zu verfolgende Person mögliche Positionen von Ellipsen vorhergesagt, die dann als Muster mit Regionen aus einer Vorder- / Hintergrundtrennung verglichen werden. Durch die Verwendung von mehreren Systemen, auf denen jeweils ein einzelner Verfolgungsprozeß abgearbeitet wird, ist das System zur Beobachtung eines größeren Bereiches vorgestellt worden. Es bleibt jedoch unklar, in welcher Form die Informationen zur "Übergabe" eines Objektes an ein anderes System über das Netzwerk des verteilten Gesamtsystems weitergegeben wird.

Eine Kombination von Personenverfolgung und der Beobachtung von transportierten Gegenständen ist in [HCHD99] vorgestellt. Dort wird zunächst über eine Vordergrundregion die Silhouette der zu verfolgenden Personen ermittelt. In diesen Regionen wird die Mittelachse bestimmt, so daß Aussagen über die Symmetrie gemacht werden können. Unter der Annahme, daß Personen generell bezüglich der Mittelachse symmetrisch sind, werden alle Bereiche, die über eine symmetrische Kernregion herausragen, als Bereich betrachtet, in denen Objekte abgebildet sind, die von Personen getragen oder transportiert werden. Das Ziel des Systems ist es, hierüber festzustellen, ob Personen Objekte aufnehmen, Objekte abstellen oder ob zwei Personen Objekte austauschen. Diese Interpretation von Handlungen gehen über die eigentliche Personendetektion und -verfolgung hinaus und sind Gegenstand des folgenden Abschnittes.

Bewegungsanalyse und -erkennung

Der dritte Bereich aus der Einteilung der Ansätze zur Beobachtung der Bewegung des menschlichen Körpers entsprechend Tab. 1.2 ist die Interpretation der Bewegungen. Dies kann zum einen die Analyse von erfaßten Bewegungsabläufen sein, zum anderen auch die Deutung der Trajektorie, die bei der Personendetektion und -verfolgung ermittelt worden ist. Das Ziel ist dabei immer, den beobachteten Abläufen eine Bedeutung zu attribuieren, um eine Handlung zu erkennen. Oftmals wird hier von der Erkennung von Gesten gesprochen.[22] Ein verwandter Themenbereich ist die Gesichtserkennung[23], mit Ansätzen zur Detektion von Gesichtern und zur Identifikation von Personen. Ein dritter Bereich ist die Erkennung von Handzeichen und Zeichensprache. Über die Forschungsrichtungen dieser drei Bereiche sind Überblicke in [Dau97, Wex97, Edw97, Nak98] gegeben.

Stellvertretend für die Ansätze aus diesen Bereichen sollen hier einige wenige Arbeiten kurz umrissen werden, die direkt auf Ansätzen zur Bewegungserfassung bei Personen und zur Personendetektion aufsetzten. So wird in [HR95] eine natürlich sprachliche Beschreibung von Szeneninhalten angestrebt. Der Ansatz baut auf den schon erwähnten Arbeiten von Rohr [Roh93] zur Detektion von Fußgängern auf. Eine der genannten Szenenbeschreibungen ist daher: "*Auf*

[22] engl. *gesture recognition.*
[23] engl. *face recognition.*

der Straße in etwa rechts vor dem Osttrakt befindet sich ein Fußgänger. Er geht über die Straße." Ähnliche Beschreibungen von Personenbewegungen in Szenen sind in [BT97a] angegeben. Dort werden einzelne sog. Szenarien von Personenbewegungen erkannt. Dies sind z.B. "*Die Person geht geradeaus.*" oder "*Die Person ändert ihre Bewegungsrichtung.*". Diese einzelnen Szenarien werden als Zustände eines Automaten betrachtet, mit dem eine Gesamthandlung erkannt werden soll. Wird hierbei ein bestimmter Zustand erreicht, so kann Alarm ausgelöst werden, wenn damit eine "verdächtige" Handlung bestimmt wurde. Als Beispiel ist hierzu die Überwachung eines Parkplatzes angegeben, wobei davon ausgegangen wird, daß Personen, die zu ihrem eigenen Wagen gehen, ein anderes Bewegungsverhalten haben als Personen, die Fahrzeuge entwenden oder beschädigen.

In [ACF99] wird zur Bewegungserfassung von Personen ein 3D Modell verwendet, mit dem der Oberkörper, der Kopf und die Arme berücksichtigt sind. Über einen Stereoansatz werden die 3D Positionen der Hände und des Kopfes ermittelt. Aus der Abfolge der für die Hände bestimmten 3D Positionen werden mit einem Entscheidungsbaum die Bewegungen gedeutet. Die einzelnen Gesten / Signale, die das System erkennen kann, sind dabei Kombinationen aus Haltungen und Bewegungen der Arme. So wird z.B. als eine der Haltungen bestimmt, ob ein Arm nach oben oder nach unten gehalten wird. Ein Beispiel für ein dynamisches Signal ist das Kreisenlassen der Arme, wobei zusätzlich noch die Geschwindigkeit der Kreisbewegungen ermittelt und gedeutet wird. Ähnliche Bewegungen der Arme werden in [BBBC98] erfaßt. Die dort als Posen bezeichneten Signale dienen in einer Mensch-Maschine-Kommunikation zum Steuern einer mobilen Plattform, wobei die Plattform selbst das videobasierte Erkennungssystem enthält.

Bei dem Ansatz in [Bre97] ist das Ziel nicht die Deutung von Bewegungen als Signale oder Gesten, sondern die Erkennung der Bewegung an sich. So wird dort über die Detektion der Beine einer beobachteten Person entschieden, ob die Person z.B. geht oder springt. Daß eine erkannte Bewegung auch in den Prozeß der Bewegungserfassung zurückgekoppelt werden kann, wird in [WP99] gezeigt. Denn wie schon im Abschnitt zur Bewegungserfassung erläutert, können Bewegungsvorhersagen genutzt werden, um die Zuordnung von Bild zu Modellmerkmalen zu verbessern. Ist nur ein Bewegungsmodell bekannt und wird dieses vorausgesetzt, so kann nur die Bewegung der Personen erfaßt werden, die diesem Modell entspricht. Läßt das System jedoch eine frühzeitige Erkennung einer periodischen Bewegung zu, so kann auf zuvor gespeicherte Bewegungsmuster / Bewegungsdaten zur Verbesserung der Bewegungserfassung zurückgegriffen werden. Gezeigt ist dies für die Verfolgung der Bewegungen der Hände. Beide Ansätze nutzen, wie auch das schon erwähnte System des T'ai Chi Trainers [Bec96], *hidden Markov models* zur Erkennung der Bewegungen. Dies gilt ebenso für den Ansatz in [EKR98]. Dort sind für eine bestimmte Anzahl von Handbewegungen sog. Filtermodelle angelegt worden. Somit können diese Bewegungen klassifiziert werden, wobei alle weiteren Bewegungen als unbekannt zurückgewiesen werden.

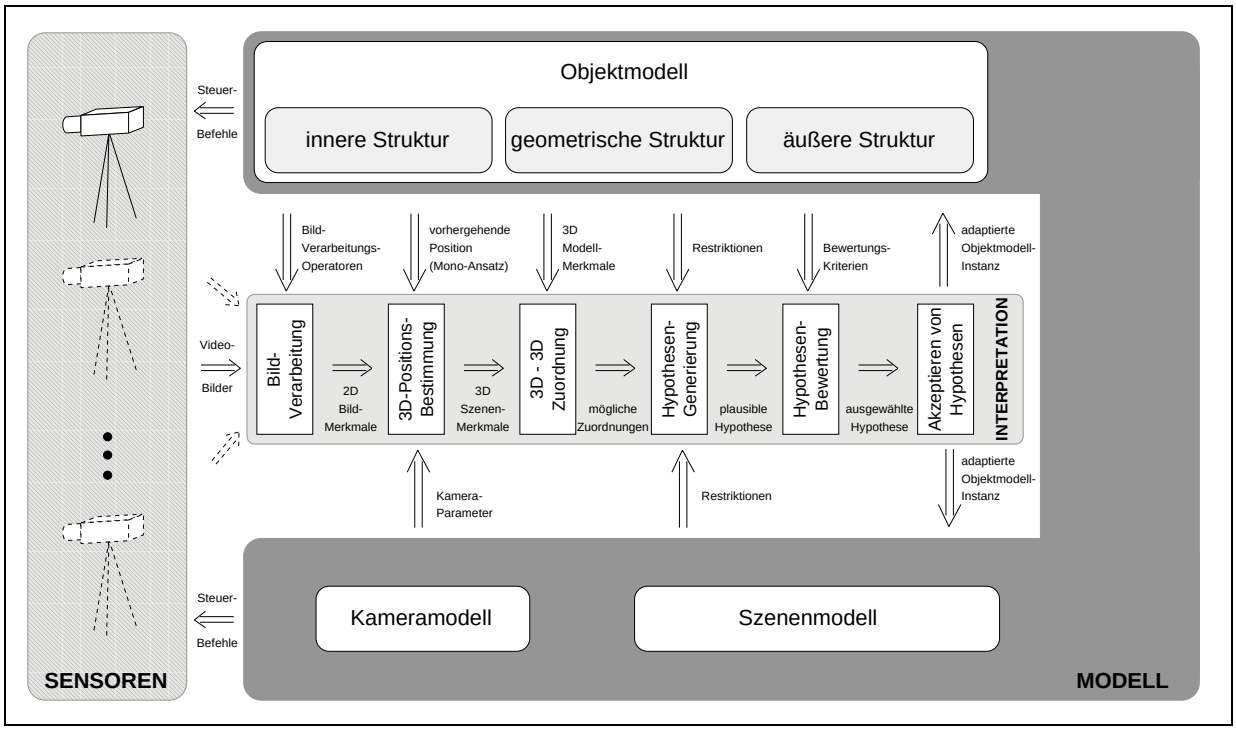


Abbildung 1.3: *Überblick über die Systemstruktur von* STABIL^{++}.

1.4 Das System STABIL^{++}

1.4.1 Systemüberblick

Das vorzustellende Konzept zur videobasierten Beobachtung artikularer Bewegung gliedert sich generell in zwei Hauptteile: Dies ist zum einen der Bereich der Modellierung und zum anderen der eigentliche Interpretationsprozeß, der sich auf die Modellierung abstützt. Entsprechend der Umsetzung des Konzeptes in das System STABIL^{++} ist in Abb. 1.3 die Systemstruktur dargestellt.

Der Interpretationsprozeß ist eingebettet in die Modellierung, wobei dieser generell von dem Objektmodell initiiert wird, sich jedoch auch auf ein Kameramodell und ein Szenenmodell abstützt. Das Objektmodell ist entsprechend der Modellierung artikularer Objekte aus einzelnen 3D Objektmodellteilen aufgebaut, die mit Gelenken verbunden sind. Entsprechend einer Dreiteilung der Modellstruktur in die innere, die geometrische und die äußere Struktur, werden diese verschiedenen Aspekte der Modellierung in den Interpretationsprozeß eingebracht.

Bei der Interpretation wird zunächst von einer zu verfolgenden Objektmodellinstanz aus den zur Verfügung stehenden Sensoren / Kameras die zu verwendenden Kameras ausgewählt und bei aktiven Kamerasystemen diese zusätzlich angesteuert. Diese Auswahl basiert auf 3D Informationen. Daher ist in STABIL^{++} eine durchgängige 3D Modellierung realisiert. Mit dem Kameramodell sind, neben der Modellierung lageunabhängiger Abbildungseigenschaften der Kameras, auch die 3D Lage der Kameras in der Szene bekannt.

Basierend auf Bildverarbeitungsoperatoren, die durch das Objektmodell bestimmt sind, werden zur Interpretation zunächst aus den aufgenommenen Bildern die Bildmerkmale extrahiert. Der nächste Schritt ist der Übergang zu 3D Szenenmerkmalen, denn das Struktur-Vergleichsverfahren nimmt einen 3D/3D Vergleich von Szenenmerkmalen mit Modellmerk-

malen vor. Für den 2D/3D Übergang wird in STABIL++ zwischen einer monokularen Tiefenschätzung und einem Stereoansatz unterschieden. Bei der Tiefenschätzung wird zusätzliches Modellwissen verwendet, um die 3D Position der Merkmale zu ermitteln, wogegen beim Stereoansatz der Schnittpunkt von Sichtstrahlen zueinander gehörender Bildmerkmale aus Bildern von zwei oder mehreren Kameras ermittelt wird. Die Anzahl der Kameras kann in dem System anwendungsabhängig variiert werden.

Aus den durch den Vergleich von Szenenmerkmalen und Modellmerkmalen ermittelten Zuordnungen werden Hypothesen unter der Verwendung von Restriktionen, die sich aus dem Objektmodell und aus dem Szenenmodell ergeben, generiert. Zur Beurteilung der Hypothesen und zur anschließenden Auswahl einer Hypothese als Interpretation der Szene, wird weiteres Wissen aus dem Objektmodell verwendet. Mit dem letzten Schritt des Interpretationsprozesses wird in STABIL++ eine Adaption der Objektmodellinstanz vorgenommen. Hiermit wird berücksichtigt, daß es sich z.B. bei der Beobachtung von Personen nicht um die Detektion und Verfolgung von exakt bekannten und vermessenen Objekten handelt.

Bevor in Kap. 2 die verwendete Modellierung und in Kap. 3 der Interpretationsprozeß erläutert wird, wird im folgenden Abschnitt eine Abgrenzung und Einordnung des realisierten Konzeptes zu den im letzten Abschnitt vorgestellten Arbeiten vorgenommen.

1.4.2 Abgrenzung und Einordnung

In dem Bericht von Ponce et al. in [PHZ96] zur Objektrepräsentation wird deutlich gemacht, daß mit Bildinterpretationssystemen Aufgaben zu lösen sind, die komplexer sind, als andere Aufgaben aus dem Bereich der künstlichen Intelligenz. Hiermit wird begründet, daß entsprechend der komplexen und vielseitigen Anforderungen an die verschiedenen videobasierten Interpretationsaufgaben jeweils neue, den Bedingungen entsprechende, Objektrepräsentationen und Modelle entwickelt werden. Vergleicht man die grundlegend verschiedenen Aufgaben zur automatischen Schrifterkennung und zur Bewegungserfassung von artikularen Objekten, so kann dies nachvollzogen werden. Jedoch sind die vorgestellten Ansätze zur Beobachtung artikularer Bewegung jeweils auch spezielle Lösungen für eine bestimmte Aufgabe.

Das Objektmodell, wie es in dem vorzustellenden Konzept zum Einsatz kommt, ist demgegenüber ein generisches Modell zur generellen Beschreibung von artikularen Objekten. Entsprechend der inneren Modellstruktur können eine beliebige Anzahl von Objektmodellteilen in einer hierarchischen Ordnung gekoppelt sein. Die Konfiguration des Modells[24] wird über die geometrische 3D Struktur anhand von Transformationen zwischen lokalen Koordinatensystemen in den Objektmodellteilen abgebildet. Die Modellierung ist damit eine *object-centred representation*, vgl. Kap. 1.3 und [Pop94]. Die äußere Erscheinung, d.h. die Ausdehnung der einzelnen Objektmodellteile und deren verschiedenartige Erscheinung wird in der äußeren Objektmodellstruktur berücksichtigt.

Bei der Modellierung des menschlichen Körpers, als ein Beispiel für ein artikulares Objekt, kann daher entsprechend der Anwendung die Anzahl der Objektmodellteile variiert werden. Ebenso können verschiedenartige, den Anforderungen und Umgebungsbedingungen entsprechende Modellmerkmale den Objektmodellteilen zugeordnet werden. Auf Grund dieser Konzeptionierung ist es möglich, das realisierte System STABIL++ zur Bewegungserfassung und zur Personendetektion und -verfolgung anzuwenden. Das System ist daher sowohl dem Bereich A und dem Bereich B in der Einteilung der verschiedenen Ansätze zur Beobachtung der Bewegung von Personen in Tab. 1.2 zuzuordnen.

[24] Bei Personenmodellen kann von der Haltung des Modells gesprochen werden.

Betrachtet man die weitere Unterteilung der Ansätze zur Bewegungserfassung (A) in die Verwendung von a priori Wissen über die Erscheinung der Objekte, so ist STABIL^{++} auch hier beiden Gruppen zuzuordnen. Für eine initiale Detektion eines Objektes und seiner Haltung wird mit dem vorzustellenden Konzept generell keine spezielle Konfiguration vorausgesetzt. Für jeden weiteren Interpretationsschritt wird jedoch das Wissen der zuvor ermittelten Konfiguration des Modells genutzt. Wie in der Übersicht von Aggawarl und Cai [AC99] angemerkt, wird hierdurch das Problem der Korrespondenzfindung von Modellmerkmalen zu Szenenmerkmalen bei einer Re-Detektion vereinfacht. Es wird hierzu jedoch kein Bewegungsmodell vorausgesetzt, mit dem die Haltungen entsprechend eines bekannten Bewegungsablaufes vorhergesagt werden oder eine feste Objektgeschwindigkeit vorausgesetzt. Vielmehr wird für jedes Objektmodellteil unabhängig ein 3D Suchraum vorhergesagt. Die Vorhersage stützt sich dabei auf die bei der Verarbeitung der Bildfolgen gewonnenen 3D Lageinformationen der Objektmodellteile. Basierend auf den Vorhersagen können bei der Verdeckung einzelner Merkmale die 3D Merkmalspositionen geschätzt werden. Daher läßt sich im Interpretationsprozeß bei der Korrespondenzfindung auch bei Verdeckungen eine gültige Zuordnung finden. Der Problematik der Verdeckung von Merkmalen und Objektmodellteilen kann in STABIL^{++} jedoch auch durch die Verwendung von einer beliebigen Anzahl von Kameras mit unterschiedlichen Ansichten begegnet werden.

Aufgrund der Dreiteilung der gewählten Objektmodellstruktur trifft das vorzustellende Konzept auch auf mehrere der weiteren Unterpunkte zur Modellstruktur / Merkmale der Anwendungen der Gruppe A aus Tab. 1.2 zu. So ist mit der inneren Objektmodellstruktur, die allgemein als “Strichmännchen” bekannte Struktur realisiert, mit der der innere Zusammenhang der Objektmodellteile beschrieben ist. Ebenso sind 3D Volumenkörper in der äußeren Objektmodellstruktur berücksichtigt. Durch eine entsprechende Wahl von Merkmalen, die den einzelnen Objektmodellteilen zugeordnet werden, können ebenfalls 2D Konturen berücksichtigt werden, wobei sich diese aus den projizierten 3D Volumenkörpern ergeben.

Für die Unterteilung des zweiten Anwendungsfeldes zur Verfolgung von Personenbewegung (B) gilt ähnliches: Bei dem vorzustellenden Konzept ist die Anzahl der Kameras generell nicht beschränkt. Wird mehr als eine Kamera eingesetzt, so sind zwei Fälle zu unterscheiden. Zum einen können die Kameras den gleichen Szenenausschnitt abbilden. Zum anderen können die Kameras so angeordnet sein, daß der zu beobachtende Szenenbereich vergrößert wird. Sobald korrespondierende Bildmerkmale in mindestens zwei Kameras abgebildet werden, wird die entsprechende 3D Position über einen Stereoansatz ermittelt. Ist das zu detektierende Objekt nur in einer Kamera sichtbar, so wird die Tiefeninformation bei der Gewinnung von 3D Positionen geschätzt. Diese Schätzungen stützen sich hierbei auf Modellwissen ab.

Diese Flexibilität bei der Auswahl der zur Interpretation zu verwendenden Kameras wird durch die Verwendung eines expliziten Szenenmodells und der damit verbundenen konsequenten 3D Modellierung erreicht. So sind dem Szenenmodell, neben einer hinreichenden Beschreibung der Umgebung, auch die Positionen und Orientierungen der zur Verfügung stehenden Kameras bekannt. Weiterhin können die 3D Informationen über die Positionen von Objekten genutzt werden um aktive Kamerasysteme so zu positionieren, daß die Objekte zur Detektion und Verfolgung optimal erfaßt werden. Bei einer Anwendung zur Personenverfolgung ist es daher selbstverständlich, daß die Trajektorie, die die Objektbewegung wiedergibt, nicht nur eine 2D Linie in der Bildebene ist, sondern den durchschrittenen Weg im 3D Raum darstellt.

Obwohl die Detektion einer Objektmodellinstanz durch diese selbst initiiert wird, verwaltet das Szenenmodell die Instanzen. Hiermit ist realisiert, daß zum einen mehrere Objekte gleichzeitig detektiert und verfolgt werden können. Die Anzahl der Instanzen ist hierbei zunächst unbeschränkt, jedoch müssen in den Kamerabildern die Bildmerkmale vereinzelt werden können. Zum anderen kann über das Szenenmodell eine Objektmodellinstanz für eine aktive Verfolgung

Anwendung	Anwendungs-gebiet	Objektmodell	Merkmale (p: primär, s: sekundär)	3D Tiefen-information
3D Personen-detektion, Lokalisation u. Verfolgung	Sicherheits-technik	"Kopf", "Oberkörper" und "Beine"	p: "hautfarbener Kopf" s: "Vorhandensein des Rumpfes" s: "Füße auf dem Boden"	(Multi-) Mono und Stereo-Ansatz
3D Bewegungs-erfassung und Gelenkwinkel-bestimmung	Ergonomie, Sportmedizin, Tiermedizin, und Virtual-reality	Modellteile entsprechend Struktur des kompletten Lebewesens	p: Landmarken (Markierung der Gelenke)	bevorzugt (Multi-) Stereo-Ansatz

Tabelle 1.3: *Beispielhafte Konfigurationen und Anwendungen von* STABIL++*.*

ausgewählt werden.

Die Modellmerkmale, mit denen die Erscheinung jedes Objektmodellteils in der äußeren Struktur bestimmt wird, sind in der gewählten generischen Modellierung austauschbar. So kann das vorzustellende Konzept auch bei der weiteren Unterteilung der Gruppe B der Anwendungen in der Tab. 1.2 bezüglich der Merkmale in verschiedene Gruppen eingeteilt werden. Ein Modellmerkmal muß lediglich ermöglichen, die Position eines Objektmodellteiles zu lokalisieren. Es ist sicherzustellen, daß die korrespondierenden Bildmerkmale aus den Kamerabildern jedoch noch zu extrahieren sind. Aus dieser Bedingung heraus ist die Anwendbarkeit der Modellierung auf artikulare Objekte beschränkt, für deren Objektmodellteile Merkmale definiert werden können, die sich noch einzeln extrahieren lassen.

Das Konzept läßt, neben den Merkmalen, mit denen die Erscheinung der Objektmodellteile beschrieben ist, noch sog. sekundäre Merkmale zu. Die sekundären Merkmale sind Heuristiken, die für jedes Objektmodellteil definiert werden können. Diese Heuristiken werden neben den Restriktionen, die sich aus der generellen Objektmodellstruktur ergeben, bei der Beurteilung von Hypothesen verwendet. Die Art der zu verwendenden sekundären Merkmale, die Beschreibung der Erscheinungsform über die weiteren, primären Merkmale und die generelle Wahl der Anzahl der Objektmodellteile sind anwendungsabhängig. In der Tab. 1.3 sind für zwei beispielhafte Anwendungen in zwei unterschiedlichen Anwendungsgebieten Konfigurationen von STABIL++ skizziert.[25]

Das vorzustellende Konzept und somit auch das System STABIL++ decken den dritten Bereich der Anwendungen in der Tab. 1.2 nicht ab. Vielmehr kann das System als Grundlage für weitere Anwendungen zur Erkennung und Deutung von Bewegungen menschlicher Körper dienen. Hierzu kann das System bei einer Anwendung zur Verfolgung der Personenbewegung die 3D Trajektorie des zurückgelegten Weges stellen und bei der Bewegungserfassung einzelner Körperteile die Trajektorien der einzelnen Objektmodellteile oder die Gelenkwinkelverläufe in den Knoten zwischen den einzelnen Objektmodellteilen zur Verfügung stellen.

[25] Für eine genauere Erläuterung der Konfigurationen sei auf die im Kap. 4 beschriebenen Anwendungen verwiesen.

2 Modellierung

2.1 Einführung

Die Detektion und Verfolgung von Objekten in Videobildfolgen basieren entsprechend der dargestellten Motivation auf einer exakten Modellvorstellung. Diese Modellvorstellung geht über das *Objektmodell*, das den Interpretationsprozeß steuert, hinaus. So ist dem System mit einem *Szenenmodell* eine Beschreibung der Umgebung bekannt. Nachdem STABIL^{++} als Sensor Videokameras verwendet, wird die Gewinnung von 3D Informationen aus den Bildern auf ein *Kameramodell* gestützt, daß die Abbildung der realen 3D Welt in das 2D Bild beschreibt. Vergleiche als Überblick zur verwendeten Modellierung auch die Darstellung der Systemstruktur in Abb. 1.3.

In den folgenden Abschnitten werden die Modelle beschrieben. Zusätzlich zum eigentlichen Abschnitt zum Objektmodell wird noch gesondert auf die *Merkmale* des Objektmodells eingegangen. Ebenso wird zusätzlich zum eigentlichen *Kameramodell*, dem Modell einer *Lochkamera*, noch auf die Einteilung verschiedener Kameratypen und deren Eigenschaften im Hinblick auf die Integration in das System eingegangen.

2.2 Szenenmodell

Das Szenenmodell steuert die Abläufe der Detektion und der Verfolgung von Objekten und kapselt Wissen über die Umgebung, in der die Detektion durchgeführt wird. Das Szenenmodell ist somit ein Abbild einer realen Szene, die interpretiert werden soll. Die Modellierung ist auf die Komponenten beschränkt, die für den Interpretationsprozeß benötigt werden. Die einzelnen Komponenten des Szenenmodells werden als *Inventar* bezeichnet. Das Szenenmodell ist als das Tupel

$$scene = \langle SSP_s, obj_0, SSP^0, OBJ, wcs, CAM, ACT \rangle \tag{2.1}$$

definiert. Die einzelnen Komponenten sind

- der Observierungsraum SSP_s,
- das initiale Objektmodell obj_0,
- initiale Modellsuchräume SSP^0,
- alle gefundenen Objektmodellinstanzen OBJ,
- ein Weltkoordinatensystem wcs,
- alle zur Verfügung stehenden Kameras CAM und
- alle Akteure ACT.

Observierungsraum

Hierbei ist $SSP_s = \{wr_1, \ldots, wr_n\}$ eine Beschreibung des 3D Observierungsraumes, in dem Objekte erwartet werden. Der Observierungsraum setzt sich aus sog. 3D Weltregionen $wr_i \in \mathbb{R}^3$ zusammen. Die können zunächst beliebige Volumenkörper sein.

Für die Beschreibung des Observierungsraumes bei der Detektion in Innenräumen werden typischerweise die Flächen bestimmt, die der Lage der Wände, Türen und Fenster entsprechen. Daher werden als Weltregionen auch Flächen zugelassen, die im 3D Raum definiert sind. Beim Interpretationsprozeß muß sichergestellt werden, daß nur Objekte detektiert werden können, die sich innerhalb des Observierungsbereichs befinden. Dies bedeutet, daß sich keine den Observierungsbereich begrenzende Fläche zwischen der 3D Position der Kamera und der 3D Position des Objektes befinden darf. Es ist eine Definition des Observierungsraumes weiterhin über Volumenkörper möglich, jedoch ist die Bestimmung, ob ein 3D Punkt innerhalb von nicht rein konvexen Volumenkörpern liegt, algorithmisch aufwendig. Daher wird eine Definition über die begrenzenden Flächen der Innenräume bevorzugt, bei der nur überprüft werden muß, ob ein 3D Sichtstrahl von der Kamera eine der Flächen schneidet.

Initiale und gefundene Modelle

Mit obj_0 ist dem Szenenmodell ein initiales Objektmodell bekannt, mit dem beschrieben ist, anhand welcher Merkmale im Interpretationsprozeß Objekte zu detektieren sind.[1] Für eine initiale Detektion muß dem System ein 3D Suchraum vorgegeben sein, in denen Objekte erwartet werden. Dem Szenenmodell werden mit $SSP^0 = \{SSP_1^0, \ldots, SSP_n^0\}$ eine oder mehrere 3D Suchräume angegeben. In diesen Räumen werden die Merkmale des initialen Objektmodells erwartet. Ein einzelner Suchraum, z.B. SSP_1^0, setzt sich aus mehreren Suchraumkugeln zusammen. Um einen einzelnen Suchraum für das initiale Objektmodell zu bestimmen, wird jeweils eine initiale 3D Position des Objektmodells und ein initialer Radius für die Suchraumkugeln angegeben. Daher ist der initiale Suchraum SSP^0 durch $\{(\vec{p}_1, r_1), \ldots, (\vec{p}_n, r_n)\}$ und die Struktur des initialen Objektmodells bestimmt. Hierbei ist $\vec{p}_i = [x_i, y_i, z_i]^T$ eine 3D Position, mit dem das initiale Objektmodell positioniert wird und r_i der initiale Radius für die Bestimmung der Suchräume der Merkmale des Objektmodells. Durch einen genügend großen Radius kann auch im gesamten Observierungsraum des Szenenmodells nach Objekten gesucht werden. Ansonsten werden die initialen Modelle sinnvollerweise an Bereiche plaziert, an denen Objekte die Szene erscheinen können. Es ist jedoch zu beachten, daß der eigentliche Raum, in dem observiert wird, neben dem Observierungsraum SSP_s durch die Lage, d.h. die Position und Orientierung und durch die Blickwinkel der im System verfügbaren Kameras eingeschränkt ist.

Alle initial detektierten Objektmodelle werden in dem Szenenmodell in einer Liste von aktuell gefundenen Objekten, den sog. Objektmodellinstanzen in $OBJ = \{obj_1, \ldots, obj_n\}$ verwaltet. Bei jedem weiteren Interpretationszyklus wird zunächst versucht, alle bisher gefundenen Objektmodellinstanzen wiederzufinden. Danach wird nach weiteren, neuen Objekten gesucht. Die Liste der gefundenen Objekte kann auf eine Objektmodellinstanz beschränkt werden, wenn das System aktiv ein Objekt verfolgen soll. In diesem Verfolgungsmodus wird im Interpretationszyklus nicht nach weiteren Objekten gesucht.

[1] Die Struktur von STABIL^{++} ist so ausgelegt, daß auch mehrere verschiedene initiale Objektmodelle verwendet werden können, jedoch ist die Beschreibung des Interpretationsprozeßes in Kap. 3 auf ein initiales Objektmodell ausgelegt.

Weltkoordinatensystem

Damit eine 3D Lagebestimmung der Objekte in der Szene möglich ist, ist in dem Szenenmodell ein kartesisches Koordinatensystem wcs als Referenzsystem festgelegt. Die Lage dieses Weltkoordinatensystems in der realen Welt ist zunächst willkürlich. Es ist jedoch sinnvoll, das Weltkoordinatensystem parallel zu den doch meist rechtwinkligen Wänden von Gebäuden zu legen. Für die Orientierung wird vorausgesetzt, daß die vom Koordinatensystem aufgespannte xy-Ebene parallel zum Boden des Observierungsraumes ist. Es ist weiter vorausgesetzt, daß bei der Bestimmung der 3D Position in einem monokularen Aufbau die Höhe des Bodens $z = 0$ ist. Werden in dem Szenenmodell mehrere Räume[2] verwaltet, wodurch sich der Observierungsraum der Szene praktisch aus mehreren einzelnen Observierungsräumen zusammensetzt, so sind alle Weltregionen auf das gleiche Weltkoordinatensystem bezogen.

Die Lagen und Suchräume der Objektmodellinstanzen und initialen Objektmodelle sind ebenfalls in dem Weltkoordinatensystem angegeben. Dies gilt auch für die Lagen der in dem Szenenmodell bekannten Kameras.[3]

Kameras

Als Sensoren für die Bildinterpretation stehen dem Szenenmodell mit $CAM = \{cam_1, \ldots, cam_n\}$ Kameras zur Verfügung. Von allen Kameras sind die Abbildungseigenschaften bekannt. Hierzu wird die Abbildung durch das Modell der Lochkamera angenähert. Durch eine Kamerakalibrierung werden innere Kameraparameter bestimmt, die die Brennweite des Objektives, die Verzerrung des Objektives, deren Hauptpunkt und die Chipgröße berücksichtigen. Im weiteren werden die externen Kameraparameter bestimmt, die die Lage der Kamera im Weltkoordinatensystem angeben. Bei aktiven Schwenk-/Neige- und/oder Zoom-Kameras muß zum Zeitpunkt der Aufnahme jeweils eine gültige Kalibrierung vorliegen.[4]

Akteure

Mit $ACT = \{act_1, \ldots, act_n\}$ sind dem Szenenmodell noch eine Reihe von sog. Akteuren bekannt. Den Akteuren wird jeweils am Ende eines Interpretationszyklus der Zugriff auf die aktuell gefundenen Objektmodellinstanzen und deren 3D Daten ermöglicht. Ein Akteur kann im einfachsten Fall die 3D Lagedaten in einer Datei speichern oder auch feststellen, ob ein Objekt, z.B. eine Person einen bestimmten 3D Bereich betritt und beispielsweise einen Alarm generieren. Ein Akteur kann auch eine aktive Schwenk-/Neige-Kamera ansteuern, so daß diese auf das detektierte Objekt gerichtet werden kann.

[2] Räume im Bezug auf das Gebäude.
[3] Ein Überblick über alle im System definierten Koordinatensysteme ist in Anh. B.1 gegeben.
[4] Vgl. Kap. 2.5.4 und Anh. C zum Kameramodell und -kalibrierung.

2.3 Objektmodell

2.3.1 Einführung

Im Interpretationsprozeß werden in der Szene Objekte anhand eines Objektmodells detektiert, mit dem seine Ausprägung bestimmt ist. Das verwendete Objektmodell setzt sich, entsprechend der Definition von artikularen Objekten, aus einzelnen Objektmodellteilen zusammen, wobei jedes einzelne Objektmodellteil als starr angenommen wird.[1] Die Verbindung zwischen den einzelnen Objektmodellteilen kann als *Gelenk* oder *Knoten* bezeichnet werden, denn es sind im Objektmodell zwischen den einzelnen Objektmodellteilen nur rotatorische Veränderungen möglich.

Das Objektmodell ist somit nur zur Modellierung von Objekten geeignet, bei denen es keine Translationen zwischen den Objektmodellteilen gibt. Dies ist bei Lebewesen gegeben. Bei der Modellierung des menschlichen Körpers können z.B. die Objektmodellteile entsprechend den Körperteilen gewählt werden, so daß die Verbindungen zwischen den Objektmodellteilen durch die Gelenke des Körpers bestimmt sind. Aufgrund der Anwendungen von STABIL^{++} zur Personendetektion Bewegungserfassung wird im weiteren das Beispiel der Modellierung des menschlichen Körpers verwendet.

Mit der Verbindung der Objektmodellteile in den Gelenken ist eine innere Zusammenhangsstruktur des Objektmodells gegeben. Eines der Objektmodellteile ist ausgezeichnet und dient dem Objektmodell als Einstieg in eine *hierarchische, innere Struktur*. Ein Objektmodell ist daher wie folgt definiert:

$$obj = \langle omp_{0.1}, {}^{wcs}\mathbf{T}_{obj}, HIST, T, COMPO \rangle \tag{2.2}$$

wobei $omp_{0.1}$ das ausgezeichnete Objektmodellteil ist.

Dem Objektmodell ist ein kartesisches Koordinatensystem zugeordnet. Die Lage dieses Koordinatensystems in dem Weltkoordinatensystem wcs des Szenenmodells $scene$ ist mit der homogenen Transformationsmatrix ${}^{wcs}\mathbf{T}_{obj}$ bestimmt.[2] Mit

$$\vec{p}_{wcs} = {}^{wcs}\mathbf{T}_{obj} \cdot \vec{p}_{obj}$$

ergibt sich die Transformation eines Punktes $\vec{p}_{obj}$ im Koordinatensystem des Objektmodells in das Koordinatensystem des Szenenmodells. Die einzelnen im System verwendeten Koordinatensysteme und Koordinatentransformationen sind im Anh. B.1 ausführlich dargestellt.

Anhand des Translationsteils in der Transformation ${}^{wcs}\mathbf{T}_{obj}$ kann das Objektmodell von dem Szenenmodell bei der initialen Interpretation an die entsprechende Stelle “gestellt” werden, an der Objekte erwartet werden. Somit ist die Position bestimmt. Mit dem Rotationsteil wird für die initiale Suche die Orientierung des Koordinatensystems des Objektmodells im Weltkoordinatensystem festgelegt.[3] Die Lage einer Objektmodellinstanz, die mit ${}^{wcs}\mathbf{T}_{obj}$ angegeben ist, reduziert sich somit auf eine Position.

Zusätzlich zur aktuellen Lage des Objektmodells ist mit

$$HIST = \left\{ {}^{wcs}\mathbf{T}_{obj,t_{(-1)}}, \ldots, {}^{wcs}\mathbf{T}_{obj,t_{(-n)}} \right\}$$

[1] Für eine spezifische Objektmodellinstanz kann in STABIL^{++} jedoch eine Adaption der Größe des Modells durchgeführt werden.

[2] Vgl. Anh. B.3 zur Verwendung der homogenen Koordinatentransformation.

[3] Diese Drehung bleibt eine Objektmodellinstanz erhalten. Die Drehung des Objektes wird mit der Rotation des ersten Objektmodellteils berücksichtigt.

die Historie seiner Lage / Position bekannt. Die Historie wird im Interpretationsprozeß für die Bewegungsvorhersage verwendet, dient jedoch auch bei Anwendungen des Systems zur Bewegungsvermessung zur Darstellung von Bewegungsabläufen. Die Zeitpunkte / Zeitstempel, zu denen das Objektmodell detektiert wurde ist diesem mit

$$T = \{t_{(0)}, t_{(-1)}, \ldots, t_{(-n)}\}$$

bekannt. Diese werden hierbei entsprechend der jeweils aktuellen Zeit eines Interpretationszyklus gesetzt. Der Zeitpunkt t_0 entspricht dem aktuellen Zeitpunkt, für den die Lage des Objektmodellteils in ${}^{wcs}\mathbf{T}_{obj}$ festgelegt ist.

Mit $COMPO = \{compo_1, \ldots, compo_n\}$ sind für ein Objektmodell Gruppen / Kompositionen von Objektmodellteilen definiert, die zur Bestimmung der Rotationen zwischen den Objektmodellteilen verwendet werden. Hierbei werden, aufgrund von drei verschiedenartigen Gruppierungen, jeweils unterschiedliche Heuristiken zur Bestimmung der Gelenkwinkel zwischen den Objektmodellteilen innerhalb einer Gruppe vorausgesetzt. Dies ist notwendig, damit die Anzahl der rotatorischen Freiheitsgrade in den Gelenken herabgesetzt wird. Es werden durch die unterschiedlichen Kompositionen die durch die Struktur der zu detektierenden Objekte vorgegebenen Freiheitsgrade berücksichtigt.[4]

2.3.2 Objektmodellteil

Die eigentlichen Modellinformationen des Objektmodells sind in den Objektmodellteilen eines Objektmodells festgehalten. Mit den Informationen der einzelnen Objektmodellteile werden drei Aspekte der Modellierung abgedeckt. Dies ist zum einen die *innere Struktur* des Objektmodells, die den Zusammenhang der Objektmodellteile untereinander beschreibt. Der zweite Aspekt ist eine *geometrische Struktur*, mit der die 3D Lage der Objektmodellteile zueinander beschrieben ist. In der weiteren sog. *äußeren Struktur* ist die Erscheinungsform des Modells mit einer Volumenrepräsentation und Merkmalen gegeben.

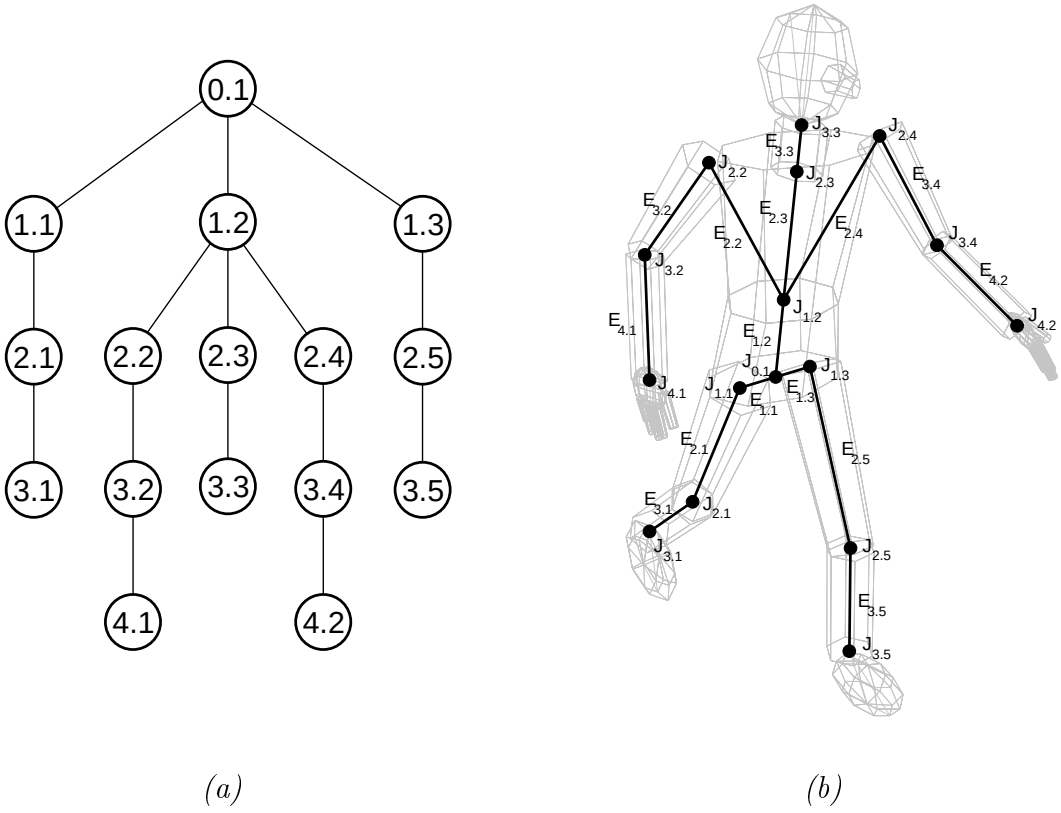
Diese drei verschiedenen Modellstrukturen werden in den folgenden Abschnitten dargestellt. Je nach Struktur werden verschiedene, durch die Objektmodellteile gekapselte, Informationen verwendet. Die Komponenten mit den Informationen werden daher in den Abschnitten zu den verschiedenen Strukturen erläutert. Es sei hier jedoch schon die zusammenfasssende Definition eines Objektmodellteils omp_ν anhand seiner Komponenten gegeben:

$$omp_\nu = \left\langle OMP_\nu, {}^{omp_\mu}\mathbf{T}_{omp_\nu}, restr_\nu^{\angle}, HIST_\nu, pred_\nu, vol_\nu, \mathbf{F}_\nu, \mathbf{F}'_\nu \right\rangle \tag{2.3}$$

Die einzelnen Komponenten sind

- die Liste der Nachfolgeobjektmodellteile OMP_ν,
- die Transformationsmatrix ${}^{omp_\mu}\mathbf{T}_{omp_\nu}$,
- die Winkelrestriktionen $restr_\nu^{\angle}$,
- die Historie $HIST_\nu$,
- der Vorhersagefilter $pred_\nu$,
- der Volumenkörper vol_ν,
- die primären Merkmale $\mathbf{F}$ und
- die sekundären Merkmale $\mathbf{F}'$.

[4] Die Bestimmung der Rotationen unter Verwendung der Kompositionen ist in Anh. A dargestellt.



(a) *(b)*

Abbildung 2.1: *Innere Objektmodellstruktur: (a) Baum der hierarchischen Struktur der Objektmodellteile und (b) Projektion der Struktur auf das 3D Modell; entspricht ungefähr der Knochenstruktur des menschlichen Körpers.*

2.3.3 Innere Objektmodellstruktur

Die Aneinanderreihung von einzelnen Objektmodellteilen bestimmt die innere Struktur des Objektmodells. Die Struktur ist hierarchisch aufgebaut und läßt sich als Baum beschreiben. Das dem Objektmodell obj bekannte ausgezeichnete Objektmodellteil $omp_{0.1}$ entspricht dem Wurzelelement des Baumes. Jedem Objektmodellteil ist eine Anzahl von nachfolgenden Objektmodellteilen in OMP_i bekannt. Somit ergibt sich ein rekursiver Aufbau des Modells aus den Objektmodellteilen. Für das Objektmodellteil omp_ν gilt z.B. $OMP_\nu = \{omp_{\nu_1}, \ldots, omp_{\nu_n}\}$.

Bei der Modellierung des menschlichen Körpers durch Objektmodellteile für die Hüfte, den Rumpf, den Hals, den Kopf und jeweils drei Teile für die Gliedmaßen kann die Hüfte als Wurzel der inneren Objektmodellstruktur angesehen werden. Die entsprechenden Nachfolger für das Objektmodell der Hüfte sind dann das Objektmodellteil für den Rumpf, den linken Oberschenkel und den rechten Oberschenkel. Die Objektmodellteile für den Kopf, für die Füße und für die beiden Hände haben dann keine weiteren Nachfolger und entsprechen in der Baumstruktur den als Blätter bezeichneten Knoten. Vgl. hierzu Abb. 2.1 (a). Die Numerierung der Objektmodellteile ist so gewählt, daß die erste Ziffer den Abstand zum Wurzelelement der Hierarchie angibt. Mit der zweiten Ziffer werden alle Objektmodellteile auf einer Hierarchieebene durchnumeriert. Für die o.a. Modellierung des menschlichen Körpers ergibt sich eine Zuordnung entsprechend Tab. 2.1. Die Reihenfolge der zweiten Ziffer ist hierbei willkürlich gewählt.

Die Abgrenzung der einzelnen Objektmodellteile zueinander erfolgt bei der Modellierung des menschlichen Körpers jeweils entsprechend den Gelenken des Körpers. D.h. an den Gelen-

Hüfte	0.1	Rumpf	1.2
Hals	2.3	Kopf	3.3
rechter Oberschenkel	1.1	linker Oberschenkel	1.3
rechter Unterschenkel	2.1	linker Unterschenkel	2.5
rechter Fuß	3.1	linker Fuß	3.5
rechter Oberarm	2.2	linker Oberarm	2.4
rechter Unterarm	3.2	linker Unterarm	3.4
rechte Hand	4.1	linke Hand	4.2

Tabelle 2.1: *Numerierung der Objektmodellteile entsprechend Abb. 2.1.*

ken beginnt jeweils ein neues Objektmodellteil. Faßt man die Gelenkpunkte als Knoten einer Baumstruktur auf und definiert zwischen den Knoten die Kanten, so kann die innere Objektmodellstruktur als Graph, entsprechend der Baumstruktur in Abb. 2.1 (a), angesehen werden.

Jeder Knoten des Baumes steht für ein Objektmodellteil. Jedem Knoten, außer dem Wurzelknoten, der das ausgezeichnete Objektmodell $omp_{0.1}$ repräsentiert, ist eine Kante zu dem Knoten des jeweiligen *Vorgängerobjektmodellteils* in der Hierarchie zugeordnet. Projiziert man diesen Graphen auf das Objektmodell, so erhält man eine "skelettartige" Darstellung der inneren Modellstruktur. Für die Modellierung des menschlichen Körpers ergibt sich die Struktur entsprechend Abb. 2.1 (b).[5] Hierbei entspricht die innere Struktur im wesentlichen der Knochenstruktur, dem Skelett des menschlichen Körpers. Eine Ausnahme bildet die Kante $E_{2.2}$ und die jeweils symmetrische in der anderen Körperhälfte $E_{2.4}$, mit der der Rumpf mit dem Oberarm verbunden ist. Dieser Kante kann kein eindeutiger Knochen zugeordnet werden. Wichtig ist jedoch, daß bei der Modellierung nur solche Gelenke zwischen den Objektmodellteilen gebildet werden, bei denen die Kanten der Inneren Modellstruktur in der Länge starr sind. Diese Bestimmung ist Grundlage einer bei der Interpretation verwendeten Restriktion.

2.3.4 Geometrische Struktur

Die Struktur des inneren Zusammenhangs wird durch eine 3D Lage der einzelnen Objektmodellteile ergänzt. Hierzu wird jeder Knoten des hierarchischen Aufbaus als Ursprung des zugehörigen Objektmodellteiles angesehen, in dem ein lokales Koordinatensystem definiert ist. Somit kann beschrieben werden, wie die einzelnen Objektmodellteile zueinander angeordnet sind und wie sich diese zueinander verdrehen können. Die Orientierung der Koordinatensysteme für die Modellierung des menschlichen Körpers ist in Abb. 2.2 (b) dargestellt.

Die Koordinatensysteme sind so orientiert, daß die z-Achse nach Möglichkeit jeweils in Richtung der Kante eines nachfolgenden Knotens zeigt. Damit zeigt die z-Achse entlang des Knochens des jeweiligen Objektmodellteiles. So liegt z.B. der Ursprung des Objektmodellteiles $omp_{2.2}$ für den rechten Oberarm im rechten Schultergelenk und die z-Achse des lokalen Koordinatensystems $Z_{2.2}$ zeigt in Richtung des Oberarmknochens, vgl. Abb. 2.1 (b) und 2.2 (a).

Ist kein eindeutiger Knochen für die Verbindung von zwei Objektmodellteilen gegeben, dann ist die Richtung der z-Achse so zu wählen, daß sie in Richtung des Nachfolgers zeigt, mit dem eine möglichst gradlinige Verbindung besteht. Dies gilt z.B. bei der Modellierung für den

[5] Der Knoten eines Objektmodellteils omp_ν ist als J_ν gekennzeichnet. Die Kante zum Knoten Vorgängerobjektmodellteil von J_ν ist mit E_ν gekennzeichnet.

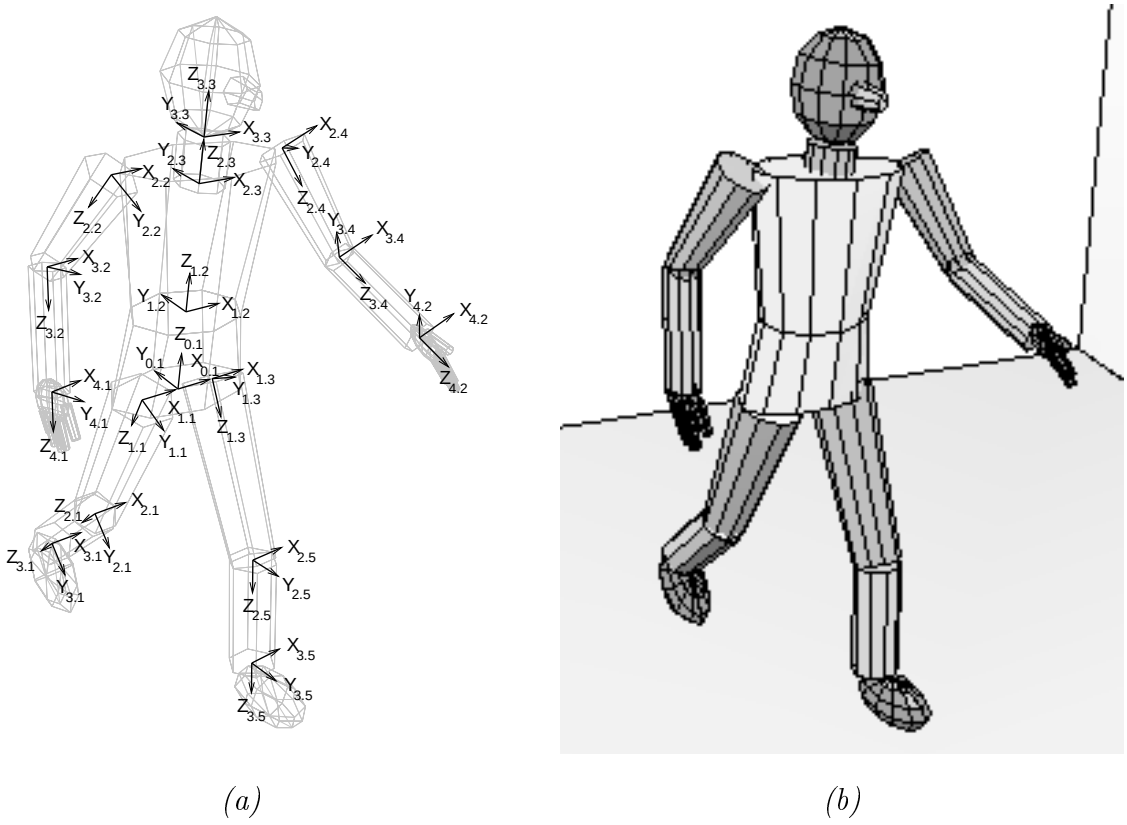


Abbildung 2.2: *(a) Geometrische Objektmodellstruktur: lokale Koordinatensysteme in den Objektmodellteilen und (b) Objektmodellstruktur mit Volumenkörpern.*

Menschen für das Objektmodellteil $omp_{1.2}$ des Rumpfes, dem die Objektmodellteile $omp_{2.3}$ für den Hals und die Objektmodellteile $omp_{2.2}$ und $omp_{2.4}$ für den linken und rechten Oberarm folgen. Für die Richtung der z-Achse $Z_{1.2}$ wird der Hals als direkter, geradliniger Nachfolger gewählt. Vergleiche hierzu auch die Beschreibung der äußeren Struktur im kommenden Abschnitt.

Mit der homogenen Transformationsmatrix ${}^{omp_\mu}\mathbf{T}_{omp_\nu}$ ist für das Objektmodellteil omp_ν die Transformation seines lokalen Koordinatensystems zum Koordinatensystem seines Vorgänger-Objektmodellteiles omp_μ angegeben.

In Kombination mit der inneren Modellstruktur kann somit für einen Punkt eines jeden Objektmodellteiles die Transformation in das Weltkoordinatensystem des Szenenmodells angegeben werden. Es gilt für die Transformation aus dem lokalen Koordinatensystem des Objektmodellteiles omp_j aus einem Objektmodell obj ins Weltkoordinatensystem wcs des Szenenmodells $scene$:

$$ {}^{wcs}\mathbf{T}_{omp_j} = {}^{wcs}\mathbf{T}_{obj} \cdot {}^{obj}\mathbf{T}_{omp_0} \cdot {}^{omp_0}\mathbf{T}_{omp_1} \cdots {}^{omp_{j-1}}\mathbf{T}_{omp_j} \tag{2.4} $$

Die Vorgängerobjektmodellteile in der hierarchischen Objektmodellstruktur sind hierbei $omp_0 \cdots omp_{j-1}$.[6]

Die Position der Ursprünge der einzelnen Koordinatensysteme und damit der Gelenke sind allein durch den Translationsanteil in den Transformationen bestimmt. Die zugehörigen Translationsvektoren entsprechen dann den Kanten E_i in der inneren Struktur, falls man eine 3D Lage

[6]Das Objektmodellteil omp_0 entspricht dem Wurzelelement der hierarchischen, inneren Modellstruktur und ist mit dem in Abschn. 2.3.3 als $omp_{0.1}$ bezeichneten Objektmodellteil identisch.

der Knoten annimmt, vgl. Abb. 2.1 (b). Mit den 3D Kanten ist die Länge der Knochen oder der Knochenstruktur bestimmt, die die zugehörigen beiden Gelenke verbindet. Im eigentlichen Interpretationsprozeß wird die feste Länge in diesen Verbindungen ausgenutzt.

Ebenso werden die möglichen Winkelstellungen in den Gelenken ausgenutzt. Hierzu hat ein Objektmodellteil omp_ν mit $restr_\nu^\angle = \{\alpha_{neg_\nu}, \alpha_{pos_\nu}, \beta_{neg_\nu}, \beta_{pos_\nu}, \gamma_{neg_\nu}, \gamma_{pos_\nu}\}$ [7] Informationen über jeweils maximal zulässige positive und negative Winkel für die Rotationen um die drei Achsen des Koordinatensystems des Vorgänger-Objektmodellteiles. Die maximalen Rotationen in den Gelenken sind entsprechend den maximalen Bewegungen in den Gelenken der modellierten Objekte zu wählen. Diese Angaben werden als Restriktionen bei der Bewertung der möglichen Stellungen der einzelnen Objektmodellteile zueinander bei der Interpretation verwendet.

Zum geometrischen Modell gehört noch die Historie der Transformationen zwischen den einzelnen Objektmodellteilen. Somit erhält jedes Objektmodellteil eine Liste von Transformationen. Für das Objektmodellteil omp_ν gilt:

$$HIST_\nu = \left\{ ({}^{omp_\mu}\mathbf{T}_{omp_{\nu,t_{(-1)}}}, \mathbf{s}_{t_{(-1)}}), \ldots, ({}^{omp_\mu}\mathbf{T}_{omp_{\nu,t_{(-n)}}}, \mathbf{s}_{t_{(-n)}}) \right\} \tag{2.5}$$

In der Historie sind somit die Gelenkwinkel der vergangenen Zeitpunkte vermerkt;[8] so wird der Bewegungsablauf einer Modellinstanz festgehalten. Die Historie wird im Interpretationsprozeß für die Bewegungsvorhersage verwendet, dient jedoch auch bei Anwendungen des Systems zur Bewegungsvermessung zur Bestimmung von Bewegungsabläufen. Die Zeiten $t_{(-1)} \ldots t_{(-n)}$ werden hierbei entsprechend der jeweils aktuellen Zeit des Interpretationszyklus gesetzt. Desweiteren ist in der Historie für die vergangenen Zeitpunkte jeweils ein sog. Szenenmerkmal $\mathbf{s}_{(-i)}$ eingetragen. Dieses Szenenmerkmal ist ein 3D Merkmal, das im Interpretationsprozeß in der Szene extrahiert und dem Objektmodellteil zugeordnet worden ist. Über die in der Historie vermerkten Szenenmerkmale ist die Position des Objektmodellteils zu dem entsprechenden Zeitpunkt bestimmt worden.[9]

Einem Objektmodellteil omp_ν ist mit $pred_\nu$ noch ein Filter zur Vorhersage der Position des Objektmodellteils bekannt. Hiermit wird im Interpretationsprozeß aufgrund der Positionsänderung über die Zeit für einen nächsten Interpretationsschritt die Position vorhergesagt, wobei diese sich unter anderem auf die Qualitäten aus der Historie $HIST$ stützen. Hierüber ist zum einen eine Einschränkung des 3D Suchraumes möglich, so daß im Interpretationsprozeß in SSP_ν für das Objektmodellteil ein 3D Suchraum ermittelt werden kann. Zum anderen wird hierdurch die Zuverlässigkeit der Interpretation verbessert und die Interpretation somit sicherer. Die Filter zur Vorhersage der Positionen stützen sich entweder einfach auf die Historie $HIST_\nu$ oder es ist mit dem Filter noch Wissen über die mögliche Bewegungsaktivität des Objektmodellteils bekannt.

2.3.5 Äußere Objektmodellstruktur

Mit der äußeren Objektmodellstruktur wird die Erscheinungsform des Objektes beschrieben. Hierzu zählen zum einen die Repräsentation der Volumen der einzelnen Objektmodellteile und deren Oberfläche. Zum anderen zählen hierzu die Merkmale, die das zu detektierende Objekt charakterisieren.

[7] Werden die Winkel über die Eulerwinkel bestimmt, so sind $\{\alpha_{neg_\nu}, \alpha_{pos_\nu}, \gamma_{1\,neg_\nu}, \gamma_{1\,pos_\nu}, \gamma_{2\,neg_\nu}, \gamma_{2\,pos_\nu}\}$ anzugeben, vgl. Anh. A.6.

[8] Die Transformationen enthalten ebenfalls die Translationsanteile, diese ändern sich jedoch nur bei einer Adaption der äußeren Objektmodellstruktur, vgl. Kap. 3.8.2.

[9] Eine Einteilung der in Merkmale und die Definition von Szenenmerkmalen ist in Kap. 2.4 zu finden.

Volumen und Oberflächen

Jedem Objektmodellteil ist ein Volumenkörper zugeordnet. Bei der Modellierung des menschlichen Körpers eignen sich beispielsweise, aufgrund der annähernden Symmetrie der Körperteile, einfache Rotationskörper, wie Ellipsoide vol_{ell} und Kegelstümpfe vol_{trCone}.[10] Die Rotationsachsen der Körper liegen parallel zu den z-Achsen der lokalen Koordinatensysteme der Objektmodellteile. Dies entspricht den Kanten E_i der inneren Struktur, unter Berücksichtigung der 3D Lage der Knoten, vgl. Abb. 2.1 (b).

Abb. 2.2(b) zeigt die 3D Volumenkörper für das Objektmodell des menschlichen Körpers. Den Objektmodellteilen, die in der inneren Struktur die Blätter des Baumes darstellen, ist jeweils ein Ellipsoid zugeordnet, allen anderen ein Kegelstumpf.[11]

Die Rotationskörper sind in den lokalen Koordinatensystemen der Objektmodellteile definiert. Es muß die Lage und ihre Ausdehnung beschrieben sein. Nachdem der Volumenkörper des Ellipsoiden vol_{ell} keine eindeutige Rotationsachse hat, wird die Lage des Mittelpunktes angegeben. Die Ausdehnung wird mit den drei Halbmessern in Richtung der drei Achsen des Koordinatensystems angegeben. Die Oberfläche des Ellipsoiden läßt sich durch eine Triangulierung approximieren. Hierzu werden parallel zur z-Achse Längengrade auf den Ellipsoiden gelegt und Breitengrade parallel zur xy-Ebene. Hierbei entstehen an den Polen Dreieck-Flächen und ansonsten Viereck-Flächen. Die Viereck-Flächen werden dann noch diagonal geteilt. Die Reihenfolge der Eckpunkte der so entstandenen Dreieck-Flächen wird so festgelegt, daß der Normalenvektor auf den Flächen vom Mittelpunkt des Ellipsoiden weg zeigt. Somit ist die Orientierung der Oberfläche bestimmt.

Die Rotationsachsen der Kegelstumpf-Volumenkörper vol_{trCone} liegen immer auf den z-Achsen der lokalen Koordinatensysteme. Die begrenzenden Ellipsenflächen liegen orthogonal zu den Rotationsachsen und damit parallel zu den xy-Ebenen. Es muß daher zur Bestimmung dieser Körper die Position der Mittelpunkte der Ellipsenflächen auf der z-Achse angegeben werden und die zugehörigen Halbmesser in Richtung der x- und der y-Achse. Die Oberflächen des Kegelstumpfes ist durch seine zwei begrenzenden Ellipsen und die Mantelfläche bestimmt. Die Mantelfläche wird ebenfalls durch eine Triangulierung approximiert. Hierzu werden beide begrenzenden Ellipsen durch Vielecke approximiert, deren Eckpunkte auf der Ellipsenlinie liegen. Beide Vielecken haben die gleiche Anzahl von Ecken, so daß die auf beiden Flächen quasi gegenüber liegenden Ecken verbunden werden können. Die somit entstandenen Viereck-Flächen werden noch diagonal geteilt. Die Reihenfolge der Eckpunkte der so entstandenen Dreieck-Flächen ist so festgelegt, daß der Normalenvektor auf den Flächen orthogonal von der z-Achse weg zeigt. Somit ist auch hier die Orientierung der Oberfläche bestimmt.

Merkmale

Die Merkmale eines Objektmodells und somit die Merkmale der Objektmodellteile charakterisieren das Modell. Anhand der Merkmale werden im Interpretationsprozeß die Objektmodelle detektiert. Hierbei wird zwischen sog. primären und sekundären Merkmalen der Objektmodellteile unterschieden. Primäre Merkmale dienen zur Bestimmung von 3D Positionen der einzelnen Objektmodellteile, wo hingegen die sekundären Merkmale zur Verifikation der vom Interpretationsprozeß aufgestellten möglichen Interpretationen, den sog. Hypothesen verwendet

[10] Die Kürzel der Komponenten sind von den zugehörigen englischen Begriffen abgeleitet: vol – **vol**ume, ell – **ell**ipsoide, $trCone$ – **tr**uncated **cone**.

[11] Zur besseren Erkennung der Orientierung der Arme sind den Händen auch noch die Finger und am Kopf eine Nase dargestellt; diese zusätzlichen Objektmodellteile erscheinen jedoch nicht in der hierarchischen, inneren Struktur.

werden.

Im Interpretationsprozeß wird zwischen verschiedenen Stufen von Merkmalen unterschieden. Die Merkmale, die den Objektmodellteilen zugeordnet sind, werden hierbei als 3D Modellmerkmale bezeichnet. Die Einteilung der Merkmale wird im folgenden Abschnitt vorgenommen.

Einem Objektmodellteil omp_ν ist mit $\mathbf{F}_\nu \in \{\emptyset, \{\mathbf{f}_\nu\}\}$ kein oder ein primäres Modellmerkmal zugeordnet.[12] Mit $\mathbf{F}'_\nu \in \{\mathbf{f}'_{\nu_1}, \ldots, \mathbf{f}'_{\nu_n}\}$ können dem Objektmodellteil sekundäre Merkmale zugeordnet werden.

[12] Die Objektmodellstrukturen lassen es zu, auch mehrere primäre Merkmale zu einem Objektmodellteil zuzuordnen, jedoch wird in Kap. 3 der Interpretationsprozeß nur unter Verwendung maximal eines primären Merkmals beschrieben.

2.4 Merkmale

2.4.1 Unterteilung der Merkmale

In STABIL^{++} wird zwischen drei Stufen von Merkmalen unterschieden:

- 2D *Bildmerkmalen* $\mathbf{i}$ im Sensorraum,
- 3D *Szenenmerkmalen* $\mathbf{s}$ im Szenenraum und
- 3D *Modellmerkmalen* $\mathbf{m}$ im Modellraum.

Aufgrund eines 3D/3D-Vergleichs zur Objektdetektion muß im Interpretationsprozeß zunächst aus den 2D Sensordaten der Videobilder 3D Informationen gewonnen werden. Man spricht hierbei von einem Übergang von den 2D Bildmerkmalen zu 3D Szenenmerkmalen. Die Szenenmerkmale sind im 3D Raum der Szene definiert und somit auf das Weltkoordinatensystem bezogen. Anschließend wird in einem sog. Struktur-Vergleichsverfahren[1] nach gültigen Zuordnungen von 3D Szenenmerkmalen zu 3D Modellmerkmalen gesucht. Für den Interpretationsprozeß muß daher jedem Objektmodellteil eines Objektmodells Informationen über seine Merkmale und die Übergänge zwischen den drei Stufen bekannt sein. Die Verwendung der Merkmale aus den drei Stufen zur Interpretation und somit der Übergang von 2D Bildmerkmalen zu 3D Szenenmerkmalen und die Zuordnung zu 3D Modellmerkmalen ist auch in Abb. 1.3 auf S. 18 dargestellt. Bei diesem Übergang und der Zuordnung ist sicherzustellen, daß nur Merkmale aus den verschiedenen Stufen mit gleichen Basisattributen ineinander übergehen / verglichen werden. Basisattribute zeichnen sich dadurch aus, daß diese keine Maßzahlen für ein Merkmal darstellen, sondern daß mit diesen Merkmalen signifikant unterschieden und in Gruppen eingeteilt werden können.

In den folgenden Abschnitten werden zunächst alle drei Merkmalsbegriffe definiert. Daran schließt sich die Beschreibung der Merkmalsinformationen, die einem Objektmodellteil zugeordnet sind, an.

2.4.2 2D Bildmerkmale

Bildmerkmale sind geometrische Primitive, die sich aus den Sensordaten / Videobild segmentieren lassen.[2] Dies sind z.B. Kanten, Ecken, Kreisbögen und Ellipsen. Zu den verschiedenen Bildmerkmalen lassen sich Attribute bestimmen. So lassen sich für Flächen ein mittlerer Grauwert oder eine Farbe bestimmen und für Regionen lassen sich Attribute der Form, wie z.B. Kompaktheit, Anisometrie und Zirkularität berechnen. Auf Konturen können die Attribute der Länge, der Fläche, der Orientierung und Maßzahlen angegeben werden. Für eine Ellipse sind die Maßzahlen die Orientierung als Winkel und die Radien der beiden Halbmesser.[3] Als Basisattribute, zur signifikanten Unterscheidung und Einteilung von Bildmerkmalen in Gruppen eignen sich z.B. der mittlere Grauwert oder die Farbe der Bildregion. Die Attribute, mit denen Maßzahlen bestimmt sind, werden dann zur Beurteilung und zur Unterscheidung von Bildmerkmalen einer Gruppe verwendet.

[1] engl. *matching*.

[2] Weitere, als Bildmerkmal zu bezeichnende Eigenschaften eines Videobildes, wie z.B der Farbe oder Textur, werden hier als Attribut eines Bildmerkmals verwendet. Das geometrische Primitiv einer bestimmt texturierten Fläche hat dann jedoch keine weiteren geometrischen Eigenschaften, als die Ausdehnung und den Mittelpunkt.

[3] Die in STABIL^{++} verwendeten Methoden der Bildverarbeitung, die damit zu bestimmenden Bildmerkmale und deren Attribute sind im Kap. 3.4 erläutert. Weitere Übersichten zu Bildmerkmalen und ihren Attributen sind z.B. in [Mun96], [Ric95] und [HS92] gegeben.

Zusätzlich zu den Attributen sind allen segmentierten Bildmerkmalen eine Bildzeile y und eine Bildspalte x zugeordnet. Der durch die Zeile und Spalte bestimmte Bildpunkt $\vec{p}_{img} = [x, y]^T$ gibt die Position des Bildmerkmals im Bild an. Dies kann bei symmetrischen, z.B. kreisförmigen Merkmalen der Mittelpunkt sein, bei einem beliebigen Merkmal ein anderer ausgezeichneter Punkt. Es muß nur sichergestellt sein, daß sich die Attribute $attr_1^{\mathbf{i}}, \ldots, attr_n^{\mathbf{i}}$ des Merkmals, sofern es zusätzlich zur Position, die Orientierung beschreibt, auf diesen ausgezeichneten Bildpunkt beziehen. Insbesondere im Hinblick auf den Übergang von 2D Bildmerkmalen zu 3D Szenenmerkmalen sind symmetrische Bildmerkmale von Vorteil, da deren Lage direkt mit dem durch x und y definierten Punkt $\vec{p}_{img}$ eindeutig beschrieben ist. Man spricht daher auch von Punktmerkmalen.

Somit ist ein Bildmerkmal $\mathbf{i}$ wie folgt definiert:

$$\mathbf{i} = \left\langle \{attr_1^{\mathbf{i}}, \ldots, attr_n^{\mathbf{i}}\}, \vec{p}_{img}, camPar, camPose \right\rangle \tag{2.6}$$

Mit $camPar$ und $camPose$ sind die Abbildungseigenschaften der Kamera beschrieben, die zum Zeitpunkt der Aufnahme des Videobildes Gültigkeit hatten. Über diese können dann die für den Übergang von den 2D Bildmerkmalen auf die 3D Szenenmerkmale notwendigen 3D Sichtstrahlen bestimmt werden. Die Abbildungseigenschaften der Kamera sind im Kap. 2.5.4 zum Kameramodell dargestellt.

2.4.3 3D Szenenmerkmale

Die 2D Bildmerkmale müssen sich in ein 3D Szenenmerkmal überführen lassen. Daher sind die Szenenmerkmale ebenfalls Kanten, Ecken, Kreisbögen und Ellipsen, jedoch ist für diese eine 3D Lage bestimmbar. Die Attribute $\{attr_1^{\mathbf{s}}, \ldots, attr_n^{\mathbf{s}}\}$ der Szenenmerkmale sind, sofern es Maßzahlen sind, auf das Weltkoordinatensystem des Szenenmodells bezogen. Beschreibende Attribute, die sich als Basisattribute eignen, wie z.B. Farbe oder z.B. Temperatur eines Szenenmerkmals, sind hier unabhängig vom Sensor bestimmt. Bei der Überführung von Bildmerkmalen zu Szenenmerkmalen werden die Basisattribute von den Bildmerkmalen übernommen, so daß die gleiche Gruppierung möglich ist.

Ein Szenenmerkmal mit dem Attribut einer bestimmten Farbe, z.B. *rot*, kann jedoch nur durch ein Bildmerkmal bestimmt werden, das auf den Sensordaten einer Farbkamera bestimmt worden ist. Für den Sensor muß ein Farbklassifikator vorhanden sein, in dem eine Klasse für die Farbe, die dem Szenenmerkmal attribuiert ist, definiert ist.

Bei dem Attribut einer bestimmten Temperatur, z.B. 37 °C, muß das zugehörige Bildmerkmal auf den Daten einer Infrarotkamera bestimmt werden. Zu dem Sensor muß eine Temperaturkalibrierung bekannt sein, damit die Temperaturangabe des Szenenmerkmals auf ein Helligkeitsintervall der Grauwerte abgebildet werden kann.

Die Abbildungseigenschaften der Kameras sind durch die Verwendung eines Kameramodells und der damit durchgeführten Kamerakalibrierung bekannt, daher können die geometrischen Attribute eines 3D Szenenmerkmals, die durch Maßzahlen im Weltkoordinatensystem angegeben sind, in das 2D Signalfeld des Kamerabildes projiziert werden. Somit wird eine 3D Kante im Szenenraum zu einer 2D Linie oder zu einem 2D Kreisbogen entsprechend der Linsenverzerrung im Videobild. Umgekehrt kann jedoch keine direkte Projektion von 2D Bildmerkmalen auf 3D Szenenmerkmale vorgenommen werden.

Aus einzelnen Punkten von Bildmerkmalen, die durch ihre Zeilen- und Spaltenposition im Bild bestimmt sind, können 3D Sichtstrahlen im Merkmalsraum des Szenenmodells erzeugt

werden. Diese Punkte können die Anfangs- und Endpunkte einer Kante oder der Mittelpunkt einer Ellipse sein. Bei komplexeren Konturen erhält man durch eine Unterabtastung ein entsprechendes Bündel von Sichtstrahlen.

Die Sichtstrahlen gehen vom Ursprung des Kamerakoordinatensystems aus und schneiden im Abstand der Brennweite die Bildebene; die Lage der Bildebene ist hierbei durch die Modellvorstellung der Lochkamera bestimmt. Hiermit ist nur bestimmt, daß der korrespondierende 3D Punkt des Szenenmerkmals auf dem Sichtstrahl liegt. Die Tiefeninformation, d.h. in welchem Abstand der Punkt des Szenenmerkmals auf dem Sichtstrahl liegt, muß über zusätzliche Information gewonnen werden.

Die Tiefeninformation kann in STABIL^{++} durch zwei verschiedene Ansätze gewonnen werden. Im ersten Ansatz werden korrespondierende Bildmerkmale aus mehreren (mind. zwei) Videobildern von mehreren verschiedenen Kameras verwendet, man spricht von einem Stereo- und Mehrfachstereoaufbau. Im Szenenraum wird nach Schnittpunkten der 3D Sichtstrahlen des Bildmerkmals aus dem ersten Bild mit den entsprechenden Sichtstrahlen aus dem zweiten (und weiteren) Bild gesucht.[4] Durch die Basisattribute der Bildmerkmale wird sichergestellt, daß hierzu nur gleichartige Bildmerkmale verwendet werden.

Steht nur die Information von dem Bildmerkmal einer Kamera zur Verfügung, so kann die Tiefeninformation nur unter Verwendung von Modellwissen geschätzt werden. Hierzu werden mindestens zwei Sichtstrahlen oder ein Sichtstrahl und eine feste Bezugsfläche, z.B. die xy-Ebene des Szenenraumes verwendet. Aufgrund der Annahme der Orientierung des Objektmodells und somit seiner Objektmodellteile im Szenenraum wird die "Tiefe" an der Position bestimmt, an der das Abstandsmaß zwischen den Sichtstrahlen oder zwischen einem Sichtstrahl und einer Bezugsfläche dem entsprechenden Maß aus dem Modell entspricht.[5]

Bei komplexeren Merkmalen, deren Kontur mehrere Sichtstrahlen zugeordnet werden, muß ein Ausgleichsverfahren angewendet werden, um die optimale Lage des Szenenmerkmals in dem Sichtstrahlbündel zu finden. Durch die Verwendung der einfacheren Punkt- oder Flächenmerkmale kann man eine komplexere Ausgleichsrechnung umgehen. Dies setzt jedoch voraus, daß ein Merkmal von vielen (allen) Kameras und in gleicher Form bzgl. der Projektion im Bild sichtbar ist.[6]

Bei Einschränkung auf die Verwendung von Punkt- oder Flächenmerkmalen reicht für die Angabe der Lage des Szenenmerkmals ein Punkt $\vec{p}_{wcs}$ im Weltkoordinatensystem aus, mit dem eine Position bestimmt ist. Ein Szenenmerkmal $\mathbf{s}$ ist somit durch die Menge seiner Attribute und seinem Punkt im Weltkoordinatensystem bestimmt:

$$\mathbf{s} = \langle \{attr_1^{\mathbf{s}}, \ldots, attr_n^{\mathbf{s}}\}, \vec{p}_{wcs}, \mathbf{I}_{extr}, q \rangle \tag{2.7}$$

Mit $\mathbf{I}_{extr} = \{\mathbf{i}_1, \ldots, \mathbf{i}_k\}$ sind einem 3D Szenenmerkmal noch die 2D Bildmerkmale bekannt, aus denen seine Position $\vec{p}_{wcs}$ bestimmt worden ist. Die Anzahl der vermerkten Bildmerkmale hängt von der Anzahl der bei der Tiefenschätzung verwendeten Bildmerkmale ab.

Weiterhin ist für jedes Szenenmerkmal eine Qualität $q \in [0, 1]$ eingetragen. Diese Qualität ist ein Maß für die Güte / Sicherheit, mit der seine Position $\vec{p}_{wcs}$ im Interpretationsprozeß bestimmt worden ist. Die Qualität hängt dabei generell von den zur Tiefenschätzung verwendeten Bildmerkmalen ab. Jedoch kann ein Szenenmerkmal auch aus einer Vorhersage geschätzt

[4] Die Problematik der Korrespondenzsuche wird im Kap. 3.5.3 zum 2D / 3D Übergang erläutert.

[5] Weitere Angaben zur Schätzung der Tiefeninformation aus Modellwissen sind unter Kap. 3.5.2 zu finden.

[6] Verdeckungen können von dem System zum einen bei der initialen Detektion durch Verwendung mehrerer Kameras akzeptiert werden, zum anderen werden durch das verwendete Interpretationsverfahren bei der Re-Detektion weitere zeitweilig verdeckte Merkmale in ihrer Position vorhergesagt.

werden, dann ist $\mathbf{I}_{extr} = \emptyset$ und das Maß für die Güte / Sicherheit q der 3D Position ist dementsprechend geringer zu setzen. Die Bestimmung der Qualität ist in dem Kap. 3 zum Interpretationsprozeß näher erläutert.

2.4.4 3D Modellmerkmale

Die 3D Modellmerkmale entsprechen in ihrer Art und den Attributen $attr_1^{\mathbf{m}}, \ldots, attr_n^{\mathbf{m}}$ im wesentlichen den 3D Szenenmerkmalen, sie sind jedoch einem festen Objektmodellteil zugeordnet und ihre Lage ist nicht im Weltkoordinatensystem des Szenenmodells, sondern im lokalen Koordinatensystem des Objektmodellteils definiert. Auch hier lassen sich die beschreibenden Attribute als Basisattribute verwenden, mit denen sichergestellt wird, daß nur Szenenmerkmale zu Modellmerkmalen zugeordnet werden, die gleichartig sind.

Die Größe und die Position des Modellmerkmals im lokalen Koordinatensystem kann *explizit* als Attribut angegeben werden oder bestimmt sich *implizit* aus dem Volumenkörper vol, der dem Objektmodellteil mit der äußeren Objektmodellstruktur zugeordnet ist. Die explizite Angabe von Attributen ist notwendig, wenn das Objektmodellteil durch eine Landmarke markiert ist. Bei diesen expliziten Merkmalen muß mit einem Verschiebungsvektor $\vec{t} = [x, y, z]^T$ der Versatz des Mittelpunktes der Marke zum Ursprung des Koordinatensystems des Objektmodellteils und, z.B. bei einer kreisförmigen Marke, der Radius angegeben werden.[7]

Für ein Objektmodellteil, dem ein Ellipsoid vol_{ell} als Volumenkörper zugeordnet ist, kann eine 3D Ellipse als implizites Modellmerkmal verwendet werden. Die Attribute der Maße ergeben sich dann implizit aus den beiden größten Halbmessern des Ellipsoiden und dessen Mittelpunkt, bezogen auf das lokale Koordinatensystem. Der Verschiebungsvektor $\vec{t}$ ist dann implizit durch die Position des Mittelpunktes des Ellipsoiden im lokalen Koordinatensystem bestimmt.

Ein Modellmerkmal $\mathbf{m}$ ist somit definiert als:

$$\mathbf{m} = \left\langle \{attr_1^{\mathbf{m}}, \ldots, attr_n^{\mathbf{m}}\}, \vec{t} \right\rangle \tag{2.8}$$

2.4.5 Primäre Merkmale der Objektmodellteile

Das primäre Merkmal $\mathbf{f}_\nu$ eines Objektmodellteiles omp_ν dient zur Positionsbestimmung des Objektmodellteiles. Daher muß dem primären Merkmal neben seinem zugehörigen 3D Modellmerkmal die Information mitgegeben werden, wie das zugehörige 2D Bildmerkmal aus den Sensordaten extrahiert und wie die Tiefeninformation für das 3D Szenenmerkmal bestimmt werden kann. Somit sind primäre Merkmale wie folgt definiert:

$$\mathbf{f}_\nu = \langle \mathbf{m}_\nu, IP_\nu, RESTR_\nu, QUAL_\nu \rangle \tag{2.9}$$

Mit den Bildverarbeitungsoperatoren $IP_\nu = \{ip_{\nu,1}(.), \ldots, ip_{\nu,n}(.)\}$ ist die Segmentierung der 2D Bildmerkmale aus einem Videobild img bestimmt. Für die im Interpretationsprozeß notwendige restriktive Beurteilung der Zuordnung von Szenenmerkmalen zu dem Modellmerkmal $\mathbf{m}_\nu$ kennt das primäre Merkmal noch Restriktionen $RESTR_\nu = \{restr_{\nu,1}(.), \ldots, restr_{\nu,m}(.)\}$. Diese Zuordnungen werden anschließend noch durch die Gütefunktionen $QUAL_\nu = \{qual_{\nu,1}(.), \ldots, qual_{\nu,m}(.)\}$ bewertet.

[7] Durch die Einschränkung auf symmetrische Punkt- und Flächenmerkmale reicht die Angabe des Verschiebungsvektors aus, um die Lage des Modellmerkmals im lokalen Koordinatensystem anzugeben; es kann die Orientierung, d.h. der Rotationsanteil der Lage vernachlässigt werden.

Im folgenden werden anhand von zwei Beispielen die Verwendung primärer Merkmale für ein Personenmodell kurz erläutert:

Ein signifikantes und charakteristisches Merkmal von Personen ist die Hautfarbe. Durch die exponierte Lage ist insbesondere der Kopf dazu geeignet, die Position einer Person zu bestimmen.[8] Das 3D Modellmerkmal, das hierbei dem Objektmodellteil $omp_{3.3}$ zugeordnet ist, ist eine "hautfarbene" 3D Ellipse. Dieses Merkmal ist ein implizites Merkmal. Somit ist die Größe der Ellipse durch die Größe des zugehörigen Volumenkörpers vol_{ell} bestimmt. Die Ellipse liegt per Definition in der xz-Ebene des lokalen Koordinatensystems des Objektmodellteiles des Kopfes. Nachdem der Ursprung des lokalen Koordinatensystems an der Stelle liegt, an der Hals und Kopf aufeinander treffen, entspricht der Verschiebungsvektor $\vec{t}$ in der Länge der halben Höhe des Kopfes und ist in Richtung der z-Achse $Z_{3.3}$ orientiert; vgl. Abb. 2.2. Das Basisattribut des Merkmals ist die Farbe. Somit werden im Interpretationsprozeß dem Modellmerkmal nur solche Szenenmerkmale zugeordnet, die aus Bildmerkmalen mit der entsprechenden Farbe bestimmt worden sind. Weitere Attribute $attr^{\mathbf{m}}$ sind neben der Größe der Ellipse und der Farbe auch Attribute der Form. Hierzu werden für eine Ellipse eine minimale und eine maximale Exzentrizität angegeben, um als Merkmal des Kopfes anerkannt zu werden.

Mit dem Bildverarbeitungsoperator, der dem primären Merkmal zugeordnet ist, werden in den zur Verfügung stehenden Videobildern "hautfarbene" Bildregionen segmentiert. Die 2D Bildmerkmale zeichnen sich durch den Mittelpunkt der Bildregion $\vec{p}_{img}$, die Kameraparameter $camPar$ und $camPose$ und die Attribute $attr^{\mathbf{i}}$, die sich für die Bildregion bestimmen lassen, aus. Die Attribute $attr^{\mathbf{i}}$ der Bildregion sind die *Farbe*, die *Fläche in Pixeln*, die beiden *Halbmesser* und der *Winkel* einer auf die Bildregion angewendeten Ellipsenanpassung, vgl. auch Kap. 3.4.3 zur Extraktion der Bildmerkmale.

Aus diesen Bildmerkmalen werden in geeigneter Weise Tiefeninformationen gewonnen, so daß ein 3D Szenenmerkmal $\mathbf{s}$ entsteht. Das Szenenmerkmal hat lediglich die Farbe der Ellipse als Attribut $attr^{\mathbf{s}}$. Alle weiteren Informationen sind in den mit Bildmerkmalen $\mathbf{i} \in \mathbf{I}_{extr}$ enthalten, die für die Tiefenschätzung verwendet wurden. Im Interpretationsprozeß werden dann die 3D Szenenmerkmale, die eine 3D Lage einer "hautfarbenen" Ellipse repräsentieren, mit dem 3D Modellmerkmal des Objektmodellteils, das den Kopf repräsentiert, verglichen. Bei diesem Vergleich werden alle zusätzlichen Attribute der verschiedenen Stufen der Merkmale verglichen und die Restriktionen $restr(.) \in RESTR$ angewendet. Hierzu werden, sobald eine Zuordnung vorgenommen wurde, die 3D Ellipsen des primären Merkmals anhand der Kameraparameter in den Sensorraum projiziert und dort mit den Daten der 2D Bildmerkmale verglichen. Weitere Einzelheiten sind hierzu in dem Kap. 3.6 zur Generierung von Hypothesen beschrieben.

Ein weiteres Beispiel für ein primäres Merkmal ist eine Markierung der Gelenke durch farbige Landmarken, wie z.B. farbige Bänder.[9] Das 3D Modellmerkmal ist dann ein "farbiger Fleck" (engl. *colored blob*). Diese Modellmerkmale sind explizite Merkmale, so daß als Attribut $attr^{\mathbf{m}}$ hier explizit die Größe anzugeben ist. Sofern die Form der Merkmale beliebig oder nicht eindeutig ist, werden keine weiteren Formattribute angegeben. Nachdem die Gelenke im inneren eines Körpers liegen, muß mit dem Verschiebungsvektor $\vec{t}$ die Position der Landmarke auf der Oberfläche des Körpers beschrieben werden. Bei der Markierung des Objektmodellteiles $omp_{1.2}$ für den Rumpf wird die Markierung beispielsweise auf dem Bauch vorgenommen. Somit muß in dem Verschiebungsvektor im lokalen Koordinatensystem des Objektmodellteiles $omp_{1.2}$ die Verschiebung in Richtung der y-Achse $Y_{1.2}$ berücksichtigt werden.

Für die zugehörigen 2D Bildmerkmale und 3D Szenenmerkmale gilt ähnliches, analog des

[8] Vgl. hierzu die in Kap. 4.2 beschriebene Anwendung.

[9] Vgl. hierzu die in Kap. 4.3.1 beschriebene Anwendung.

ersten Beispieles für das Modellmerkmale des Kopfes.

2.4.6 Sekundäre Merkmale der Objektmodellteile

Die sekundären Merkmale eines Objektmodellteils dienen im Gegensatz zu den primären Merkmalen nicht zur 3D Positionsbestimmung des Objektmodellteiles, sondern zur Überprüfung der im Interpretationsprozeß aufgestellten Hypothesen.

Hierzu ist die Lage des Objektmodellteils zuvor bestimmt worden, so daß bei den sekundären Merkmalen eine Zuordnung von 3D Szenenmerkmalen zu 3D Modellmerkmalen nicht mehr vorgenommen wird. Es ergibt sich vielmehr aus dem 3D Modellmerkmal zusammen mit der angenommenen Lage des Objektmodellteils ein 3D Szenenmerkmal, das dann im 3D Szenenraum oder 2D Sensorraum überprüft wird.

Es kann sich hierbei um die reine Überprüfung der Position eines Objektmodellteils im Weltkoordinatensystem handeln oder um eine weitergehende Verifikation mit den Sensordaten. Dies bedeutet, daß das entsprechende 3D Szenenmerkmal in den 2D Sensorraum des Videobildes projiziert und dort mit einem entsprechenden 2D Bildmerkmal verglichen werden muß. Diese Abbildung von 3D auf 2D ist eindeutig und direkt, basierend auf den Informationen der Kamerakalibrierung, ausführbar.

Somit ergibt sich für die sekundären Merkmale eines Objektmodellteils omp_ν folgende Definition:

$$\mathbf{f}'_\nu = \left\langle \mathbf{m}_\nu, IP_\nu, qual_{\mathbf{f}'_\nu(.)} \right\rangle \tag{2.10}$$

Mit $qual_{\mathbf{f}_\nu}(.)$ ist den sekundären Merkmalen eine Bewertungsregel bekannt, mit der das Ergebnis der Verifikation der Hypothese bestimmt wird. Sofern hierbei der Vergleich mit Sensordaten notwendig ist, kennt das sekundäre Merkmal mit $IP_\nu = \{ip_{\nu,1}(.), \ldots, ip_{\nu,n}(.)\}$ zusätzlich Bildverarbeitungsoperatoren zur Bestimmung von 2D Bildmerkmale. Im Interpretationsprozeß müssen weiterhin die Kameraparameter $camPar$ und $camPose$ für die Bildverarbeitungsmethoden zur Verfügung gestellt werden, damit die 3D Modellmerkmale in 2D Bildmerkmale projiziert werden können.

Bei der Verifikation der Lage eines Objektmodellteils kann es sich z.B. bei der Personendetektion um die Überprüfung des Abstandes der Objektmodellteile für die Füße von der xy-Ebene des Weltkoordinatensystems handeln. Ist im Interpretationsprozeß beispielsweise die "hautfarbene" Ellipse einer Hand dem Objektmodellteil des Kopfes zugeordnet worden, so kann über die vermeintliche Position der Füße festgestellt werden, daß diese hypothetische Zuordnung falsch ist. Ist die Hand unterhalb des "richtigen" Kopfes, so ergibt sich für die Füße eine negative Koordinate auf der z-Achse des Weltkoordinatensystems. Ist die Hand oberhalb des Kopfes, so "schweben" die Füße über der xy-Ebene.

Ein weiteres Beispiel für ein sekundäres Merkmal ist das Überprüfen des Vorhandenseins von bestimmten Objektmodellteilen, denen kein primäres Merkmal zugeordnet ist. So kann z.B. geprüft werden, ob unterhalb des Objektmodellteiles eines Kopfes, für den es eine hypothetische Zuordnung aufgrund des primären Merkmals der "hautfarbenen" Ellipse gibt, eine entsprechende Repräsentation des Objektmodellteils für den Rumpf vorhanden ist.[10] Nachdem dem Objektmodellteil des Rumpfes kein primäres Merkmal zugeordnet ist, kann lediglich das Vorhandensein anhand von unspezifischen Bildmerkmalen geprüft werden. Hierzu wird z.B. die zur Aufmerksamkeitssteuerung verwendete Vordergrundregion[11], die für die aktuellen Video-

[10] vgl. Beispiel zu primärem Merkmal in Kap. 2.4.5.

[11] vgl. Kap. 3.4.2 zur Bildsegmentierung.

bilder bestimmt worden ist, verwendet. Aufgrund der vermeintlichen Lage des Kopfes wird, basierend auf der bekannten Lage der Objektmodellteile zueinander, der Volumenkörper, der den Rumpf repräsentiert, in das Bild projiziert. Hierzu müssen die Parameter der Kamerakalibrierung bekannt sein. Die projizierte Region wird mit der Vordergrundregion verglichen. Falls sich nicht ein zuvor festgelegter Anteil der projizierten Region mit der Vordergrundregion deckt, so ist eine hypothetische Zuordnung für den Kopf nicht sehr wahrscheinlich und kann verworfen werden.

2.5 Kameramodell

Die Kameras $CAM = \{cam_1, \ldots, cam_n\}$ als Inventar des Szenenmodells $scene$ sind die Sensoren des Bildinterpretationssystems STABIL^{++}. Im Hinblick auf die Verwendung des Begriffes der *Kamera* in der formalen Beschreibung des Interpretationsprozesses und der Modellierung wird dieser hier näher erläutert und abgegrenzt.

Im folgenden werden zunächst die generellen und die vom Interpretationsprozeß genutzten Eigenschaften der bildgebenden Sensoren beschrieben. Daran schließt sich eine Klassifizierung der in STABIL^{++} realisierten Kameras bezüglich der Videoquelle und ihrer Aktivität an. Im letzten Abschnitt wird dann auf das eigentliche Modell einer *Lochkamera* eingegangen, mit dem die Abbildung der realen 3D Welt in das 2D Bild beschrieben wird.

2.5.1 Eigenschaften der Kameras

Als Kamera wird in dem System eine *bildgebende* Einheit betrachtet, die ein digitales Videobild img liefert. Der eigentliche Sensor dieser Einheit ist ein CCD-Array (CCD-Chip). Jedoch wird hier immer die komplette Abbildungskette von Objektiv, CCD-Chip, Signalübertragung bis hin zur Digitalisierung als Kamera bezeichnet. Dies gilt auch bei der *off-line* Verarbeitung von zuvor aufgezeichneten Bildern.

Eine der wichtigsten Eigenschaften der Kamera als Sensor in einem System zur 3D Objektdetektion sind die Abbildungseigenschaften. Mit den Abbildungseigenschaften einer Kamera wird beschrieben, wie die reale 3D Welt auf das zweidimensionale Videobild abgebildet wird. Um diese Abbildung zu beschreiben, verwendet man ein Kameramodell. Als Modellparameter wird zwischen den internen Kameraparametern $camPar$ und den externen Kameraparametern $camPose$ unterschieden. Mit $camPar$ wird generell die Abbildung eines 3D Punktes bestimmt, wobei $camPose$ die Lage, d.h. die Position und die Orientierung der Kamera im Weltkoordinatensystem angibt.

Einer Kamera werden in STABIL^{++} direkt sog. *low-level* Bildverarbeitungsoperatoren $IP = \{ip_1(.), \ldots, ip_n(.)\}$ zugeordnet. Mit diesen Operatoren wird direkt bei der Generierung der Videobilder eine Bildvorverarbeitung vorgenommen. Zur Vorverarbeitung zählt z.B. eine Beseitigung des *Interlace*-Effektes und eine Kontrastverstärkung.

Darüber hinaus werden die von einer Kamera aufgenommenen Videobilder in Bildbereiche[1] $\{REG^{attr_1}, \ldots, REG^{attr_n}\}$ mit verschiedenen Attributen segmentiert. Diese Segmentierung kann als Klassifikation bezeichnet werden, denn es werden die einzelnen Bildpunkte bestimmten Regionen zugeordnet, die aufgrund ihrer Attribute einer Klasse zugeordnet sind. Der Kamera sind daher Klassifikatoren $CLC = \{clc^{attr_1}, \ldots, clc^{attr_n}\}$ zugeordnet mit denen die entsprechenden Regionen $\{REG^{attr_1}, \ldots, REG^{attr_n}\}$ bestimmt werden können.[2] Die einer Kamera zugeordneten Klassifikatoren sind vom Objektmodell und dessen Merkmalen zunächst unabhängig und haben einen direkten Bezug zu einer einzelnen Kamera. Es muß jedoch sichergestellt sein, daß für alle durch die Objektmodelle verwendeten primären und sekundären Merkmale und die zugehörigen Bildverarbeitungsoperatoren die zu verwendenden Bildregionen mit den entsprechenden Attributen von der Kamera zur Verfügung gestellt werden.[3]

[1] Die Bildbereiche werden auch als Regionen bezeichnet, das verwendete Kürzel kann aus dem zugehörigen englischen Begriff abgeleitet werden: reg / REG – **reg**ion.

[2] Eine Ausnahme bilden die Farbklassifikatoren, denn mit einem Farbklassifikator clc^{color} werden entsprechend der diesem bekannten Farben / Farbklassen $color_i \in COLOR$ die Regionen $\{REG^{color_1}, \ldots, REG^{color_n}\}$ erzeugt, vgl. Kap. 3.4.2.

[3] Als Beispiel: Wird ein primäres Merkmal mit dem Basisattribut der Farbe “blau” verwendet, so muß durch einen

Eine mögliche Klassifikation ist die Einteilung des Bildes in Bereiche von Vorder- und Hintergrund. Dies wird auch als Aufmerksamkeitssteuerung bezeichnet, denn im Interpretationsprozeß kann die Suche der Bildmerkmale auf die Vordergrundbereiche eingeschränkt werden. Der direkte Bezug dieser Vorder- / Hintergrundtrennung zu einer Kamera ist durch eine bildfolgenbasierte Verarbeitung begründet.

Eine weitere Klassifikation ist die Segmentierung von Farbvideobildern in Bereiche mit dem Attribut einer bestimmten Farbe. Man spricht hier von einer Farbklassifikation. Auch hier kann die Verarbeitung von Bildfolgen ausgenutzt werden. Darüber hinaus ist jedoch die direkte Zuordnung der Farbklassifikation zu einer Kamera durch die unterschiedliche Farbabbildung der Sensoren begründet. Die Verfahren zur Aufmerksamkeitssteuerung und zur Farbklassifikation sind im Kap. 3.4 erläutert.

Für die Definition einer Kamera cam_ν ergibt sich somit das Tupel:

$$cam_\nu = \langle camPar_\nu, camPose_\nu, IP_\nu(.), CLC_\nu, type_\nu \rangle \, . \tag{2.11}$$

Wobei $type_\nu$ einer Einteilung der Kameras entspricht, die in Abschn. 2.5.3 vorgenommen wird.

2.5.2 Sensordaten: Bilder

Die verwendeten Sensordaten in STABIL^{++} sind generell die Helligkeits- und Farbinformationen in Videobildern. Jedoch wird als Bild img die Einheit aus den eigentlichen Helligkeits- und Farbinformationen und allen Zusatzinformationen, die einem Bild bei der Generierung durch die Kamera mitgegeben werden, bezeichnet. Das Bild ist somit als das Tupel

$$img = \left\langle CAN, camPar, camPose, \{REG^{attr_1}, \ldots, REG^{attr_n}\}, t \right\rangle \tag{2.12}$$

definiert.

Mit $CAN = \{can_1, \ldots, can_n\}$ beinhaltet ein Bild einzelne Bildkanäle. Entsprechend der bei der Aufzeichnung in der Kamera ausgeführten Bildvorverarbeitung sind die Inhalte der Bildkanäle aufbereitet. Die Anzahl der Kanäle ist von der Kamera und der verwendeten Digitalisierungskarte (*framegrabber*) bestimmt. Bei der Anwendung des Systems z.B. zur Personendetektion werden sinnvollerweise Schwarz-Weiß-Kameras (SW-Kameras) oder Farbkameras verwendet. Das Signal einer SW-Kamera wird bei der Digitalisierung in einen Grauwertkanal aufgezeichnet. Bei der Verwendung von Farbkameras werden in der Regel drei Grauwertkanäle aufgezeichnet. Die Bedeutung der einzelnen Kanäle ist durch ein Farbmodell auch Farbraum genannt festgelegt. Für die rechnerinterne Darstellung wird meist der *RGB*-Farbraum verwendet. In diesem Farbraum wird in einem Kanal der Rotanteil, in einem weiteren Kanal der Grünanteil und in dem dritten Kanal der Blauanteil kodiert. Weitere, gebräuchliche Farbräume und deren Verwendung sind in [Haf99] nachzulesen.

Dem Bild sind die Modellparameter des Kameramodells, die zum Zeitpunkt der Aufnahme des Bildes für die entsprechende Kamera aktuell Gültigkeit hatten, mit $camPar$ und $camPose$ bekannt. Damit sind für das Bild die Abbildungsvorschriften der 3D Welt auf der Grundlage des verwendeten Kameramodells bekannt.

Die Ergebnisse der bei der Kamera definierten Klassifikatoren sind in dem Bild in den Regionen $\{REG^{attr_1}, \ldots, REG^{attr_n}\}$ bekannt. Daher kann bei der Verwendung einer Vorder- /

Klassifikator der Kamera eine Bildregion mit dem Attribut “blau” bestimmt werden können. Der Klassifikator und die Farbklasse “blau” ist von dem Objektmodell und den Merkmalen unabhängig und hat nur einen Bezug zur Kamera.

Hintergrundtrennung in der Kamera im Interpretationsprozeß direkt auf Bildregionen zugegriffen werden, die als Vordergrund klassifiziert worden sind. Bei der Verwendung eines Farbklassifikators sind dem Bild zusätzlich Regionen mit dem Attribut einer bestimmten Farbe bekannt. Die einzelnen Regionen werden wie folgt entsprechend der Attribute gekennzeichnet:

$reg^{(fg)} \in REG^{(fg)}$	Einzelregion, mit dem Attribut “dem Vordergrund zugehörig”
$reg^{(red)} \in REG^{(red)}$	Einzelregion, mit dem Attribut “der Klasse ROT zugehörig”
$reg^{(cyan)} \in REG^{(cyan)}$	Einzelregion, mit dem Attribut “der Klasse CYAN zugehörig” .

Desweiteren ist dem Bild mit dem Zeitstempel t der Zeitpunkt der Generierung bekannt. Für diesen Zeitpunkt kann die Systemzeit zum Zeitpunkt der Digitalisierung verwendet werden. Jedoch muß über den Zeitstempel lediglich die Differenz der Aufnahmezeitpunkte aufeinander folgender Bilder einer Bildsequenz erfasst sein.

Im Interpretationsprozeß werden die Modellmerkmale aus den primären Merkmalen eines Objektmodellteils zu den Szenenmerkmalen zugeordnet. Über die in den Szenenmerkmalen vermerkten Bildmerkmale besteht daher ein eindeutiger Bezug zu einem Bild. Der Zeitstempel des Bildes hat somit auch Gültigkeit für die ermittelten Positionsdaten des Objektmodellteils, vgl. Glg. 2.5.

2.5.3 Einteilung der Kameras

In STABIL^{++} wird zwischen verschiedenen Arten der Kameras unterschieden. Es wird daher eine Einteilung oder Klassifizierung nach der Art der Videoquelle und nach einer möglichen Aktivität der Kamera vorgenommen. Die Zugehörigkeit zu einer bestimmten Gruppe ist einer Kamera cam_ν mit $type_\nu$ bekannt.

Videoquelle

Es kann generell zwischen einer *on-line* und einer *off-line* Verarbeitung der Videobilder unterschieden werden. Bei der *on-line* Verarbeitung ist das analoge Videosignal einer realen Kamera an eine sog. Digitalisierungskarte (*framegrabber*) angeschlossen. Im Bildinterpretationsprozeß wird zu Beginn eines Verarbeitungsschrittes von der Kamera ein Bild angefordert, welches zu diesem Zeitpunkt digitalisiert wird. Man spricht daher in STABIL^{++} auch von einer sog. *Live-Camera*. Der Zeitstempel des Bildes t entspricht der Live-Camera der aktuellen Systemzeit.

Bei der *off-line* Verarbeitung werden die Videobilder vor dem eigentlichen Bildinterpretationsprozeß aufgezeichnet und als Folge von Dateien gespeichert. Bei der Bildinterpretation werden dann die Bilder in der Reihenfolge der Aufzeichnung verarbeitet. Die Kameras der *off-line* Verarbeitung werden in STABIL^{++} als *File-Camera* bezeichnet.

Die File-Camera muß sich im System mit einer Live-Camera identisch verhalten. Daher ist sicherzustellen, daß für jedes Bild die Parameter des Kameramodells $camPar$ und $camPose$ bekannt sind. Es wird sinnvollerweise bei der Aufzeichnung nur mit stationären Kameras mit fester Brennweite gearbeitet. Somit sind die Kameraparameter für eine komplette Bildfolge identisch. Ein aktives Ansteuern der Kamera ist bei der *off-line* Verarbeitung nicht möglich.

Im weiteren muß für jedes Bild ein Zeitstempel t gesetzt werden. Die Zeitstempel zweier in einer Bildfolge aufeinander folgender Bilder muß relativ der Differenz der Aufnahmezeit entsprechen und nicht der Verarbeitungszeit. Daher ist es sinnvoll die Bilder in einem festen Zeitraster aufzunehmen. Für das erste Bild der Bildfolge wird der Zeitstempel auf null gesetzt werden, wobei er sich für jedes weitere Bild um die zeitliche Differenz der Aufnahmezeitpunkte erhöht.

Die *off-line* Verarbeitung kann zum einen zu Testzwecken verwendet werden, bei der unterschiedliche Parametrierung des Systems auf gleichbleibenden Sensordaten überprüft werden. Zum anderen kann die *off-line* Verarbeitung genutzt werden, um die maximale Anzahl der von den Kameras zur Verfügung stehenden Bilder zu verarbeiten, wenn bei komplexen Objektmodellen die Verarbeitungsgeschwindigkeit des Systems abnimmt und bei einer *on-line* Verarbeitung nicht mehr alle Bilder verarbeitet werden können. Dies ist insbesondere bei Anwendungen zur Analyse von Bewegungsabläufen notwendig, wenn eine hohe Genauigkeit erzielt werden muß. Zur Analyse von kurzen Bewegungsabläufen können die zuvor aufgenommenen Bilder in genügend großem Hauptspeicher gehalten werden und anschließend sofort interpretiert werden. Man spricht in STABIL^{++} dann von einer sog. *Memory-Camera*. Es gelten hier jedoch die gleichen Einschränkungen der File-Camera. Mit diesem Konzept der Einteilung der Kameras mit den entsprechenden Eigenschaften ist es auch möglich, Hochgeschwindigkeitskameras zu verwenden. Hierbei werden die digitalen Bilder in einer Spezial-HW zwischengespeichert und nach und nach von dem System abgearbeitet.

Aktivität

In STABIL^{++} wird desweiteren zwischen stationären und aktiven Kameras unterschieden. Bei einer stationären Kamera kann weder die Brennweite noch die Lage der Kamera verändert werden, d.h. die Parameter des Kameramodells $camPar$ und $camPose$ verändern sich nicht. Diese Kameras werden im System als *Fix-Camera* bezeichnet. Wie im letzten Abschnitt angedeutet, gelten für alle Bilder einer zuvor gespeicherten Bildfolge sinnvollerweise die gleichen Parameter des Kameramodells. Damit ist eine File-Camera immer eine Fix-Camera.

Live-Cameras können jedoch auch aktive Kameras sein. Man spricht dann von den sog. *Active-Cameras*. Die möglichen Freiheitsgrade einer aktiven Kamera unterteilen sich in Freiheitsgrade der Optik und die Freiheitsgrade der Lage. Die Optik ist generell variabel in

- der Brennweite (Zoom-Objektive)
- der Fokusierung und
- der Blendenöffnung.

Es ist zu beachten, daß Veränderungen der Optik auch die Abbildungseigenschaften der Kamera verändern, daher müssen die internen Kameraparameter $camPar$ bei veränderter Optik entsprechend angepaßt werden. Die Veränderungen haben ebenso Auswirkungen auf die Klassifikatoren CLC, die einer Kamera zugeordnet sind, denn durch eine Veränderung der Abbildung kann es zu einer Verschiebung der Klassen im Merkmalsraum kommen.

Die Kameralage und -position kann durch Translationen und Rotationen verändert werden. So kann es sich bei der Kamera um eine Schwenk- / Neigekamera handeln oder sogar um eine Kamera, die auf einem mobilen System montiert ist und bei der alle sechs Freiheitsgrade der Bewegung variabel sind. Dies sind

- drei Freiheitsgrade der Translation (in drei Richtungen des kartesischen Koordinatensystems)
- drei Freiheitsgrade der Rotation (um drei Achsen des kartesischen Koordinatensystems)

Beim Erzeugen eines Bildes muß die Lage der Kamera bezüglich des Weltkoordinatensystems bekannt sein. Daher müssen die äußeren Kameraparameter $camPose$ bei einer Veränderung der Lage der Kamera angepaßt werden.

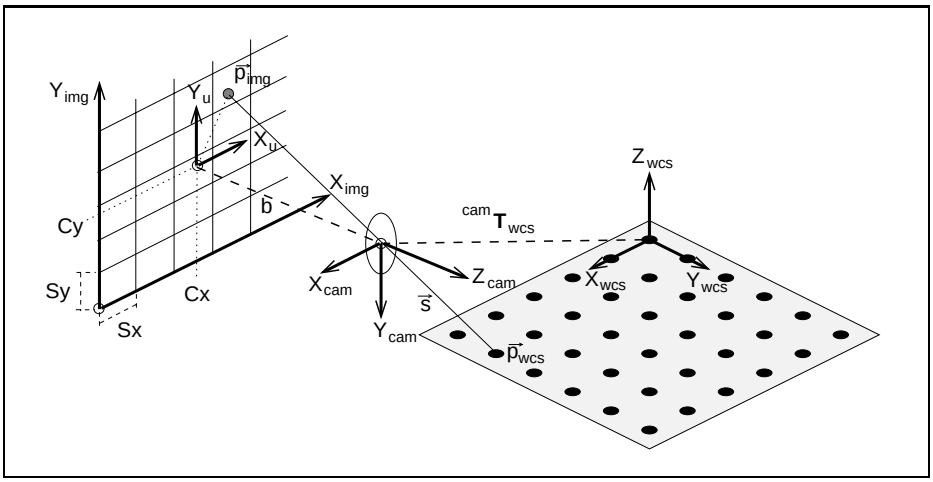


Abbildung 2.3: *Modell einer Lochkamera mit radialer Verzerrung: Abbildung eines Punktes $\vec{p}_{wcs}$ im Weltkoordinatensystem auf einen Bildpunkt $\vec{p}_{img}$ im Rechnerkoordinatensystem; die Koordinatensysteme $[X_u, Y_u]$ und $[X_{img}, Y_{img}]$ sind entsprechend des Lochkameramodells spiegelverkehrt zum Kamerakoordinatensystem $[X_{cam}, Y_{cam}]$ orientiert.*

Entsprechend der Anwendung werden unterschiedliche Kameras eingesetzt. Bei der Detektion und der Verfolgung von Personen mit STABIL^{++} ist ein Anwendungsbereich die Videoüberwachung in der Sicherheitstechnik. In diesem Bereich ist der Einsatz von Schwenk- / Neigekameras mit Zoom-Objektiven üblich. Bei diesen Kameras sind nicht alle möglichen Freiheitsgrade der allgemeinen Active-Camera ausgenutzt, so daß es sich hier um eine Einschränkung handelt.

2.5.4 Lochkamera

Mit einer Videokamera wird ein zweidimensionales Abbild der dreidimensionalen Umwelt erzeugt. Man erhält vereinfacht ausgedrückt ein zweidimensionales Signalfeld von Beleuchtungsintensitäten. Um jedoch diese Signalinformation, insbesondere im Hinblick auf die 2D/3D Abbildung, zu deuten, muß man mit Hilfe eines mathematischen Modells den Vorgang der Abbildung beschreiben. Im STABIL^{++} wird hierzu ein Lochkameramodell mit radialer Verzerrung verwendet, wie es von in [Len87] vorgestellt ist. Im einzelnen wird hier die Implementierung verwendet, wie sie von in [Lan98] vorgestellt ist. Daher beziehen sich die Ausführungen in den folgenden Abschnitten auch auf diese Arbeiten.[4]

Projektion

Die Abbildung / Projektion eines 3D Punktes im Weltkoordinatensystem $\vec{p}_{wcs}$ in das diskrete Kamerabild im Rechner und somit auf einen 2D Punkt[5] $\vec{p}_{img}$ läßt sich im Lochkameramodell

[4] Das Kameramodell mit zugehöriger Kalibrierung ist ursprünglich im Rahmen des Teilprojektes L9 des SFB 331 in der Forschungsgruppe Bildverstehen an der Technischen Universität München von St. Lanser und Ch. Zierl entwickelt worden, vgl. auch [SF96], den Konferenzband zur AMS'96 mit Sonderbeiträgen zu den technischen Demonstrationen aus dem Sonderforschungsbereich 331 "Informationsverarbeitung in autonomen, mobilen Handhabungssystemen" und [LZB95].

[5] pixel.

mit radialer Verzerrung in 4 Schritte einteilen, vgl. hierzu Abb. 2.3[6]:

1. Transformation des 3D Weltpunktes $\vec{p}_{wcs}$ in einen 3D Punkt $\vec{p}_{cam}$ im lokalen Koordinatensystem der Kamera
2. Perspektivische Projektion des 3D Punktes $\vec{p}_{cam}$ in einen 2D Punkt $\vec{u}$ in der Bildebene (Bildkoordinatensystem)
3. Berücksichtigung der Linsenfehler in Form der radialen Verzerrung mit Überführung des Punktes $\vec{u}$ in $\vec{v}$
4. Diskretisierung des Punktes $\vec{v}$ und damit Abbildung auf den Bildpunkt $\vec{p}_{img}$ im Rechnerkoordinatensystem.

Im ersten Schritt wird bei der Transformation des abzubildenden Punktes vom Weltkoordinatensystem in das lokale Koordinatensystem der Kamera die Lage der Kamera und damit des Kamerakoordinatensystems im Weltkoordinatensystem als bekannt vorausgesetzt. Die Lage und somit die Transformation kann durch eine homogene Transformationsmatrix ${}^{cam}\mathbf{T}_{wcs}$ beschrieben werden, so daß gilt:

$$\vec{p}_{cam} = {}^{cam}\mathbf{T}_{wcs} \cdot \vec{p}_{wcs} \tag{2.13}$$

Die sechs Freiheitsgrade der Transformation werden als die *äußeren Kameraparameter*

$$camPose = \left\langle [x, y, z]^T, \alpha, \beta, \gamma \right\rangle \tag{2.14}$$

des Kameramodells bezeichnet. Im einzelnen sind das der Translationsvektor $[x, y, z]^T$ und die drei Rotationswinkel α, β, γ.

Mit der Angabe des Translationsvektors und der drei Winkel ist die Transformation noch nicht eindeutig bestimmt. Es muß zusätzlich noch die Basis der Transformation angegeben werden, d.h. ob zuerst die Translation oder zuerst die Rotation ausgeführt wird, in welcher Reihenfolge die Rotationen angewendet werden und welchen Bezug die Winkelangaben haben. Es hat sich im Allgemeinen als sinnvoll erwiesen, die Winkel in der $zy'z''$-Notation anzugeben und die Translation vor der Rotation auszuführen. Somit kann man sich die äußeren Kameraparameter wie folgt verdeutlichen: Um die Lage des Kamerakoordinatensystems im Weltkoordinatensystem zu beschreiben, wird zunächst ein Koordinatensystem mit dem Weltkoordinatensystem in Deckung gebracht. Dieses System wird dann im Weltkoordinatensystem um den Vektor $[x, y, z]^T$ verschoben. Anschließend wird eine Drehung des verschobenen Koordinatensystems mit dem Winkel γ um die z-Achse des verschobenen Koordinatensystems vorgenommen. Daran schließt sich eine Drehung mit dem Winkel β um die y-Achse des bereits einmal gedrehten Koordinatensystems an. Schließlich wird noch eine Drehung mit dem Winkel α um die z-Achse des bereits zweimal gedrehten Koordinatensystems ausgeführt. Das so transformierte Koordinatensystem entspricht dann dem Kamerakoordinatensystem mit den durch $[x, y, z]^T$ und α, β, γ gegebenen äußeren Kameraparametern. Vergleiche hierzu die Ausführungen im Anh. B zur Koordinatentransformation.

Die weiteren Parameter des Kameramodells, die in den anderen drei Abbildungsschritten verwendet werden, werden als die *internen Kameraparameter*

$$camPar = \left\langle b, \kappa, S_x, S_y, [C_x, C_y]^T \right\rangle \tag{2.15}$$

[6] Abb. 2.3 ist in ursprünglicher Darstellung in [LZB95] zu finden.

bezeichnet. Im einzelnen sind dies:

- die *Kammerkonstante* b,
- der *Verzerrungskoeffizient* κ,
- die *Skalierungsfaktoren* S_x, S_y und
- der *Hauptpunkt* $[C_x, C_y]^T$.

Alle Parameter des Kameramodells werden im Zuge der Kamerakalibrierung bestimmt; diese ist im Anh. C beschrieben.

Mit diesen Parametern ergeben sich die weiteren Abbildungsschritte wie folgt: Im zweiten Schritt wird unter Berücksichtigung der sog. *Kammerkonstante* b, der 3D Punkt aus dem Kamerakoordinatensystem $\vec{p}_{cam} = [x_{cam}, y_{cam}, z_{cam}]^T$ durch die perspektivische Projektion in die Bildebene projiziert. Man erhält damit das zweidimensionale Abbild $\vec{u} = [u_x, u_y]^T$. Es gilt:

$$\vec{u} = \begin{pmatrix} u_x \\ u_y \end{pmatrix} = \begin{pmatrix} b\ \frac{x_{cam}}{z_{cam}} \\ b\ \frac{y_{cam}}{z_{cam}} \end{pmatrix} \tag{2.16}$$

Es sei noch angemerkt, daß die Kammerkonstante b des Lochkameramodells mit der optischen Brennweite des Objektives zu vergleichen ist, jedoch nicht gleichzusetzen ist.

In dem dritten Schritt der Abbildung werden die radialen Verzerrungen der Linsen berücksichtigt. Es gilt:

$$\vec{v} = \begin{pmatrix} v_x \\ v_v \end{pmatrix} = \begin{pmatrix} \frac{2\ u_x}{1\ +\ \sqrt{1\ -\ 4\ \kappa\ \left(u_x^2\ +\ u_y^2\right)}} \\ \frac{2\ u_y}{1\ +\ \sqrt{1\ -\ 4\ \kappa\ \left(u_x^2\ +\ u_y^2\right)}} \end{pmatrix} \tag{2.17}$$

Mit dem Verzerrungskoeffizienten κ ist die Verzeichnung modelliert, es gilt für eine kissenförmige Verzeichnung $\kappa > 0$ und für eine tonnenförmige Verzeichnung $\kappa < 0$.

Im vierten Schritt der Abbildung erfolgt die Diskretisierung, d.h. die Abbildung des Punktes $\vec{u}$ aus dem Bildkoordinatensystem in das Rechnerkoordinatensystem. Hierbei werden zum einen die Skalierungsfaktoren S_x und S_y berücksichtigt, mit denen das Seitenverhältnis der virtuellen Bildpunkte des Kameramodells beschrieben ist. Zum anderen wird der Hauptpunkt mit $[C_x, C_y]^T$ berücksichtigt, mit dem das Zentrum der Abbildung bestimmt ist. Dieses Zentrum ist der Mittelpunkt der radialen Verzerrung. Für den Bildpunkt $\vec{p}_{img} = [p_{img_x}, p_{img_y}]^T$ im diskreten Kamerabild des Rechners gilt dann:

$$\vec{p}_{img} = \begin{pmatrix} x_{img} \\ y_{img} \end{pmatrix} = \begin{pmatrix} \frac{v_x}{S_x}\ +\ C_x \\ \frac{v_y}{S_y}\ +\ C_y \end{pmatrix} \tag{2.18}$$

Umgekehrte Projektion / Bestimmung von Sichtstrahlen

Eine eigentliche Umkehrung der Projektion gibt es nicht, man kommt lediglich von dem 2D Punkt $\vec{p}_{img}$ auf einen Sichtstrahl $\vec{s}$. Dieser Sichtstrahl verläuft von dem Punkt $\vec{p}_{img}$ in der Bildebene durch den Ursprung des Kamerakoordinatensystems. Vgl. hierzu auch die Darstellung in

Abb. 2.3. Definiert man einen Punkt $\vec{p}^{\,i}_{cam}$, der die Lage des Bildpunktes $\vec{p}_{img}$ im Kamerakoordinatensystem angibt, so ergibt sich für diesen unter Berücksichtigung der Kammerkonstante b:

$$\vec{p}^{\,i}_{cam} = \begin{pmatrix} 0 \\ 0 \\ -b \end{pmatrix} - \begin{pmatrix} x_{img} \\ y_{img} \\ 0 \end{pmatrix} = - \begin{pmatrix} x_{img} \\ y_{img} \\ b \end{pmatrix} \tag{2.19}$$

Für den Sichtstrahl $\vec{s}_{cam}$ im Kamerakoordinatensystem gilt dann:

$$\vec{s}_{cam} = [0,0,0]^T - \vec{p}^{\,i}_{cam} = -\vec{p}^{\,i}_{cam} = \begin{pmatrix} x_{img} \\ y_{img} \\ b \end{pmatrix} \tag{2.20}$$

Mit den äußeren Kameraparametern erhält man über die zugehörige Transformationsmatrix den Sichtstrahl $\vec{s}_{wcs}$ im Weltkoordinatensystem entsprechend:

$$\vec{s}_{wcs} = {}^{wcs}\mathbf{T}_{cam} \cdot \vec{s}_{cam} \tag{2.21}$$

Hierbei ist mit ${}^{wcs}\mathbf{T}_{cam}$ die inverse Koordinatentransformation zu der in Glg. 2.13 angewendeten Transformation beschrieben, vgl. hierzu die Ausführungen im Anh. B.3 zur homogenen Koordinatentransformation.

2.6 Zusammenfassung

Die in diesem Kapitel vorgestellte Modellierung ist die Grundlage für die modellbasierte Interpretation der Szene mit STABIL^{++}. Mit dem vorgestellten Objektmodell ist definiert, welche Objekte zu detektieren sind. Hierbei ist das Modell anhand einer hierarchischen, inneren Struktur aus einzelnen Objektmodellteilen aufgebaut. Für die Interpretation sind jedem Objektmodellteil Merkmale zugeordnet, mit denen die Erscheinung des Modells charakterisiert ist. Über diese Merkmale wird daher auch beschrieben, wie sich das Objektmodell im 2D Videobild darstellt. Die eigentliche Interpretation wird jedoch im 3D Raum vorgenommen, so daß man einen Übergang von 2D Merkmalen im Sensorraum zu 3D Merkmalen im Szenenraum bilden muß. Es wird daher zwischen 2D Bildmerkmalen, 3D Szenenmerkmalen und 3D Modellmerkmalen unterschieden.

Mit dem eingeführten Modell der Szene wird Wissen über die Observierungsräume zur Verfügung gestellt, das im Interpretationsprozeß genutzt wird. Ebenso sind mit dem Inventar des Szenenmodells die zur Verfügung stehenden Kameras bekannt, mit denen der Sensorraum bestimmt ist. Zur Beschreibung der Abbildungseigenschaften der Kameras und damit der Projektion der 3D Szene in das 2D Videobild ist ein entsprechendes Kameramodell eingeführt worden.

Im folgenden Kapitel wird nun beschrieben, wie die Interpretation der Szene zur Detektion und Verfolgung der Objekte ausgeführt wird, dies basiert auf der eingeführten Modellierung der Objekte, der Szene und der Sensoren.

3 Interpretationsprozeß

3.1 Prozeßablauf

3.1.1 Grundlagen

Der Interpretationsprozeß in STABIL^{++} ist die auf dem Modellwissen basierende Interpretation der Sensordaten, aus der die Interpretation der Szene resultiert. Grundlage ist hierbei das Objektmodell, das so ausgelegt ist, daß es alle Informationen zur Detektion und Verfolgung eines Objektes beinhaltet. Das Modell der Szene stellt zusätzlich mit dem Wissen über die Observierungsräume ein "Weltwissen" zur Verfügung. Dem Szenenmodell sind ebenfalls die zur Verfügung stehenden Kameras CAM und die initialen Suchräume SSP^0 bekannt. Weiterhin werden von dem Szenenmodell die gefundenen Objektmodellinstanzen OBJ verwaltet. Somit steuert das Szenenmodell in STABIL^{++} den Interpretationsprozeß, obwohl das Objektmodell sich, aufgrund des dort gekapselten Wissens, selbst in den Sensordaten finden oder wiederfinden kann.[1] Es wird im weiteren von der initialen Detektion eines Objektmodells und der Re-Detektion von Objektmodellinstanzen gesprochen.

Der Interpretationsprozeß teilt sich in sich immer wiederholende Interpretationsschritte, man spricht daher von einem Interpretationszyklus. Für jeden neuen Interpretationszyklus wird ein Satz neuer Sensordaten verwendet, daher werden zunächst neue Videobilder aufgenommen oder aus dem Speicher geholt. Daran schließt sich die Detektion oder Re-Detektion der Objektmodellinstanzen an. Das Ergebnis der Interpretation sind die einzelnen Objektmodellinstanzen, denen ihre Position in dem Weltkoordinatensystem des Szenenmodells und ihre Haltung, in Form der Objektmodellstrukturen, bekannt sind. Zur Auswertung des Ergebnises gibt das Szenenmodell am Ende eines Interpretationszyklus die Informationen über die Objektmodellinstanzen an ihre Akteure ACT weiter.

Entsprechend Abb. 3.1 teilt sich ein Interpretationszyklus in die drei Teilschritte der Bildgenerierung, der Detektion und der Aktion. Zusätzlich zu den drei Teilschritten des Interpretationszyklus werden übergeordnet von dem Szenenmodell noch die Objektmodellinstanzen verwaltet. Der Ablauf eines Interpretationszyklus läßt sich entsprechend Alg. 3.1 darstellen, wobei vordergründig die Verwaltung der Objektmodellinstanzen zu erkennen ist. Die drei Teilschritte sind dort als weiter zu spezifizierende Funktionen verwendet. So werden in Zeile 13 die Bilder von den Kameras eingezogen oder aus dem Speicher geholt. In Zeile 15 wird die Detektion von schon bestehenden Objektmodellinstanzen und in Zeile 22 für das initiale Objektmodell aufgerufen. Der Teilschritt der Aktion wird dann ab der Zeile 28 durchgeführt.

Auf den Großteil der Schritte des Algorithmus und somit auf die Verwaltung der Objektmodellinstanzen durch das Szenenmodell wird in dem folgenden Abschnitt genauer eingegangen. Die Spezifikation der drei Teilschritte eines Interpretationszyklus schließt sich daran an. Aufgrund der aus Abb. 3.1 und dem Alg. 3.1 erkennbaren prozeduralen Abarbeitung der ein-

[1] Vgl. hierzu auch die Systemübersicht in Abb. 1.3.

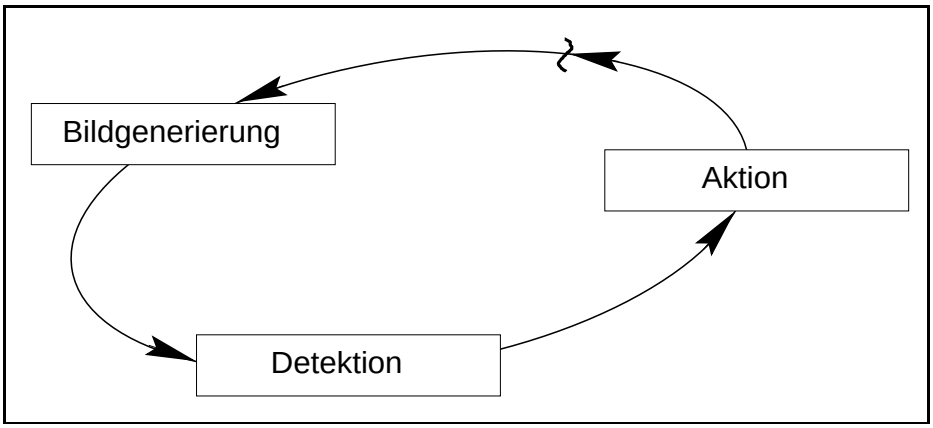


Abbildung 3.1: *Der Interpretationszyklus mit den drei Teilschritten.*

zelnen Schritte ist der zeitliche Ablauf der Interpretation unmittelbar durch die Komplexität der einzelnen Schritte bestimmt. Maßgeblich hängt die Zeitdauer eines Interpretationszyklus, neben den Sensordaten, von der Feinheit der verwendeten Modellierung ab. Die notwendige Feinheit ist wiederum von der Anwendung des Systems abhängig; es sei hierzu auf das Kap. 4 der Anwendungen verwiesen.

In STABIL^{++} muß jedem einzelnen Interpretationszyklus eine Zeit zugeordnet werden. Dies begründet sich zum einen durch die in STABIL^{++} bei der Detektion verwendeten Vorhersage der Positionen der Objektmodellteile. Zum anderen ist dies durch die Möglichkeit begründet, über die Historien $HIST$ der Objektmodellinstanzen und den zugehörigen Objektmodellteilen deren Bewegungen nachzuvollziehen. Es wird daher für die aktuell in der Szene gefundenen Objektmodellinstanzen die Zeit des Interpretationszyklus jeweils übernommen.

Für den Interpretationszyklus ist diese Zeit mit den Sensordaten, den Videobildern, festgelegt. Entsprechend den Ausführungen in Kap. 2.5.2 ist jedem Bild img ein Zeitstempel t mitgegeben, der bei einer Live-Camera der aktuellen Systemzeit und bei einer File- oder Memory-Camera jedoch den Aufnahmezeitpunkten entspricht. Der Zeitstempel und somit die Zeiten zweier aufeinanderfolgender Interpretationszyklen sind nur bei der Live-Camera von der Zeitdauer eines Interpretationszyklus abhängig.

3.1.2 Verwaltung der Objektmodellinstanzen

Mit der Verwaltung der Objektmodellinstanzen durch das Szenenmodell sind auch die verschiedenen Verarbeitungsmodi des Systems stark verbunden. Es wird zwischen einer initialen Detektion, der Re-Detektion von Objektmodellinstanzen, dem Verwerfen von Objektmodellinstanzen und der Verfolgung von Objektmodellinstanzen unterschieden. Die einzelnen Modi werden im folgenden anhand des Ablaufes eines Interpretationszyklus entsprechend Alg. 3.1 beschrieben.

Initiale Detektion

Beim Start des Systems ist die Liste der gefundenen Objektmodellinstanzen dem Szenenmodell $OBJ = \emptyset$ und dem Szenenmodell ist mit obj_0 nur ein initiales Objektmodell bekannt. Anhand des initialen Objektmodells wird dann im ersten Interpretationszyklus versucht, ein Objekt zu detektieren. Hierzu muß das initiale Modell an eine oder mehrere Positionen im Observierungsraum SSP_s der Szene gestellt werden, so daß sich die initialen Modellsuchräume SSP^0 des

Eingabe: $scene = \langle SSP_s, obj_0, SSP^0, OBJ, wcs, CAM, ACT \rangle$ // *Inventar der Szene*
Eingabe: q_{min} // *minimale Qualität einer Objektmodellinstanz*
Eingabe: $mode$ // *Verarbeitungsmodus z.B.: 'Verfolge'*
Eingabe: f // *Nummer der zu verfolgenden Objektmodellinstanz*
Berechne: $OBJ \Leftarrow$ Interpretation der Szene.

1: $IMG \Leftarrow \emptyset$ // *Liste von Bildern*
2: $SSP \Leftarrow \emptyset$ // *Gesamtsuchraum*
3: **falls** $(mode =$ 'Verfolge' $) \wedge (OBJ \neq \emptyset)$ **dann**
4: $OBJ \Leftarrow \{obj_f\}$ // *Verwerfe alle Objektmodellinstanzen* $\in OBJ$ *bis auf* obj_f
5: •
6: **für alle** $obj_i \in OBJ$ **wiederhole**
7: $SSP_i \Leftarrow$ Bestimme Suchraum für obj_i // ***s. Kap. 3.2***
8: $SSP \Leftarrow SSP \cup SSP_i$ // *Suchraum aller Objektmodelle*
9: •
10: **falls** $(OBJ = \emptyset) \vee (mode \neq$'Verfolge') **dann**
11: $SSP \Leftarrow SSP \cup SSP^0$ // *initiale Suchräume hinzufügen*
12: •
13: $IMG \Leftarrow$ Generiere Bilder von allen $cam_k \in CAM$ für SSP // ***s. Kap. 3.1.3***
14: **für alle** $obj_i \in OBJ$ **wiederhole**
15: $obj_i \Leftarrow$ Detektion von obj_i basierend auf IMG // ***s. Kap. 3.1.4***
16: **falls** $q_i < q_{min}$ **dann**
17: $OBJ \Leftarrow OBJ \setminus \{obj_i\}$ // *behalte Objektmodellinstanzen nur mit Mindestgüte*
18: •
19: •
20: **falls** $(OBJ = \emptyset) \vee (mode \neq$'Verfolge') **dann**
21: **wiederhole**
22: $obj_{neu} \Leftarrow$ Detektion von obj_0 basierend auf IMG // ***s. Kap. 3.1.4***
23: **falls** $q_{neu} \geq q_{min}$ **dann**
24: $OBJ \Leftarrow OBJ \cup obj_{neu}$ // *füge neue Objektmodellinstanz hinzu*
25: •
26: **bis** $q_{neu} < q_{min}$
27: •
28: **für alle** $act_l \in ACT$ **wiederhole**
29: Führe Aktion mit allen $obj_i \in OBJ$ aus // ***s. Kap. 3.1.5***
30: •

Algorithmus 3.1: *Ablauf eines Interpretationszyklus / Verwaltung der Objektmodellinstanzen.*

Szenenmodells ergeben.[2]

Die initialen Modellsuchräume setzen sich aus den kugelförmigen Suchräumen für die einzelnen Objektmodellteile des initialen Objektmodells, für die auch Merkmale definiert sind, zusammen. Der Durchmesser der Einzelsuchräume kann variiert werden, so daß Objekte oder sogar Objektmodellteile durch das System entweder nur an bestimmten Positionen oder, bei der Wahl von sehr großem Durchmesser, im gesamten Observierungsraum erwartet werden. Die

[2] Anm.: Die Systemstruktur läßt es auch zu, verschiedene initiale Objektmodelle gleichzeitig zu verwenden, diese werden jedoch entsprechend ihrer Reihenfolge bei der Detektion bevorzugt behandelt, vgl. auch die Beschreibung der Re-Detektion von Objektmodellinstanzen.

Wahl der verwendeten Positionen des initialen Objektmodells und die Wahl der Durchmesser für die initialen Suchräume sind anwendungsabhängig.

Bei der initialen Detektion kommen die Zeilen 4 und 7 - 8 des Alg. 3.1 nicht zur Ausführung. Mit der Zeile 11 wird dem Gesamtsuchbereich des Interpretationszyklus SSP daher nur der initiale Suchraum SSP^0 zugewiesen. Nachdem in Zeile 13 die Bilder eingezogen worden sind, wird für die initiale Detektion dann erst wieder die Schleife in den Zeilen 20 - 27 ausgeführt.

Ist ein Objektmodell anhand seiner Merkmale in den zur Verfügung stehenden Sensordaten detektiert worden (Zeile 22) und ist eine Mindestqualität q_{min} erreicht, so wird eine Objektmodellinstanz obj_1 generiert und in die Liste der gefundenen Objektmodellinstanzen eingetragen (Zeile 24). Mit dem initialen Objektmodell wird anschließend versucht, auf den nicht von der ersten Objektmodellinstanz belegten Szenenmerkmalen, ein weiteres Objekt zu detektieren. Dieser Vorgang wird wiederholt, bis in dem Teilschritt der Detektion des Interpretationsprozeß keine weitere gültige Hypothese ($q < q_{min}$) mehr aufgestellt werden kann.

Re-Detektion von Objektmodellinstanzen

Ist die Liste der Objektmodellinstanzen $OBJ \neq \emptyset$, so wird in einem Interpretationszyklus versucht, die einzelnen Objektmodellinstanzen wieder zu finden. Zunächst wird dabei die Detektion für die Instanz mit dem kleinsten Index und somit der "ältesten" Instanz durchgeführt.[3]

Vor der eigentlichen Detektion muß man die wiederzufindende Objektmodellinstanz "altern" lassen. Hierzu werden in der Historie $HIST$ der Instanz und in den Historien $HIST_i$ seiner Objektmodellteile die entsprechenden Transformationsmatrizen übernommen, die im letzten Interpretationszyklus bestimmt worden sind.[4] Die Suchräume für die Detektion der Objektmodellinstanz werden hier nun direkt von den Objektmodellteilen anhand einer Vorhersage aus den Historien bestimmt.

Mit den, nach der Detektion der ersten Objektmodellinstanz verbleibenden Szenenmerkmalen wird dann jeweils die Detektion für die nächste Instanz aus der Liste vorgenommen, vgl. auch Zeilen 14 - 19. Anschließend wird, wie bei der initialen Detektion, versucht weitere Objektmodellinstanzen zu finden. Hierbei werden die Suchräume wieder entsprechend den initialen Suchräumen des Szenenmodells verwendet (Zeilen 21 - 26).

Verwerfen von Objektmodellinstanzen

Kann für eine bestehende Objektmodellinstanz im aktuellen Interpretationszyklus keine gültige Hypothese aufgestellt werden, so wird zunächst angenommen, daß die Merkmale der Instanz im aktuellen Satz von Sensordaten zeitweise nicht zu detektieren waren. Aufgrund der Vorhersage der Positionen wird die Position der Objektmodellteile und somit die komplette Lage der Objektmodellinstanz geschätzt. Die Güte / Qualität einer geschätzten Objektmodellinstanz ist geringer, als die einer anhand von Szenenmerkmalen detektierten Modellinstanz. Kann eine Instanz in einem weiteren Schritt nicht detektiert werden, so sinkt ihre Qualität weiter. Entsprechend eines anwendungsabhängigen Schwellenwerts werden Instanzen mit zu schlechter Qualität verworfen, vgl. Zeilen 16 - 18 im Alg. 3.1.[5]

[3] Es wird hiermit eine Bevorzugung der jeweils "ältesten" Objektmodellinstanz bei der Re-Detektion in Kauf genommen. Bei der "ältesten" Objektmodellinstanz handelt es sich um die Instanz, bei der der erste Eintrag in der Historie $HIST$ den kleinsten Zeitstempel hat.

[4] Für die Objektmodellteile werden zusätzlich noch die Szenenmerkmale $\mathbf{s}$ in der Historie vermerkt, die im letzten Interpretationszyklus detektiert und zugeordnet wurden. Ebenso werden die Zeitstempel in T für die Objektmodellinstanz fortgeschrieben.

[5] Die Bestimmung der Qualitäten von Hypothesen ist in Kap. 3.7 erläutert.

Verfolgung von Objektmodellinstanzen

Anwendungsabhängig kann es erforderlich sein, daß einzelne Objektmodellinstanzen einzeln verfolgt werden sollen. Hiermit wird zum einen, im Hinblick auf die Verwaltung der Objektmodellinstanzen, erreicht, daß ein einzelner Interpretationszyklus kürzer wird, denn es wird nicht versucht, weitere Instanzen oder das initiale Objektmodell zu finden. Zum anderen kann der Sichtbereich des Systems / der Kamera(s) gezielt verändert werden. Bei der Verfolgung wird gezielt auf eine weitere Kamera, die in dem Szenenmodell bekannt ist, umgeschaltet. Weiter ist bei der Verwendung einer aktiven Kamera, z.B. einer Schwenk- / Neigekamera, die Möglichkeit gegeben, hinter einem Objekt herzuschwenken.

Beim Start des Verfolgungsmodus muß aus der Liste der Objektmodellinstanzen eine Instanz ausgewählt werden. Alle anderen Instanzen werden verworfen und bei der Re-Detektion der zu verfolgenden Objektmodellinstanz werden auch keine weiteren initialen Objektmodelle gesucht. Die Auswahl der zu verfolgenden Objektmodellinstanz muß anwendungsabhängig über eine geeignete Schnittstelle erfolgen.

Wird der Verfolgungsmodus gleich zu Beginn des Interpretationszyklus aktiviert, so ist die initiale Detektion auf das Finden einer Objektmodellinstanz beschränkt. Wird im Verfolgungsmodus die einzelne Instanz verworfen, so wird im anschließenden Interpretationszyklus wieder im initialen Modellsuchraum SSP^0 des Szenenmodells die Detektion aufgenommen. Das bedeutet bei der Verwendung einer Schwenk- / Neigekamera eine entsprechende Positionierung auf eine Ausgangsstellung.

Im Alg. 3.1 werden bei der Verfolgung von Objektmodellinstanzen in den Zeilen 3 - 5 alle die Objektmodellinstanzen verworfen, die nicht verfolgt werden sollen. Dies ist notwendig, wenn zunächst im gesamten Observierungsraum Objekte detektiert wurden und dann eine Instanz ausgewählt wurde, die verfolgt werden soll. Somit werden die Zeilen 7 - 8 nur für die zu verfolgende Objektmodellinstanz ausgeführt und der Gesamtsuchraum des Interpretationszyklus $SSP = SSP_f$, der Suchraum der zu verfolgenden Objektmodellinstanz, gesetzt. SSP wird in Zeile 11 auch nicht mehr durch den initialen Suchraum ergänzt.

Die Schritte zur Detektion in der Zeile 25 werden nur für die zu verfolgende Objektmodellinstanz obj_f aufgerufen. Ist die Qualität / Güte der Instanz zu gering, wird in Zeile 17 die Liste der Objektmodellinstanzen $OBJ \Leftarrow \emptyset$ gesetzt.

3.1.3 Bildgenerierung

Im Teilschritt der Bildgenerierung werden die für den Interpretationsprozeß benötigten Bilder von den in dem Szenenmodell zur Verfügung stehenden Kameras eingezogen und die Bildvorverarbeitung durchgeführt. Der Ablauf der Bildgenerierung in STABIL^{++} läßt sich mit dem Alg. 3.2 darstellen.

Dort ist zu erkennen, daß nicht generell von allen Kameras aus CAM Bilder eingezogen werden, sondern nur entsprechend dem im Interpretationsprozeß bestimmten Suchraum SSP; vgl. auch Alg. 3.1. Die entsprechende Abfrage auf Sichtbarkeit des Suchraumes in der Zeile 12 im Alg. 3.2 wird im Kap. 3.3.4 näher erläutert. Anschließend wird von allen ausgewählten Kameras ein Bild eingezogen (Zeilen 16 - 19). Auf allen eingezogenen Bildern wird dann in den Zeilen 20 - 22 eine Bildvorverarbeitung ausgeführt.

Desweiteren können, falls Objektmodellinstanzen verfolgt und aktive Kameras verwendet werden, die Kameras zuvor ausgerichtet und der Blickwinkel ausgewählt werden. Im Alg. 3.2 findet dies in den Schritten 3 - 10 statt.[6]

[6] Für die Ausrichtung und Blickwinkelbestimmung von Kuppelkameras (Schwenk-/Neigekameras mit Motor-

Eingabe: CAM *// Liste der zur Verfügung stehenden Kameras*
Eingabe: SSP *// für den Interpretationszyklus bestimmter Suchraum*
Eingabe: $mode$ *// Verarbeitungsmodus z.B.: 'Verfolge'*
Berechne: $IMG = \{img_1, \ldots, img_n\} \Leftarrow$ Generierung der Bilder.
1: $IMG \Leftarrow \emptyset$ *// Liste von Bildern*
2: $CAM_{tmp} \Leftarrow \emptyset$ *// temporäre Liste von Kameras*
3: **falls** $mode =$ 'Verfolge' **dann**
4: **für alle** $cam_i \in CAM$ **wiederhole**
5: **falls** $type_i =$ 'Active-Camera' **dann**
6: Ausrichten von cam_i *// anhand von SSP,* ***s. Kap. 3.3.2***
7: Wahl des Blickwinkels von cam_i *// anhand von SSP,* ***s. Kap. 3.3.3***
8: •
9: •
10: •
11: **für alle** $cam_j \in CAM$ **wiederhole**
12: **falls** SSP mit cam_j sichtbar **dann**
13: $CAM_{tmp} \Leftarrow CAM_{tmp} \cup \{cam_j\}$ *//* ***s. Kap. 3.3.4***
14: •
15: •
16: **für alle** $cam_k \in CAM_{tmp}$ **wiederhole**
17: $img_{neu} \Leftarrow$ Ziehe Bild von cam_k ein *//* ***s. Kap. 3.4.1***
18: $IMG \Leftarrow IMG \cup img_{neu}$
19: •
20: **für alle** $img_l \in IMG$ **wiederhole**
21: $img_l \Leftarrow$ Bildvorverarbeitung für img_l *//* ***s. Kap. 3.4.2***
22: •

Algorithmus 3.2: *Bildgenerierung im Interpretationsprozeß.*

3.1.4 Detektion

Die Detektion einer Objektmodellinstanz obj, auf der Grundlage der entsprechend des Suchraumes SSP generierten Bildern $IMG = \{img_1, \ldots, img_n\}$, erfolgt entsprechend dem Alg. 3.3. In den Schritten 5 - 10 werden alle primären Merkmale der Objektmodellteile von obj rekursiv entsprechend der hierarchischen, inneren Objektmodellstruktur zusammengefaßt, so daß in den Schritten 11 - 15 auf allen Kamerabildern die entsprechenden 2D Bildmerkmale extrahiert werden können. Hierzu werden die den primären Merkmalen zugeordneten Bildverarbeitungsoperatoren $IP = \{ip_1(.), \ldots, ip_n(.)\}$ für jedes Bild ausgeführt.

Anschließend werden im Schritt 16 aus den 2D Bildmerkmalen $\mathbf{i}_j \in \mathbf{I}$ 3D Szenenmerkmale $\mathbf{s}_k \in \mathbf{S}$ bestimmt. Dieser 2D / 3D Übergang ist im Kap. 3.5 beschrieben. Es werden dann die 3D Szenenmerkmale den 3D Bildmerkmalen der Objektmodellinstanz zugeordnet. Jede mögliche, komplette Zuordnung ergibt im Detektionsschritt (Zeile 17) eine Hypothese für die innere Objektmodellstruktur und für die Translationen aus der geometrischen Struktur. Die genaue Definition der Hypothese, sowie der Prozeß der Generierung der Hypothesen ist im Kap. 3.6 beschrieben. Für alle so generierten Hypothesen wird in Zeile 19 eine Qualität bestimmt. Hierbei werden dann ebenfalls noch die Rotationen der geometrischen Objektmodellstruktur ermittelt. Desweiteren wird zur Bestimmung der Qualität die äußere Objektmodellstruktur herangezogen,

zoom-Objektiven) sind diese Schritte in den Kap. 3.3.2 und 3.3.3 näher erläutert.

Eingabe: obj // *Objektmodellinstanz*
Eingabe: $IMG = \{img_1, \ldots, img_n\}$ // *aufgrund des Suchraumes ausgewählte Bilder*
Eingabe: q_{min} // *minimale Qualität einer Objektmodellinstanz*
Berechne: $obj \Leftarrow$ Detektion der Objektmodellinstanz in den Bildern.
Berechne: $IMG \Leftarrow$ Ausblenden der für die Detektion verwendeten Bildmerkmale

```
 1: F ⇐ ∅    // Liste von primären Merkmalen
 2: I ⇐ ∅    // Liste von 2D Bildmerkmalen
 3: S ⇐ ∅    // Liste von 3D Szenenmerkmalen
 4: H ⇐ ∅    // Liste von Hypothesen
 5: omp_x ⇐ omp_0.1
 6: F ⇐ F ∪ {f_x}    // primäres Merkmal von omp_x
 7: für alle omp_i ∈ OMP_x wiederhole
 8:    omp_x ⇐ omp_i    // Parameter der Rekursion
 9:    rekursiv Zeilen 6 - 10
10: •
11: für alle img_j ∈ IMG wiederhole
12:    für alle f_k ∈ F wiederhole
13:       I ⇐ I ∪ IP_k(img_j)    // Extraktion der Bildmerkmale, s. Kap. 3.4.3
14:    •
15: •
16: S ⇐ Bestimme 3D Szenenmerkmale aus I    // s. Kap. 3.5
17: H ⇐ Generiere Hypothesen für obj aus S    // s. Kap. 3.6
18: für alle h_l ∈ H wiederhole
19:    q_l ⇐ Bestimme Qualität von h_l    // Hypothesenbewertung, s. Kap. 3.7
20: •
21: Sortiere H nach absteigender Qualität
22: falls q_1 ≥ q_min dann
23:    Akzeptiere h_1 für obj    // s. Kap. 3.8
24:    omp_x = omp_0.1
25:    für alle img_m ∈ IMG wiederhole
26:       img_m ⇐ img_m reduziert um Region der Projektion von f_x
27:    •
28:    für alle obj_n ∈ OMP_x wiederhole
29:       omp_x = omp_n    // Parameter der Rekursion
30:       rekursiv Zeilen 25 - 31
31:    •
32: •
```

Algorithmus 3.3: *Detektion im Interpretationsprozeß.*

indem die Lage der Volumina der einzelnen Objektmodellteile zueinander verglichen werden.

In Zeile 21 werden die Hypothesen nach absteigender Qualität sortiert. Falls die Qualität der besten Hypothese h_1 mindestens der minimal zu erreichenden Qualität q_{min} entspricht, so wird diese Hypothese in Zeile 23 ausgewählt und als Interpretation für die Szene für obj akzeptiert. Die Auswahl einer Hypothese für eine Objektmodellinstanz beinhaltet auch eine Adaption des Modells an die gemessenen 3D Szenenmerkmale.

Da entsprechend dem Alg. 3.1 die Detektion für die verschiedenen Objektmodellinstanzen oder das initiale Objektmodell mehrmals durchgeführt wird, ist sicherzustellen, daß die einzel-

Eingabe: obj *// Objektmodellinstanz*
Berechne: Speicherung der geometrischen Objektmodellstrukturen in einer Datei
1: Schreibe ${}^{obj}\mathbf{T}_{omp_{0.1}}$ *// Position von obj*
2: $omp_x = omp_{0.1}$
3: **für alle** $omp_i \in OMP_x$ **wiederhole**
4: $omp_y \Leftarrow omp_i$
5: Schreibe ${}^{omp_x}\mathbf{T}_{omp_i}$ *// Lage von omp_i im lokalen KS von omp_x*
6: $omp_x = omp_i$ *// Parameter der Rekursion*
7: rekursiv Zeilen 3 - 8
8: •

Algorithmus 3.4: *Aktion im Interpretationsprozeß, z.B. Speichern der geometrischen Objektmodellstruktur.*

nen Szenenmerkmale und somit die zugehörigen Bildmerkmale nur einmal verwendet werden. In den Schritten 24 - 31 werden daher rekursiv für alle Objektmodellteile der ausgewählten Hypothese die Projektionen der den primären Merkmalen zugehörigen 3D Modellmerkmale aus den Bildern ausgeblendet. Die notwendige 3D / 2D Projektion kann entsprechend der für das Bild bekannten Parameter des Kameramodells durchgeführt werden. Das Ausblenden der Modellmerkmale umfaßt nicht nur die einzelnen Kanäle des Bildes $CAN = \{can_1, \ldots, can_n\}$, sondern auch die Ergebnisregionen $\{REG^{attr_1}, \ldots, REG^{attr_n}\}$ der Klassifikatoren $CLC = \{clc^{attr_1}, \ldots, clc^{attr_n}\}$ der verwendeten Kameras.

Entsprechend der formalen Darstellung in den Alg. 3.1 und 3.3 werden somit in einem Interpretationszyklus für jede einzelne Objektmodellinstanz die Extraktion der Bildmerkmale wiederholt durchgeführt. Die mehrfache Ausführung von gleichen Bildverarbeitungsoperatoren $ip(.) \in IP$ auf immer wieder den gleichen Bildern ist jedoch nicht sinnvoll. Daher ist es empfehlenswert, entgegen der gewählten formalen Darstellung, die primären Merkmale der Objektmodellteile aller Objektmodellinstanzen und, falls erforderlich, des initialen Objektmodells zusammenzufassen und zu gruppieren. Diese Gruppierung erfolgt entsprechend Modellmerkmalen mit gleichen Attributen und somit gleichen Bildverarbeitungsoperatoren IP. Es kann daher vor der ersten Detektion im Interpretationsprozeß (vor Zeile 14 im Alg. 3.1) die Extraktion der Bildmerkmale vorgenommen werden. Es muß dann jedoch eine Verwaltung der schon verwendeten Bildmerkmale eingeführt werden.

3.1.5 Aktion

Mit allen dem Szenenmodell bekannten Akteuren $act_l \in ACT$ werden am Ende des Interpretationszyklus die Ergebnisse der Detektion verwendet. Dies wird als die Aktion im Interpretationszyklus bezeichnet. Es können hier die verschiedensten Akteure Verwendung finden. Die Art der auszuführenden Aktionen und somit der Akteure sind anwendungsabhängig, daher sei hier auf die Anwendungen in Kap. 4 verwiesen.

Im Alg. 3.4 ist beispielhaft ein einfacher Akteur dargestellt, der die Transformationen zwischen den lokalen Koordinatensystemen der einzelnen Objektmodellteile in eine Datei schreibt. Hierzu wird zunächst die Matrix der Transformation zwischen dem Koordinatensystem des Objektmodells obj und dem lokalen Koordinatensystem des ersten Objektmodellteils $omp_{0.1}$ ausgegeben (Zeile 1). Daran schließt sich in den Schritten 2 - 8 die Ausgabe der Matrizen der Transformation zwischen jeweils zwei aufeinander folgenden Objektmodellteilen an. Hierzu wird die hierarchische, innere Objektmodellstruktur von obj rekursiv durchlaufen.

3.2 3D Suchräume / Positionsvorhersage

3.2.1 Eigenschaften der Suchräume

Es werden in STABIL^{++} verschiedene 3D (Such-)Bereiche verwendet, mit denen die Detektion der Objekte auf bestimmte Bereiche der Szene begrenzt wird. Dies ist zum einen der 3D Observierungsraum SSP_s und die initialen 3D Modellsuchräume SSP^0 des definierten Szenenmodells, zum anderen der 3D Sichtbereich jeder Kamera. Desweiteren ist für jedes einzelne Objektmodellteil, dem ein Merkmal zugewiesen ist, ein 3D Suchraum bestimmbar. Der Suchraum eines Objektmodells setzt sich daher aus den Suchräumen seiner Objektmodellteile zusammen.

Der Observierungsraum SSP_s umfaßt einen 3D Raum, der im Weltkoordinatensystem wcs definiert ist, in dem das System Objekte erwartet. Diese Information wird im Interpretationsprozeß genutzt, um Fehlsegmentierungen von Bildmerkmalen auszuschließen. Hierzu wird als Heuristik verwendet, daß die 3D Szenenmerkmale der Objekte innerhalb des Observierungsraumes liegen und sichtbar sein müssen. Ist der Observierungsraum durch Weltregionen wr bestimmt, die die begrenzenden 3D Flächen beschreiben, so darf keine der Weltregionen $wr \in SSP_s$ zwischen der Kamera und dem vermeintlichen 3D Punkt des Szenenmerkmals liegen.

Der Sichtbereich einer Kamera ist durch die Pyramide bestimmt, die sich durch die vier Sichtstrahlen entsprechend der Eckpunkte des Bildes ergeben. Durch das verwendete Kameramodell sind diese 3D Strahlen eindeutig bestimmt. Aufgrund der bekannten Lage der Kamera im Weltkoordinatensystem ist die Sichtpyramide durch den Observierungsraum des Szenenmodells begrenzt. Der Sichtbereich der Kamera ist bei starren Kameras fix. Bei der Verwendung von aktiven, z.B. Schwenk- / Neigekameras kann der Sichtbereich der Kamera jedoch im Interpretationsprozeß zur Verfolgung von einzelnen Objektmodellinstanzen angepaßt werden.

Bei dem Suchraum des Objektmodells handelt es sich um eine Vereinigung der 3D Räume, in denen die Szenenmerkmale erwartet werden, die bei der Generierung der Hypothesen den Modellmerkmalen zugeordnet werden. Es muß hierbei zwischen dem Suchraum für die Detektion eines initialen Objektmodells und dem Suchraum einer Objektmodellinstanz unterschieden werden.

Die initialen Modellsuchräume des Szenenmodells sind jeweils durch die Angabe einer Position und eines Radius bestimmt. Mit der angegebenen Position wird die Position des Koordinatensystems des Objektmodells bestimmt. Somit wird das Objektmodell an eine (oder mehrere) bestimmte Stelle(n) im Raum gestellt und ist mit einer vorläufigen Objektmodellinstanz vergleichbar. Damit wird erreicht, daß sich die Re-Detektion von Objektmodellinstanzen nicht von der Detektion des initialen Objektmodells unterscheidet, vgl. Zeilen 15 und Zeile 22 in Alg. 3.1.

Um die lokalen Koordinatensysteme der Objektmodellteile, denen ein Merkmal zugewiesen ist, wird ein kugelförmiger oder zylindrischer Suchraum gebildet. Die Lage der Suchräume ist durch die Lage des Objektmodellteils und somit durch die vorgegebene geometrische Struktur des Objektmodells bestimmt. Mit dem anzugebenden Radius wird die Größe der Einzelsuchräume bestimmt. Der Radius muß anwendungsabhängig gewählt werden. Kennt man die exakte Ausgangsposition eines zu detektierenden Objektes, so kann der Suchraum exakt vorpositioniert und die Suchräume sehr klein gewählt werden. Hiermit wird der Vorgang der Generierung der Hypothesen beschleunigt.

Soll jedoch in einem größeren Bereich observiert werden, so kann man zum einen das initi-

ale Objektmodell an die Stellen positionieren, an denen Objekte auftauchen können.[1] Der Radius muß dann entsprechend einer anzunehmenden Ungenauigkeit gewählt werden. Zum anderen kann man nur eine Position des initialen Objektmodells in der Mitte des Observierungsraumes setzen und einen Radius wählen, der dann den ganzen Observierungsraum umfaßt. Es sei hier noch erwähnt, daß der Suchraum jedoch immer durch den Observierungsraum und die Sichtbereiche der Kameras beschränkt ist.

Ist ein Objektmodell einmal detektiert worden, so sind die Suchräume der einzelnen Objektmodellteile durch die Objektmodellinstanzen selbst festgelegt. Die Position der Suchräume wird durch eine Positionsvorhersage[2] bestimmt. Somit erhält man jeweils für den kommenden Interpretationszyklus Suchräume, in dem die Objektmodellinstanz zu erwarten ist. Die Größe der Suchräume hängt hierbei von der Güte / Qualität der letzten Detektion ab.

Die Suchräume der Objektmodellinstanzen werden in drei Schritten des Interpretationsprozesses genutzt: Erstens wird in dem Teilschritt der Bildgenerierung anhand der Suchräume bei der Verwendung von aktiven Kameras die Kameraausrichtung und Blickwinkeleinstellung vorgenommen. Weiterhin werden nur von den Kameras Bilder eingezogen, in deren Sichtbereichen die Suchräume liegen.

Zweitens werden durch eine 3D/2D Projektion des Suchraumes die eingezogenen Bilder auf Regionen eingeschränkt, in denen Bildmerkmale zu erwarten sind. Drittens werden die Suchräume zur Beschleunigung der Generierung der Hypothesen und zur Bewertung der Hypothesen verwendet. Die Positionsvorhersagen für die Objektmodellteile, aus denen die Position der Suchräume bestimmt wird, werden zusätzlich beim Generieren der Hypothesen auch zur Schätzung von nicht zu detektierenden Szenenmerkmalen verwendet. D.h., es wird als 3D Position des Ursprungs des lokalen Koordinatensystems des entsprechenden Objektmodellteils die Vorhersage als Schätzungen herangezogen. Man spricht hier von einer Positionsvorhersage für das Objektmodellteil.[3]

3.2.2 Bestimmung der Suchräume

Für die Vorhersage der Position des Ursprungs des lokalen Koordinatensystemes steht einem Objektmodellteil omp_ν ein Vorhersagefilter $pred_\nu$ zur Verfügung. Mit diesem Filter soll die Position eines 3D Suchraums bestimmt werden, der einen 3D Raum beschreibt, in dem mit großer Wahrscheinlichkeit im nächsten Interpretationszyklus der Ursprung des lokalen Koordinatensystems und die Merkmale des Objektmodellteiles zu liegen kommen. Daher müssen neben einem Parameter der 3D Position auch Parameter der Form und der 3D Orientierung für einen Suchraum bestimmt werden.

Zur Bestimmung dieser Parameter können Unsicherheiten berücksichtigt werden, die die Suchräume beeinflussen. Durch eine Modellierung der Unsicherheiten durch stochastische Verteilungsfunktionen ergeben sich für die Parameter des Suchraumes ein Zufallsvariablen-Problem. So läßt sich entsprechend der Ausführungen in [Lan98] zur Modellierung und Propagierung von Unsicherheiten mit Hilfe des *Mahalanobis-Abstandes* ein Hyperellipsoid definieren, der den Suchraum bestimmt. Dort wird ein 2D Suchraum im Videobild ermittelt, in dem dann die projizierten Modellmerkmale liegen müssen; nur in diesem wird dort bei der Interpretation eine 2D / 2D Zuordnung von Merkmalen vorgenommen. Aufgrund der bei der

[1] Das initiale Modell wird im Interpretationsprozeß nacheinander an mehrere Stellen positioniert. Eine detektierte Objektmodellinstanz wird jedoch pro Interpretationszyklus nur einmal verwendet, vgl. Alg. 3.1.

[2] Wird das komplette Objektmodell betrachtet, so kann auch von einer Bewegungsvorhersage gesprochen werden.

[3] Die Anwendung der 3D Suchräume im Interpretationsprozeß ist in Kap. 3.3.4 zur Sichtbarkeit der Suchräume, in Kap. 3.4.2 zur Bildsegmentierung und in Kap. 3.6 und 3.7 zur Generierung und Bewertung von Hypothesen beschrieben.

Zuordnung zugrunde gelegten Starthypothese der gesuchten Lage ergibt sich die erste zu modellierende Unsicherheit. Im weiteren werden in dem dort verwendeten Ansatz die Unsicherheit in der Modellinformation, in der Lage der Kamera, in den internen Kameraparametern und in den extrahierten Bildmerkmalen berücksichtigt.

In STABIL^{++} wird jedoch ein 3D Suchraum ermittelt, denn die Interpretation wird im 3D Raum durchgeführt, d.h. es wird eine 3D / 3D Zuordnung von Merkmalen durchgeführt. Es entfallen daher zunächst die Unsicherheiten, die die Kamera betreffen. Es wird jedoch auch hier in einem weiteren Schritt der 3D Suchraum in die Videobilder projiziert, um dadurch eine Einschränkung des Bildbereiches vornehmen zu können. Dabei wird jedoch keine das Kameramodell betreffende Unsicherheit berücksichtigt, denn entsprechend den Ausführungen in [Lan98] lassen sich diese Unsicherheiten bei einer "sorgfältigen Kalibrierung sowohl der inneren Parameter als auch der Lage der Kamera vernachlässigen". Dies gilt für STABIL^{++} insbesondere, da die Kamera nicht auf einem autonomen mobilen System montiert ist und im allgemeinen in der Lage zum Weltkoordinatensystem fix ist. Bei dem Einsatz von Schwenk- / Neigekameras, die in einer kontinuierlichen Fahrt ein Objekt verfolgen, ist dies jedoch zu überdenken. Hierbei muß dann jeweils beim Einzug eines Bildes immer eine exakte Kalibrierung vorliegen.

Die Unsicherheiten für die Bestimmung der 3D Suchräume ergeben sich daher in STABIL^{++} in erster Linie aus dem Objektmodell selbst und somit durch Bewegung einer Objektmodellinstanz und Veränderungen in seiner geometrischen Objektmodellstruktur über die Zeit, da hier Bildfolgen ausgewertet werden. Entsprechend einer pragmatischen Betrachtung der Anforderungen an die 3D Suchräume, lassen sich die drei Parameter des Suchraumes für ein Objektmodellteil omp_ν anhand der folgenden Restriktionen und den sich dadurch ergebenden Abhängigkeiten bestimmen:

Geometrische Objektmodellstruktur: Mit der Historie $HIST_\nu$ der geometrischen Objektmodellstruktur kann der Bewegungsablauf des Objektmodellteils in der Vergangenheit rekonstruiert werden. Hieraus können auch eine wahrscheinliche Bewegungsrichtung und Bewegungsgeschwindigkeit vorausgesagt werden, so daß eine Position und Orientierung des Suchraumes bestimmt werden kann.

Äußere Objektmodellstruktur: Für die äußere Objektmodellstruktur muß berücksichtigt werden, daß die Merkmale des Objektmodellteiles innerhalb des Suchraumes liegen. Sofern die Form des Objektmodellteiles ein Merkmal ist, so muß hier auch der Volumenkörper berücksichtigt werden.

Zeit: Eine zeitliche Abhängigkeit ergibt sich schon durch die Vorhersage der Bewegungsrichtung / -geschwindigkeit über die Historie der geometrischen Struktur und damit der Position. Desweiteren muß berücksichtigt werden, daß eine Vorhersage unsicherer wird, je länger der Vorhersagezeitraum ist; somit muß der Suchraum entsprechend größer sein.

Detektionsgüte / Unsicherheiten: Der Suchraum muß ebenfalls ein größeres Ausmaß annehmen, wenn die Daten, auf die sich die Vorhersage stützt, unsicherer sind. Hier bedeutet dies, daß aufgrund einer schlechteren Qualität / Güte der Detektion des Objektmodellteils im letzten Interpretationszyklus der Suchraum größer werden muß.

Beweglichkeit: Restriktionen bezüglich der Beweglichkeit eines Objektmodellteiles können zur Bestimmung der Orientierung, der Form und der Größe des Suchraumes verwendet werden. Die Beweglichkeit ist zum einen durch die Bewegungsfreiheitsgrade des Objektmodellteils bestimmt, zum anderen durch die Beweglichkeit seiner Vorgängerobjektmodellteile. Ist die Bewegungsrichtung eingeschränkt, so sollte sich der Suchraum nur in die mögliche Richtung der Bewegung ausdehnen. Kann ein Objektmodellteil aufgrund geringer Freiheitsgrade seine Lage nur mit kleiner Geschwindigkeit verändern, so kann der Suchraum entsprechend kleiner gewählt werden.

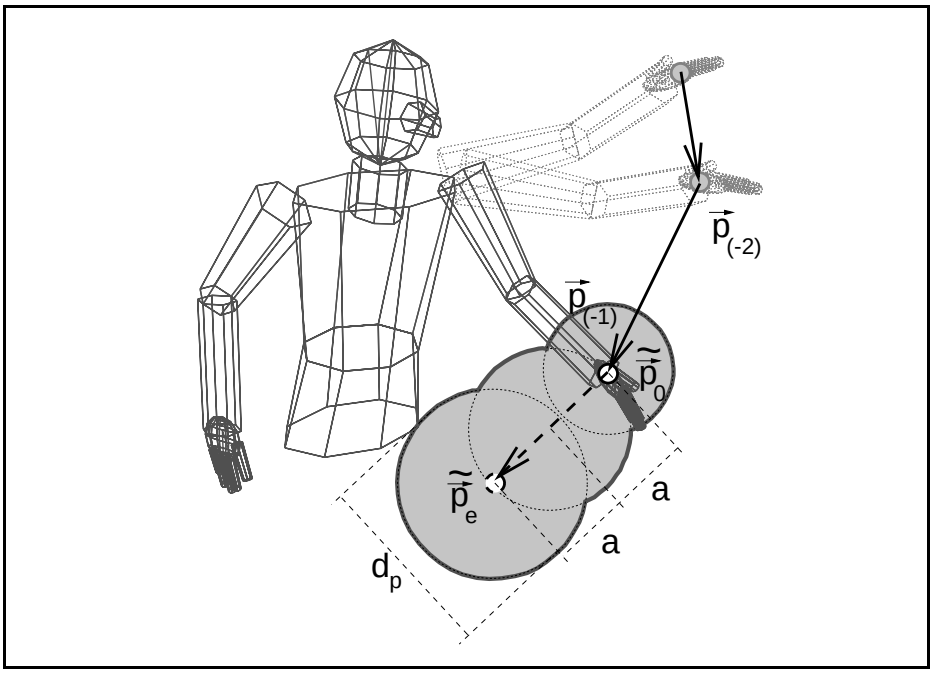


Abbildung 3.2: *Suchbereich für primäres Merkmal des Objektmodellteiles der linken Hand.*

Aufgrund der angegebenen Bedingungen und Abhängigkeiten muß sich für ein Objektmodellteil ein Suchraum ergeben, dessen Position zunächst ausgehend von der Position des Ursprungs des lokalen Koordinatensystems des Objektmodellteils im letzten Interpretationszyklus bestimmt ist. Der Suchraum muß sich dann in Richtung der bisher durch das Objektmodellteil ausgeführten Bewegung ausdehnen, wobei die Form entsprechend der Unsicherheiten in der Positionsvorhersage zu wählen ist. Die Modellierung von beliebigen 3D Volumenkörpern gestaltet sich schwierig, so daß im Hinblick auf eine algorithmisch einfach handhabbare Lösung in STABIL^{++} nur kugelförmige und zylindrische 3D Einzelsuchräume verwendet werden. Dies begründet sich zum einen durch die einfache 3D / 2D Projektion dieser geometrischen Primitiven. Zum anderen werden in STABIL^{++} Punktmerkmale verwendet, so daß ein kugelförmiger Suchraum die primären Merkmale gleichmäßig umschließt. Zur Approximierung eines Suchraums, der die Bewegungsrichtung und -geschwindigkeit berücksichtigt, können dann mehrere kugelförmige Suchräume in Kombination verwendet werden, vgl. Abb. 3.2.

Anhand des Vorhersagefilters $pred_\nu$ wird zunächst für ein Objektmodellteil omp_ν eine Position vorhergesagt. Für die Verwendung dieser Positionen zur Bestimmung der Suchräume ist es wichtig, ein Maß zu ermitteln, das angibt, mit welcher Wahrscheinlichkeit der Punkt im nächsten Interpretationszyklus dort zu liegen kommt. Ist die Wahrscheinlichkeit groß, so müßte auch der Suchraum entsprechend groß sein, damit bei der Bestimmung der Szenenmerkmale auch gesichert in diesem Bereich gesucht wird.

Dies widerspricht jedoch der Verwendung der Suchräume zur Beschleunigung der Interpretation. Hierbei muß zum einen erreicht werden, daß die ins Bild projizierten Bereiche möglichst klein sind, damit die Bildverarbeitungsoperatoren nur in einer begrenzten Region angewendet werden müssen. Zum anderen werden die Suchräume verwendet, um die mögliche Zuordnung von Szenen- zu Modellmerkmalen zu beurteilen und den hierbei für dieses Korrespondenzproblem aufgespannten Suchraum einzuschränken.[4] Durch kleine 3D Suchräume werden eindeuti-

[4]Die Verwendung des Begriffs “Suchraum” kann hier zu Verwirrung führen: Mit der Suchraumgröße des Kor-

Eingabe: obj // *Objektmodellinstanz*
Berechne: $SSP \Leftarrow$ Suchraumvorhersage für Objektmodellinstanz obj
1: $SSP \Leftarrow \emptyset$
2: $omp_x \Leftarrow omp_{0.1}$ // $omp_{0.1} \in OMP_i$
3: $SSP_x \Leftarrow \emptyset$ // *Suchraum des* omp_x
4: $\tilde{P} \Leftarrow$ Bestimme Vorhersage mit Vorhersagefilter $pred_x$ // $\tilde{P} \in \{\emptyset, \{\tilde{\vec{p}}\}\}$
5: **falls** $\tilde{P} = \emptyset$ **dann**
6: $\quad \tilde{P} \Leftarrow$ Vorhersage aus letzter geometrischer Modellstruktur
7: •
8: **für alle** $\mathbf{f}_j \in \mathbf{F}_x$ **wiederhole**
9: $\quad ssp \Leftarrow$ Bestimme Suchraum für primäres Merkmal $\mathbf{f}_j$ an Vorhersageposition $\tilde{\vec{p}}$
10: $\quad SSP_x \Leftarrow SSP_x \cup ssp$
11: •
12: **für alle** $\mathbf{f}'_k \in \mathbf{F}'_x$ **wiederhole**
13: $\quad ssp \Leftarrow$ Bestimme Suchraum für sekundäres Merkmal $\mathbf{f}'_k$
14: $\quad SSP_x \Leftarrow SSP_x \cup ssp$
15: •
16: $SSP \Leftarrow SSP \cup SSP_x$ // *Suchraum des Objektmodells*
17: **für alle** $omp_j \in OMP_x$ **wiederhole**
18: $\quad omp_x \Leftarrow omp_j$ // *Parameter der Rekursion*
19: $\quad$ rekursiv Zeilen 3 - 20
20: •

Algorithmus 3.5: *Positionsvorhersage und Bestimmung der 3D Suchräume einer Objektmodellinstanz.*

gere Aussagen ermöglicht. Daher sollte bei einer gesicherten Vorhersage, der eine hohe Wahrscheinlichkeit entspricht, der Suchraum entsprechend kleiner sein, als bei einer unsicheren. Ein minimaler 3D Suchraum muß erreicht werden, wenn eine maximale Qualität der Vorhersage vorliegt. Ist die Vorhersage sehr unsicher, dann muß der 3D Suchraum anwachsen, denn die wahre Position kann in einem größeren Radius um die Position liegen.

Es wird daher in STABIL^{++} von einer Qualität einer Vorhersage gesprochen, die ein Maß für die Sicherheit der Vorhersage ist. Mit dieser Qualität wird zum einen über die Zulässigkeit der Vorhersage entschieden und zum anderen dann, entsprechend der o.a. Überlegungen, die Größe der Suchraumkugeln bestimmt. Bei der Verwendung der Positionsvorhersagen als Schätzung für 3D Szenenmerkmale kann der Begriff der Qualität direkt verwendet werden und ist dann mit den Qualitäten der real gefundenen Szenenmerkmale vergleichbar.

Nach der Positionsvorhersage mit entsprechender Qualitätsbeurteilung werden die Suchräume selbst bestimmt, dies jedoch getrennt für die primären und sekundären Merkmale des Objektmodellteils. Es ergibt sich daher für die Bestimmung des Gesamtsuchraumes eines Objektmodells in STABIL^{++} ein Ablauf entsprechend dem Alg. 3.5.

Dort wird, entsprechend der hierarchischen, inneren Objektmodellstruktur, zunächst für das erste Objektmodellteil der Objektmodellinstanz ein Suchraum bestimmt und anschließend rekursiv für die anderen Objektmodellteile. In Zeile 4 wird versucht, für das Objektmodellteil omp_x eine Position mit einer entsprechenden Mindestqualität vorherzusagen. Kann keine Vor-

respondenzproblems wird ein Maß für seine Komplexität angegeben. Dies ist nicht mit den 3D Suchräumen SSP, die für die Lage der Objektmodellteile vorhergesagt werden, zu verwechseln.

hersage mit einer minimalen Qualität bestimmt werden, so ist $\tilde{P} = \emptyset$ und es wird in Zeile 6 die Vorhersage entsprechend der Position aus der letzten geometrischen Modellstruktur gesetzt. Die Qualität für diese "Vorhersage" wird mit der minimal zu akzeptierenden Qualität gesetzt, wodurch um die alte Position ein 3D Suchraum mit der maximalen Größe gelegt wird.

Mit diesen vorhergesagten Positionen und den dazugehörigen Wahrscheinlichkeiten / Qualitäten werden dann in den Zeilen 9 - 13 für alle primären Merkmale des Objektmodellteils die Suchräume bestimmt. In den Zeilen 14 - 17 werden für die sekundären Merkmale die Suchräume bestimmt. Im folgenden wird der Schritt der Positionsvorhersage erläutert, daran schließt sich die Beschreibung der eigentlichen Bestimmung der Suchräume an.

3.2.3 Positionsvorhersage

Die Positionsvorhersage des Ursprungs des lokalen Koordinatensystems eines Objektmodellteils omp_ν erfolgt durch den ihm bekannten Vorhersagefilter $pred_\nu$. Mit den Vorhersagefiltern der Objektmodellteile wird eine Positionsvorhersage für das zugehörige Objektmodell getroffen. Die vorhergesagten Positionen werden zur Bestimmung der Positionen von 3D Suchräumen und als Schätzung für die Positionen der Szenenmerkmale verwendet, falls kein Bildmerkmal extrahiert werden kann oder keine gültigen 3D Positionen aus diesen bestimmt werden können. Die Positionsvorhersage kann durch einfache Extrapolation aus den in der Historie $HIST_\nu$ vorhandenen Positionen für die vergangenen Interpretationszyklen erfolgen.

Ein über Extrapolation vorhergesagter Punkt $\tilde{\vec{p}}_e$ soll berücksichtigen, daß sich das Objektmodellteil mit gleicher Geschwindigkeit und Richtung weiter bewegt. Für die Extrapolation des Punktes $\tilde{\vec{p}}_e$ für ein Objektmodellteil omp_ν werden daher aus der Historie $HIST_\nu$ die letzten Positionen des Ursprungs des lokalen Koordinatensystems zu den Zeitpunkten $t_{(-1)}$ und $t_{(-2)}$ ermittelt. Man erhält die beiden Punkte:

$$\begin{aligned} \vec{p}_{(-1)} &= {}^{wcs}\mathbf{T}_{\nu,t_{(-1)}} \cdot [0,0,0]^T \\ \vec{p}_{(-2)} &= {}^{wcs}\mathbf{T}_{\nu,t_{(-2)}} \cdot [0,0,0]^T \end{aligned} \tag{3.1}$$

Es muß hierbei entsprechend der hierarchischen, inneren Objektmodellstruktur aus den in den Historien vermerkten Transformationsmatrizen rekursiv die Transformation zum Weltkoordinatensystem wcs ermittelt werden, vgl. Glg. 2.4.

Zur Bestimmung von $\tilde{\vec{p}}_e$ wird zunächst eine Bewegungsgeschwindigkeit vorausgesagt. Die Vorhersage des Geschwindigkeitsvektors $\tilde{\vec{v}}$ kann hierzu aus den beiden vorhergehenden Positionen entsprechend

$$\tilde{\vec{v}} = \frac{dx}{dt} = \frac{\vec{p}_{(-1)} - \vec{p}_{(-2)}}{t_{(-1)} - t_{(-2)}}$$

bestimmt werden. Mit dieser Geschwindigkeit ergibt sich für die Vorhersage:

$$\tilde{\vec{p}}_e = \vec{p}_{(-1)} + \tilde{\vec{v}} \cdot (t_0 - t_{(-1)}) \tag{3.2}$$

Die Qualität q_e der vorhergesagten Position $\tilde{\vec{p}}_e$ muß zum einen die Güte / Qualität der, der Vorhersage zugrunde gelegten, Punkte berücksichtigen. Aus der Historie $HIST$ des Objektmodellteiles werden daher aus den dort vermerkten Szenenmerkmalen $\mathbf{s}_{(-1)}$ und $\mathbf{s}_{(-2)}$ die Qualitäten q verwendet. Diese beiden Qualitätsmaße sollen hier, entsprechend des Alters, mit $q_{(-1)}$ und $q_{(-2)}$ bezeichnet werden. Zum anderen muß die Unsicherheit, die sich durch einen langen

Bildrate [Bilder / sek]	25	12,5	8	4	2	1	0,5
Taktlänge [sek]	0,04	0,08	0,125	0,25	0,5	1	2
Alterungsfaktor [%]	95,0	77,7	61,7	38,1	21,3	11,3	5,82

Tabelle 3.1: *Alterungsfaktoren der Qualitätsvorhersage in Abhängigkeit der Bildwiederholrate.*

Vorhersagezeitraum ergibt, ebenfalls mit eingehen. Daher werden die beiden Qualitätsmaße gemittelt und mit der sog. Alterungsfunktion $a(t)$ gewichtet:

$$q_e = a(t_0 - t_{(-1)}) \cdot \frac{q_{(-1)} + q_{(-2)}}{2}$$

Die Alterungsfunktion hat einen Wertebereich von $[0, 1]$, wobei der Wert mit größer werdendem t exponentiell abnimmt. Entsprechend den Anwendungserfahrungen ist die Alterungsfunktion so ausgelegt, daß sich bei einer Bildwiederholrate von 25 Bildern / Sekunde ein Faktor von ca. 95% ergibt; damit gilt für $a(t)$:

$$a(t) = 1 - e^{-\frac{0,12}{t}} \tag{3.3}$$

In Systemen zur Bildfolgenverarbeitung wird teilweise anstelle der expliziten Zeit t mit einer Numerierung der einzelnen Bilder gearbeitet. Um dann einen Alterungsfaktor entsprechend der Alterungsfunktion $a(t)$ ermitteln zu können, muß die mittlere Bildwiederholrate bekannt sein. Für verschiedene Bildwiederholraten ergeben sich die Faktoren für $a(t)$ entsprechend Tab. 3.1 für jeweils ein Δt von einer Taktlänge.[5]

Sind in der Historie weniger als zwei Einträge vorhanden, so kann $\tilde{\vec{p}}_e$ nicht ermittelt werden. Ferner wird $\tilde{\vec{p}}_e$ nicht als Vorhersage akzeptiert, wenn die Qualtität $q_e < q_{min}$, hierbei ist q_{min} eine anwendungsabhängig zu setzende minimale Qualität. In diesen beiden Fällen wird ein Punkt $\tilde{\vec{p}}_0$ ermittelt, der sich an der zuletzt sicher detektierten Position des Objektmodellteils orientiert. Hierzu muß $\tilde{\vec{p}}_0$ einen Punkt aus der Historie annehmen, dem zuletzt ein Szenenmerkmal zugewiesen worden ist. Daher werden die Einträge der Qualität der Szenenmerkmale $\mathbf{s}_{(-1)} \ldots \mathbf{s}_{(-n)}$ in der Historie $HIST_\nu$ des Objektmodellteils omp_ν soweit durchlaufen, bis ein Eintrag mit einer Qualität $q_{(-i)} \geq q_{min}$ gefunden wird.[6]

Entsprechend dem zugehörigen Zeitpunkt $t_{(-i)}$ wird

$$\tilde{\vec{p}}_0 = \vec{p}_{(-i)} \tag{3.4}$$

gesetzt. Um die unterschiedliche Sicherheit der Vorhersagen $\tilde{\vec{p}}_e$ und den Punkt $\tilde{\vec{p}}_0$ berücksichtigen zu können, muß für beide ein unterschiedliches Qualitätsmaß bestimmt werden. Entsprechend der geringeren Wahrscheinlichkeit, daß der Punkt die alte Position beibehält, ist die Qualität q_0 geringer als q_e zu wählen:

$$q_0 = a(t_{(-i)}) \cdot 0,95 \cdot q_{(-i)}$$

Daher ergibt sich, für die Qualität q_0 ein Gesamtalterungsfaktor von 90%, bei einer Bildwiederholrate von 25 Bildern / Sekunde, gegenüber 95% einer vergleichbaren Qualität q_e.

[5] Die Taktlänge ist die Zeit zwischen zwei aufeinanderfolgenden Bildern, z.B. $t_0 - t_{(-1)}$.

[6] $q_{(-i)}$ ist hierbei die Qualität des Szenenmerkmals $\mathbf{s}_{(-i)}$.

Die vorhergesagten Punkte werden nur für den weiteren Interpretationsprozeß verwendet, wenn die Qualität dieser Punkte mindestens einer für das Objektmodellteil festgelegten Mindestqualität entsprechen. Hierüber läßt sich die "Beweglichkeit" des Objektmodellteiles berücksichtigen. Die Bestimmung der Vorhersagen in der Zeile 4 des Alg. 3.5 kann entweder $\tilde{\vec{p}}_e$, $\tilde{\vec{p}}_0$ oder keine vorhergesagte Position enthalten. Falls für das Objektmodellteil keine gesicherte Vorhersage getroffen werden konnte, wird in Zeile 6 eine Position bestimmt, die sich aus der geometrischen Modellstruktur ergibt. Somit wird unter der Annahme, daß sich weder Translation, noch Rotation zwischen den Objektmodellteilen verändert hat, die zuletzt durch den Ursprung des lokalen Koordinatensystems eingenommene Position im Weltkoordinatensystem verwendet.

Der Vollständigkeit halber sei hier erwähnt, daß für ein Objektmodellteil, für das kein primäres Merkmal definiert ist und somit auch kein Suchraum benötigt wird, die Vorhersage der Position nicht durchgeführt wird.

3.2.4 Suchräume der primären Merkmale

Die Suchräume für die primären Merkmale müssen sich an der vorhergesagten Position $\tilde{\vec{p}}$ mit ihrer Position orientieren. Die Form und die Orientierung des Suchraumes muß entsprechend der Unsicherheit in der verwendeten Bewegungsvorhersage bestimmt werden. Es wird daher zwischen der Bestimmung der Suchräume für die primären Merkmale aus extrapolierten Punkten $\tilde{\vec{p}}_e$ und den den alten Positionen entsprechenden Punkten $\tilde{\vec{p}}_0$ unterschieden.

Generell setzen sich die Suchräume jedoch aus einer oder mehreren Suchraumkugeln zusammen, für die Mittelpunkte $\tilde{\vec{p}}_i$ zu bestimmen sind. Diese Mittelpunkte werden noch entsprechend des Verschiebungsvektor $\vec{t}$ des primären Merkmales korrigiert, vgl. Glg. 2.8. Mit $\vec{t}$ ist angegeben, wo sich das Punktmerkmal im lokalen Koordinatensystem des Objektmodellteils befindet. Der Ursprung des lokalen Koordinatensystems liegt z.B. bei der Modellierung des Menschen in den Gelenken, wobei jedoch bei der Verwendung von expliziten Markierungen, diese um $\vec{t}$ versetzt auf der Oberfläche des Körpers liegen. Dies gilt auch für aus den Volumenkörpern abgeleiteten Merkmale. So muß z.B. der Suchraum für das Merkmal einer "hautfarbenen" 3D Ellipse für das Objektmodellteil des Kopfes $omp_{3.3}$ um die Länge des Vektors $\vec{t}$ in Richtung der z-Achse $Z_{3.3}$ verschoben werden, vgl. Kap. 2.4.5.

Die Suchraumbestimmung wird zu Beginn des Interpretationsprozeßes durchgeführt, daher ist zu diesem Zeitpunkt die Lage des lokalen Koordinatensystems eines Objektmodellteiles omp_ν noch nicht bekannt. Es besteht die Möglichkeit, die Lage des Koordinatensystems entsprechend des letzten Eintrages aus der Historie $HIST_\nu$ zu verwenden. Die Erfahrung hat jedoch gezeigt, daß insbesondere bei der Verwendung von expliziten Markierungen als Merkmale, sich diese nicht mit einer Rotation des Objektmodellteils selbst verschieben, sondern ihren Bezug im lokalen Koordinatensystem des Vorgängerobjektmodellteils omp_μ haben. Z.B. werden bei der Modellierung des menschlichen Körpers die Markierungen an den Hüften für die Objektmodellteile $omp_{1.1}$ und $omp_{1.3}$ der Oberschenkel kaum durch die Rotation um die z''-Achse des Objektmodellteils des Oberschenkels beeinflußt. Es werden daher die vorhergesagten Punkte wie folgt korrigiert:

$$\tilde{\vec{p}'}_i = \tilde{\vec{p}}_i \cdot {}^{wcs}\mathbf{T}_{omp_\mu} \cdot \vec{t} \tag{3.5}$$

Suchraumkugeln aus $\tilde{\vec{p}}_e$

Wird eine Bewegungsvorhersage in Form der Extrapolation verwendet, so kann die Form des Gesamtsuchraums für das primäre Merkmal lediglich aus der Bewegungsrichtung abgeleitet werden. Der vorhergesagte Punkt $\tilde{\vec{p}}_e$ bestimmt den Punkt, an dem das Merkmal mit großer Wahrscheinlichkeit zu liegen kommt. Es soll jedoch noch berücksichtigt werden, daß die Bewegung sich auch verlangsamen kann und daß mit größer werdender Entfernung von der zuletzt detektierten Position für den das primäre Merkmal die Unsicherheit zunimmt. Entsprechend der Abb. 3.2 wird daher ein "keulenförmiger" Gesamtsuchraum angestrebt, der z.B. mit drei Suchraumkugeln approximiert werden kann.

Die Kugel mit dem größten Durchmesser liegt in dem vorhergesagten Punkt $\tilde{\vec{p}}_e$, die kleinste Kugel liegt in dem Punkt, der sich aus $\vec{p}_{(-1)}$ bestimmen läßt, vgl. Glg. 3.1. Entsprechend Glg. 3.2 lassen sich zwischen den beiden Punkten $\tilde{\vec{p}}_e$ und $\vec{p}_{(-1)}$ die Mittelpunkte weiterer Suchraumkugeln bestimmen.

Die Größe der Suchräume und somit der Durchmesser wird anhand der Qualität q des vorhergesagten Punktes bestimmt. Hier soll bei einer maximalen Qualität von $q_e = 1$ ein minimaler Durchmesser d_{min} verwendet werden, wird jedoch nur eine minimale Qualität q_{min} erreicht, so muß der Suchraum seine maximale Größe d_{max} erreichen. Damit bestimmt sich der Suchraumdurchmesser d_p für die Suchraumkugel an der Position $\tilde{\vec{p}}_e$ mit

$$d_p = d_{max} - (d_{max} - d_{min}) \cdot \frac{q_e - q_{min}}{1 - q_{min}} \tag{3.6}$$

Der minimale Suchraumdurchmesser d_{min} ergibt sich bei expliziten primären Merkmalen aus dem größten Durchmesser des Punktmerkmals. Bei impliziten Merkmalen ist dieser durch das Ausmaß des Volumenkörpers gegeben und muß so bemessen sein, daß der Körper komplett im Suchraum zu liegen kommen kann. Der maximale Durchmesser d_{max} wird anwendungsabhängig für das komplette Objektmodell gesetzt. Mit q_{min} muß daher bei Objektmodellteilen mit großer "Beweglichkeit" eine hohe Schwelle gesetzt werden, damit die maximale Suchraumgröße schon bei geringer Unsicherheit erreicht wird.

Für die weiteren Suchraumkugeln, mit denen der "keulenförmige" Gesamtsuchraum approximiert wird, müssen kleinere Durchmesser verwendet werden, vgl. Abb. 3.2. Es werden entsprechend des Abstandes zur Position $\tilde{\vec{p}}_e$ kleinere Durchmesser gewählt. Für die Suchraumkugel an der Position $\vec{p}_{(-1)}$ wird z.B. $0,75 \cdot d_p$ gewählt. Für die Suchraumkugeln, die zwischen den Positionen $\tilde{\vec{p}}_e$ und $\vec{p}_{(-1)}$ zu liegen kommen, ist somit ein Durchmesser zwischen d_p und $0,75 \cdot d_p$ zu wählen.

Suchraumkugeln aus $\tilde{\vec{p}}_0$

Konnten keine Positionen vorhergesagt werden, so wird der Punkt $\tilde{\vec{p}}_0$ als Vorhersage verwendet, vgl. Glg. 3.4. Es wird dann nur eine Suchraumkugel um diesen Punkt gelegt. Der Suchraumdurchmesser bestimmt sich aus der Qualität dieser Vorhersage q_0. Den Durchmesser erhält man entsprechend der Glg. 3.6, wenn man anstelle der Qualität q_e die Qualität q_0 einsetzt.

3.2.5 Suchräume der sekundären Merkmale

Die Verwendung von 3D Suchräumen für die sekundären Merkmale eines Objektmodellteiles beschränkt sich auf die Auswahl der Bildregionen für die Bildverarbeitungsoperatoren. Somit

Bildrate [Bilder / sek]	25	12,5	8	4	2	1	0,5
Taktlänge [sek]	0,04	0,08	0,125	0,25	0,5	1	2
Durchmesser [m]	1,10	1,54	2,00	2,73	3,28	3,62	3,80

Tabelle 3.2: *Suchraumdurchmesser d_s für sekundäre Merkmale bei $d_v = 1$m in Abhängigkeit der Bildwiederholrate.*

ergibt sich, daß nur für die sekundären Merkmale ein Suchraum zu bestimmen ist, für die auch Bildverarbeitungsoperationen definiert sind.

Für das in Kap. 2.4.6 vorgestellte Beispiel, der Überprüfung des Vorhandenseins von Objektmodellteilen anhand der Vordergrundregion im Bild, ist der Suchraum durch den Volumenkörper des Objektmodellteils bestimmt. Für einen Volumenkörper der äußeren Modellstruktur vol_{ell} wird der Suchraum durch eine Kugel approximiert und für vol_{trCone} durch einen Zylinder.

Aufgrund der fehlenden Positionsvorhersagen für Objektmodellteile ohne primäre Merkmale, wird die Lage der Suchräume für die sekundären Merkmale des Objektmodellteils omp_ν generell durch den letzten Eintrag in der Historie $HIST_\nu$ bestimmt. Damit wird die Lage des Objektmodellteils im letzten Interpretationszyklus zugrundegelegt. Diese einfache Positionsbestimmung kann durch die geringe Anforderung an die Präzision der Suchräume der sekundären Merkmale gerechtfertigt werden, da diese nicht als Restriktionen beim Aufbau der Hypothesen im Interpretationsprozeß verwendet werden.

Zur Bestimmung der Durchmesser d_s der Suchräume für die sekundären Merkmale wird festgelegt, daß der Suchraum nicht größer als der vierfache Durchmesser d_v des Volumenkörpers des Objektmodells sein soll und mit größerer Bildwiederholrate entsprechend kleiner werden soll. Desweiteren muß der Suchraum eine Mindestgröße entsprechend der Größe des Volumenkörpers haben. Aus den Anwendungserfahrungen hat sich ergeben, daß bei einer Bildwiederholrate von z.B. 8 Bildern / Sekunde $d_s = 2 \cdot d_v$ ein sinnvolles Maß ist. Aufgrund dieser Überlegungen und der Forderung, daß der Suchraumdurchmesser mit größerer Taktlänge exponentiell zunehmen soll, ergibt sich für $d_s(t)$:

$$d_s(t) = d_v \cdot \left(1 + 3 \cdot e^{-\frac{0{,}137}{t}}\right)$$

Somit erhält man für einen Durchmesser $d_v = 1$m einen Suchraumdurchmesser d_s in Abhängigkeit der Bildwiederholrate entsprechend der Tab. 3.2.

3.3 Selektion der Kameras

Die Detektion und die Verfolgung von Personen in STABIL^{++} basieren auf einer kompletten 3D Interpretation der Szene. Das hierzu verwendete Objektmodell kapselt daher 3D Wissen über die zu detektierenden Objekte. Ebenso sind über das verwendete Kameramodell von allen Kameras, die in das System integriert sind, die 3D / 2D Abbildungseigenschaften, sowie die 3D Lage bekannt. Entsprechend dem, anhand der Alg. 3.1 - 3.4, skizzierten Ablauf des Interpretationsprozesses werden mittels der 3D Suchräume die für einen Interpretationszyklus zu verwendenden Kameras ausgewählt oder sogar positioniert.

Hierdurch wird sichergestellt, daß für einen Interpretationszyklus nur von den Kameras Bilder eingezogen werden, in deren Sichtbereich Objektmodellinstanzen erwartet werden oder initiale Modellsuchräume liegen. Das bedeutet eine Einschränkung des Aufwands für die Extraktion der Bildmerkmale und somit auch für die Generierung der Hypothesen. Liegt der zu Beginn des Interpretationszyklus bestimmte Suchraum SSP in den 3D Sichtbereichen mehrerer Kameras, so wird von all diesen Kameras ein Bild eingezogen.

Hinsichtlich des bei der Bestimmung der 3D Szenenmerkmale notwendigen 2D / 3D Übergangs ist zu beachten, von wievielen Kameras für einen Interpretationszyklus Bilder einzuziehen sind. Dies hängt unmittelbar mit der zu wählenden Positionierung der Kameras zusammen: In STABIL^{++} werden entweder anhand von 2D Bildmerkmalen aus nur einem Videobild die Tiefeninformation der 3D Szenenmerkmale geschätzt oder es wird anhand von 2D Bildmerkmalen aus zwei oder mehreren Videobildern die 3D Szenenmerkmale mittels eines Stereoansatzes vermessen. Sind die Kameras so angeordnet, daß sich in den Sichtbereichen Überlappungen ergeben, so werden in dem Überlappungsbereich die Bildmerkmale aus (mindestens) zwei Kameras in dem Stereoansatz zur Bestimmung der 3D Szenenmerkmale herangezogen.

In dem folgenden Abschnitt wird erläutert, wie in STABIL^{++} eine implizite Objektübergabe durch das Konzept der Kameraauswahl realisiert ist. In den weiteren Abschnitten wird die vor dem Bildeinzug notwendige Ausrichtung und Brennweiteneinstellung[1] bei der Verwendung von aktiven Schwenk- / Neigekameras mit Motorzoom-Optik und die zur Auswahl der Kameras notwendige Überprüfung der Sichtbarkeit der Suchräume erläutert.

3.3.1 Implizite Objektübergabe

Im Gegensatz zu Systemen, in denen einzelne Interpretationssysteme pro Kamera realisiert sind und somit die Objektmodellinstanzen beim Verlassen des Sichtbereiches einer Kamera an ein weiteres System “übergeben” werden müssen, ist dies in STABIL^{++} implizit durch die Kameraauswahl gegeben. Dies begründet sich auf der strikten 3D Modellierung. Die Objektmodellinstanz ist, aufgrund der 3D Modellierung, im 3D Raum des Szenenmodells definiert und somit wird die Auswahl der Kameras, die für die Re-Detektion zu verwenden sind, über die 3D Suchräume der Objektmodellteile der Instanz durchgeführt. Durchschreitet z.B. eine Person einen Raum, der durch mehrere versetzt angeordnete Kameras eingesehen wird, so wird für die zugehörige Objektmodellinstanz in jedem Interpretationszyklus ein 3D Suchraum bestimmt, der sich in Richtung der Bewegung durch den Raum verschiebt. Liegt dieser Suchraum aufgrund des Fortschreitens der Bewegung nicht mehr in dem Sichtbereich einer Kamera, so wird kein Bild von dieser Kamera eingezogen. Kommt der Suchraum jedoch in Sichtbereichen weiterer Kameras zu liegen, so werden Bilder dieser Kameras in den Interpretationsprozeß mit einbezogen.

[1] Zoom-Funktion.

Bei der Verfolgung von Objektmodellinstanzen ergibt sich quasi ein “Weiterreichen” dieser Instanzen von einer Kamera zur nächsten. Für den Interpretationsprozeß ist es jedoch bei der Auswahl der Kameras immer nur ein Wechsel zu der jeweils in der 3D Lage benachbarten Kamera.

3.3.2 Ausrichtung der Kameras

Stehen dem System aktive Kameras zur Verfügung, so können vor dem Einzug der Bilder im Interpretationsprozeß die Kameras so ausgerichtet werden, daß die Suchräume optimal im Sichtbereich der Kameras liegen. Dies ist jedoch nur möglich, wenn kompakte Suchräume vorliegen, auf die die Kameras ausgerichtet werden können. Aufgrund der unabhängigen Objektbewegungen ist das Nachführen einer Kamera nur bei der Verfolgung einer Objektmodellinstanz sinnvoll. Es ergibt sich bei der Verfolgung einer einzelnen Objektmodellinstanz innerhalb des Observierungsraumes des Szenenmodells ein einzelner Suchraum, der zudem kompakt ist, wenn für das entsprechende Objektmodell anwendungsbedingt nur wenige Merkmale definiert sind.

In STABIL^{++} sind Kuppelkameras (*Dome-Cameras*), wie sie aus der Überwachungstechnik bekannt sind, integriert worden. Die Ansteuerung erfolgt über eine bidirektionale serielle Schnittstellenkomunikation, bei der zum einen eine kontinuierliche Veränderung der Orientierung mit einer bestimmten Winkelgeschwindigkeit und der Brennweite gestartet und wieder gestoppt werden kann. Zum anderen können explizite Positionen angefahren werden, wobei hierzu entsprechende Schrittmotorpositionen zu setzen sind. Umgekehrt kann die Stellung der Kamera über die Positionen der Schrittmotoren abgefragt werden.

Bei diesen Kameras sind die Freiheitsgrade beschränkt auf Schwenken, Neigen und Verändern der Brennweite. Somit ist die Position der Kamera fixiert und nur die Orientierung in zwei Winkeln veränderbar. Zur Beschreibung der Bewegungen sind für die Kuppelkameras neben dem Kamerakoordinatensystem weitere Koordinatensysteme definiert. Dies sind ein *Basiskoordinatensystem*, das die Lage der Grundeinheit im Weltkoordinatensystem angibt und ein *Manipulatorkoordinatensystem*, mit dem die Lage der Schwenk- / Neigeeinrichtung bestimmt ist, vgl. hierzu Anh. C.5 und insbesondere Abb. C.6. Diese Einschränkung in den Freiheitsgraden der Bewegung kann genutzt werden, um den Schwenkwinkel ψ und den Neigewinkel θ zu bestimmen. Desweiteren wird angenommen, daß der Ursprung des Kamerakoordinatensystems mit dem Zentrum der Drehbewegung zusammenfällt.[2] Diese Annahme ist zulässig, da sich zum einen der Versatz bei den handelsüblichen Schwenk- /Neigekameras, die als Kuppelkameras ausgeführt sind, im Bereich von wenigen Zentimetern bewegt. Zum anderen soll ein Suchraum abgebildet werden, dessen Position und Größe aus einer Vorhersage bestimmt worden ist, die verfahrensbedingt eine Ungenauigkeit aufweist.

Die Kamera muß so ausgerichtet werden, daß der mittlere Sichtstrahl der Kamera, der entlang der z-Achse des Kamerakoordinatensystems verläuft, auf den Mittelpunkt des Suchraumes zeigt. Dies ist, entsprechend der Annahmen, gleichbedeutend mit der Ausrichtung der x-Achse des Manipulatorkoordinatensystems in Richtung des Suchraumes. Vgl. hierzu Abb. C.6, in der die Koordinatensysteme von Schwenk- / Neigekameras dargestellt sind, und Abb. 3.3.

[2] Der Versatz zwischen dem Zentrum der Drehbewegung und dem Ursprung des Kamerakoordinatensystems wird, in Analogie zum Einsatz von Videokameras an Manipulatoren von Robotern, als *Hand-Auge-Versatz* bezeichnet. Die Transformation eines Punktes $\vec{p}_{wcs}$ aus dem Weltkoordinatensystem in einen Punkt $\vec{p}_{cam}$, der im Kamerakoordinatensystem einer aktiven Kamera definiert ist, ist in Glg. C.2 angegeben. Durch die Annahme, daß der Versatz zwischen dem Drehzentrum und dem Ursprung des Kamerakoordinatensystems wegfällt, werden mit der Transformationsmatrix ${}^{cam}\mathbf{T}_{manu}$ eine statische Rotation beschrieben, die somit hier unberücksichtigt bleiben kann.

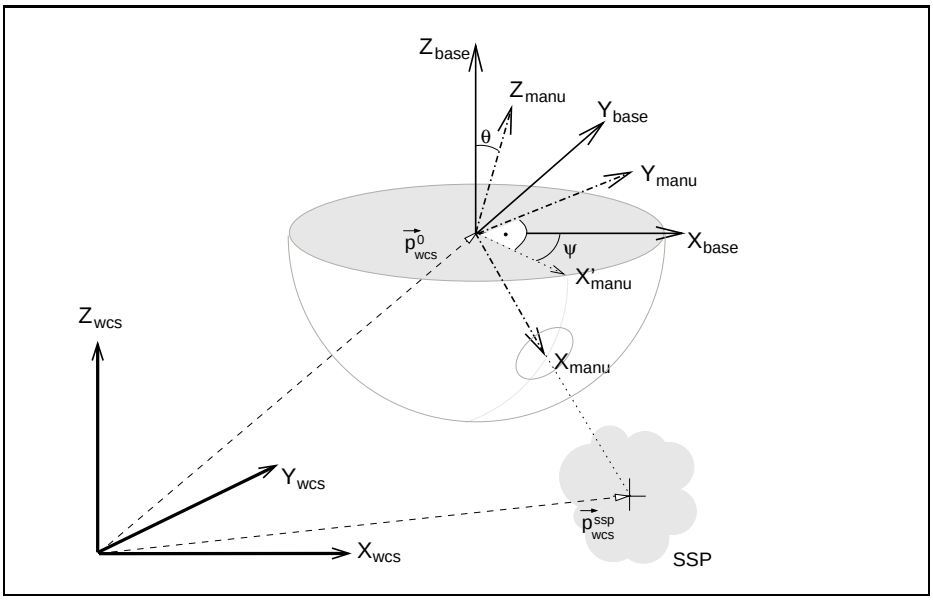


Abbildung 3.3: *Ausrichtung von Schwenk- / Neigekameras.*

Basierend auf der bekannten Lage des Basiskoordinatensystems $base$ im Weltkoordinatensystem und des Mittelpunktes des Suchraumes $\vec{p}_{wcs}^{\,ssp}$ läßt sich die Transformation zwischen dem Manipulatorkoordinatensystem $manu$ und dem Basiskoordinatensystem bestimmen. Es wird festgelegt, daß die y-Achse des Manipulatorkoordinatensystems, aufgrund der fehlenden Rotation um die optische Achse, in der xy-Ebene des Basiskoordinatensystems liegen muß. Desweiteren muß aufgrund des Aufbaus der verwendeten Schwenk- / Neigekameras die z-Achse des Manipulatorkoordinatensystems immer oberhalb der xy-Ebene des Basiskoordinatensystems zum liegen kommen. Es werden die Vektoren $\vec{x}$, $\vec{y}$ und $\vec{z}$ definiert, die in Richtung der drei Achsen des X_{manu}, Y_{manu} und Z_{manu} des Manipulatorkoordinatensystems zeigen und auf das Basiskoordinatensystem $base$ bezogen sind.

Projiziert man nun $\vec{x}$, der vom Ursprung des Basiskoordinatensystems $\vec{p}_{wcs}^{\,0}$ in Richtung des Mittelpunktes des Suchraumes $\vec{p}_{wcs}^{\,ssp}$ zeigt, in die xy-Ebene des Basiskoordinatensystems, so erhält man den Vektor $\vec{x'}$, der in Richtung der Achse X'_{manu} zeigt. Zu $\vec{x'}$ liegt $\vec{y}$ im rechten Winkel und gleichzeitig in der xy-Ebene des Basiskoordinatensystems. $\vec{z}$ steht zu $\vec{x}$ und $\vec{y}$ senkrecht und alle drei Vektoren bilden ein Orthonormalsystem. Diese drei Basisvektoren lassen sich wie folgt bestimmen:

$$
\begin{aligned}
\vec{x} &= {}^{base}\mathbf{T}_{wcs} \cdot (\vec{p}_{wcs}^{\,0} - \vec{p}_{wcs}^{\,ssp}) \\
\vec{y} &= \vec{x'} \times \vec{x} \\
\vec{z} &= \vec{x} \times \vec{y} \\
&\text{mit} \\
\vec{x'} &= \vec{x} \cdot [1,1,0]^T
\end{aligned}
$$

Der Rotationsanteil der Transformationsmatrix ${}^{base}\mathbf{T}_{manu}$ repräsentiert den Schwenkwinkel ψ und den Neigewinkel θ und läßt sich aus den Vektoren $\vec{x}$, $\vec{y}$ und $\vec{z}$ entsprechend der Glg. B.4 - B.9 direkt erzeugen. Hierbei werden die Vektoren auf die Länge eins normiert und ergeben somit die spaltenweisen Einträge in der Rotationsmatrix. Aufgrund des Zusammenfallens der Ursprünge der beiden Koordinatensysteme entspricht der Translationsteil der Matrix dem Vektor $[0,0,0]^T$.

Anhand der so bestimmten Schwenk- und Neigewinkel kann die aktive Kamera positioniert werden. Es sollte sichergestellt werden, daß die Kamera bei großen Schwenk- und Neigebewegungen vor dem Einzug des Bildes sich wieder in Ruhelage befindet, um eine Verwischung des Bildinhaltes aufgrund der Kamerabewegung zu vermeiden. Es ist daher ratsam, die Ausrichtung bei der Verfolgung nur dann auszuführen, wenn sich der zu verfolgende Suchraum weit aus der Mitte des Bildes entfernt hat. Hierdurch reduziert man die Anzahl der Kameraansteuerungen.

Eine weitere Möglichkeit ist ein kontinuierliches Verfahren der Kamera, wobei die Ansteuerung gedämpft über einen PD- oder PID-Regler erfolgen sollte. Hierbei muß jedoch sichergestellt werden, daß zu jedem Zeitpunkt, zu dem von der Kamera ein Bild eingezogen wird, der exakte Schwenk- und Neigewinkel bekannt ist. Die Winkeldaten werden benötigt, um die äußeren Kameraparameter $camPose$ zu bestimmen, die für das eingezogene Bild vermerkt werden müssen.

3.3.3 Wahl des Blickwinkels (Zoom)

Die in STABIL^{++} integrierten Kuppelkameras erlauben neben dem Schwenken und Neigen noch eine Veränderung der Brennweite (Zoom) und somit des Blickwinkels der Kamera. Hiermit kann man das Segmentierungsproblem[3] entschärfen und eine längere Brennweite / kleineren Blickwinkel wählen, wenn die Objektmodellinstanz in einer größeren Entfernung erwartet wird, und umgekehrt. Es wird angestrebt, daß der Suchraum fast formatfüllend in dem Bild der Kamera abgebildet wird. Der Blickwinkel ergibt sich entsprechend des verwendeten Kameramodells aus der Größe der virtuellen Bildebene und der Kammerkonstante b. Die Größe der virtuellen Bildebene ergibt sich wiederum aus der Anzahl der horizontalen und vertikalen Bildpunkte und den entsprechenden Skalierungsfaktoren S_x, S_y. Man erhält aufgrund der unterschiedlichen Bildhöhe und -breite einen vertikalen und einen horizontalen Blickwinkel. Entspricht die Größe der virtuellen Bildebene der Größe der real genutzten Fläche des CCD-Chips der Kamera, so kann die Kammerkonstante b mit der Brennweite des Objektives gleichgesetzt werden. Man kann dann direkt über den Blickwinkel die Brennweite bestimmen. Da die Chipgröße konstant ist, reduziert sich das Problem der Bestimmung des Blickwinkels auf die Bestimmung der Kammerkonstante b.

Das Verstellen der Blickwinkel eignet sich nur im Zusammenhang mit dem Ausrichten der aktiven Kamera und kommt daher bei der Verfolgung einzelner Objektmodellinstanzen zum Einsatz. Hierbei ergeben sich einzelne, kompakte Suchräume. Aufgrund der Ungenauigkeit in der Vorhersage des Suchraumes und der damit verbundenen Ungenauigkeit bei der Ausrichtung des mittleren Sichtstrahls der Kamera auf den Mittelpunkt des 3D Suchraumes, muß die Brennweite so gewählt werden, daß eine um 10% in der Ausdehnung größere Fläche des Suchraumes abgebildet werden kann.

Um den 3D Suchraum der zu verfolgenden Objektmodellinstanz wird hierzu eine umschließende Kugel mit dem Radius r gelegt. Der für die Kamera zu wählende Blickwinkel ist dann so zu wählen, daß die Kreisprojektion dieser Kugel mit einem Radius $r_2 = 1,1 \cdot r$ abgebildet werden kann. Bei den üblichen Bildformaten von 4:3 (Breite:Höhe) ist verständlich, daß zur Bestimmung der Kammerkonstante b der vertikale und somit kleinere Blickwinkel verwendet wird.

Der Punkt $\vec{p}_{wcs}$ am oberen Rand des Kreises, der durch den Mittelpunkt $\vec{p}^{\,0}_{wcs}$ des Suchraumes und den Radius r_2 bestimmt ist, soll in der Bildebene an dem Punkt $\vec{p}_{img} = [C_x, 0]^T$

[3] Das Segmentierungsproblem, auch Segmentationsproblem, tritt bei der Objektdetektion auf, wenn die Bildmerkmale nicht mehr im Bild zu extrahieren sind; der hier angesprochene Sachverhalt bezieht sich auf die Größe der Merkmale im Bild.

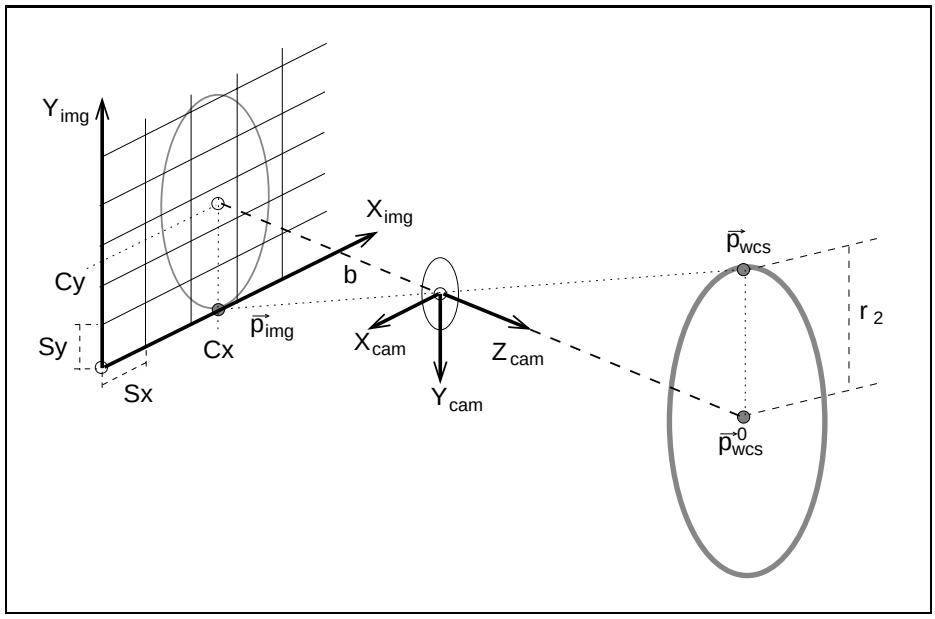


Abbildung 3.4: *Bestimmung des Blickwinkels; die Lage des Rechnerkoordinatensystems* $[X_{img}, Y_{img}]$ *ist aufgrund des Lochkameramodells spiegelverkehrt zum Kamerakoordinatensystem* $[X_{cam}, Y_{cam}]$ *orientiert.*

abgebildet werden, vgl. hierzu Abb. 3.4. Vernachlässigt man die Verzerrung und setzt somit den Verzerrungskoeffizienten $\kappa = 0$, so erhält man entsprechend der Glg. 2.17 und 2.18 für y_{img}:

$$y_{img} = \frac{u_y}{S_y} + C_y \tag{3.7}$$

Da $y_{img} = 0$ ist ergibt sich aus Glg. 2.16 dann für die Kammerkonstante b:

$$b = \frac{z_{cam} \cdot C_y \cdot S_y}{y_{cam}}$$

Dies kann auch aus Abb. 3.4 abgelesen werden, wenn man die Strecken von b und den Abstand des Mittelpunktes des Suchraumes zum Ursprung des Kamerakoordinatensystems mit den Strecken von $C_y \cdot S_y$ und r_2 ins Verhältnis setzt. Der hierzu notwendige Punkt $\vec{p}_{cam}$ ergibt sich aus $\vec{p}^{\,0}_{wcs}$ und der bekannten Transformationsmatrix ${}^{cam}\mathbf{T}_{wcs}$ entsprechend:

$$\vec{p}_{cam} = \begin{pmatrix} x_{cam} \\ y_{cam} \\ z_{cam} \end{pmatrix} = {}^{cam}\mathbf{T}_{wcs} \cdot \left[\vec{p}^{\,0}_{wcs} + \begin{pmatrix} 0 \\ -r_2 \\ 0 \end{pmatrix} \right] \tag{3.8}$$

Aus der so bestimmten Kammerkonstante b und dem entsprechenden Blickwinkel läßt sich eine für die aktive Kamera einzustellende Brennweite ableiten. Es ist hierbei zu beachten, daß sich durch die Brennweitenveränderung neben der Kammerkonstante b auch noch weitere Parameter der internen Kameraparameter $camPar$ ändern. So verändert sich durch die Drehbewegung bei den Motorzoom-Objektiven grundsätzlich die Position des Hauptpunktes $[C_x, C_y]^T$.

Um die Modellierung der Veränderung in den internen Kameraparametern bei kontinuierlicher Brennweitenänderung zu umgehen, kann man die internen Kameraparameter durch Kalibrierung für verschiedene Brennweiten bestimmen. Die hierzu auszuwählenden Brennweiten

sollten gleichmäßig abgestuft im Brennweitenverstellbereich der Optik der aktiven Kamera liegen. Wird eine Brennweite / ein Blickwinkel entsprechend der Glg. 3.7 - 3.8 bestimmt, so muß die entsprechend größere Brennweite, also der entsprechend kleinere Blickwinkel gewählt werden, für den die internen Kameraparameter vorliegen.

Es ist weiterhin zu beachten, daß sich aufgrund der Brennweitenänderung in handelsüblichen Motorzoom-Objektiven die Position des Kamerakoordinatensystems entlang der optischen Achse verschiebt und somit sich auch die äußeren Kameraparameter $camPose$ ändern. Um dies zu umgehen, werden in Spezialkameras bei einer Veränderung der Brennweite nicht das Linsensystem verschoben, sondern die Bildebene. In STABIL^{++} muß, bei der Verwendung von handelsüblichen Kuppelkameras, nach Ausrichtung der Kamera auf den Mittelpunkt des Suchraumes und anschließender Veränderung der Brennweite, die Transformationsmatrix ${}^{cam}\mathbf{T}_{wcs}$, die den Hand-Auge-Versatz beschreibt, um die Translation in Richtung der optischen Achse korrigiert werden. Die hierdurch bewirkte geringere Genauigkeit ist in den Anwendungen von STABIL^{++} akzeptabel.

3.3.4 Sichtbarkeit der Suchräume

Um die Kameras auszuwählen, von denen im aktuellen Interpretationszyklus ein Bild einzuziehen ist, wird für jede in dem Szenenmodell bekannte Kamera geprüft, ob in deren Sichtbereich ein Teil des Suchraumes SSP sichtbar ist, vgl. Zeilen 11 - 19 im Alg. 3.2. Die kleinste Einheit des Suchraumes SSP ist hierbei der Suchbereich eines Merkmals eines einzelnen Objektmodellteiles einer der zu detektierenden Objektmodellinstanzen, vgl. Alg. 3.5.

Für diese Überprüfung wird getestet, ob der Sichtstrahl $\vec{s}$, der vom Ursprung ${}^{\circ}\vec{p}^{\,cam}$ des Kamerakoordinatensystems durch den Mittelpunkt $\vec{p}^{\,ssp}$ eines Einzelsuchraumes verläuft,[4] eine der Weltregionen $wr \in SSP_s$ schneidet. Die Weltregionen des Inventars des Szenenmodells beschreiben hierbei die Wände, die geschlossenen Türen, die Decke und den Boden der Räume, in denen observiert wird. Alle Weltregionen zusammen beschreiben daher den Observierungsraum SSP_s der Szene. Zusätzlich zu den genannten Weltregionen kann der Observierungsraum durch weitere "Sichthindernisse", wie z.B. Schränke eingeschränkt werden. Die Weltregionen, aus denen sich der Observierungsraum aufbaut, sind in STABIL^{++} als Ebenen $\in \mathbb{R}^3$ definiert und durch einen Eckpunkt $\vec{a}$ und zwei weitere Vektoren $\vec{u}$ und $\vec{v}$ bestimmt. Ein Beispiel ist hierzu in Abb. 3.5 gegeben.

Der Schnittpunkt $\vec{p}^{\,i}$ des Sichtstrahls $\vec{s}$ mit der durch die Vektoren $\vec{a}$, $\vec{u}$ und $\vec{v}$ bestimmten Ebene ist durch:

$$
\begin{aligned}
\vec{p}^{\,i}_{wcs} &= {}^{\circ}\vec{p}^{\,cam}_{wcs} + \frac{\vec{n}_{wcs} \cdot (\vec{a}_{wcs} - {}^{\circ}\vec{p}^{\,cam}_{wcs})}{\vec{n}_{wcs} \cdot \vec{s}_{wcs}} \cdot \vec{s}_{wcs} \\
\text{mit} & \\
\vec{s}_{wcs} &= \vec{p}^{\,ssp}_{wcs} - {}^{\circ}\vec{p}^{\,cam}_{wcs} \\
\vec{n}_{wcs} &= \vec{u}_{wcs} \times \vec{v}_{wcs}
\end{aligned}
$$

gegeben, wobei $\vec{n}_{wcs}$ der Normalenvektor der Ebene mit Bezug zum Weltkoordinatensystem wcs ist. Faßt man die Vektoren $\vec{u}_{wcs}$, $\vec{v}_{wcs}$, $\vec{n}_{wcs}$ als die Basisvektoren eines Orthonormalsystems auf, so kann man entsprechend den Glg. B.4 - B.9 den Rotationsanteil einer Transformationsmatrix ${}^{wcs}\mathbf{T}_{area}$ bestimmen. Setzt man den Translationsanteil der Matrix entsprechend des Vektors $\vec{a}_{wcs}$, so beschreibt ${}^{wcs}\mathbf{T}_{area}$ die Transformation zwischen dem Weltkoordinatensystem und dem auf der Ebene der Weltregion definierten Koordinatensystem. Mit ${}^{area}\mathbf{T}_{wcs}$ ist die dazu inverse Transformation bestimmt.

[4] Vgl. auch die Bestimmung des Sichtstrahls in den Glg. 2.19 - 2.21.

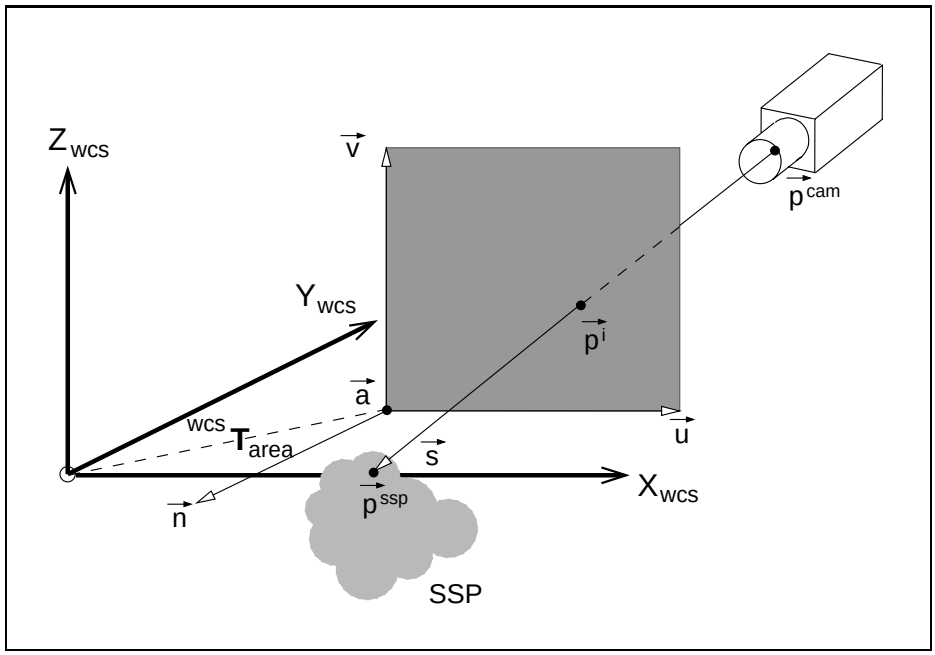


Abbildung 3.5: *Überprüfung der Sichtbarkeit von Suchräumen.*

Transformiert man nun $\vec{p}^{\,i}_{wcs}$, $\vec{u}_{wcs}$ und $\vec{v}_{wcs}$ entsprechend:

$$\begin{aligned}
\vec{p}^{\,i}_{area} &= \begin{pmatrix} p_x \\ p_y \\ p_z \end{pmatrix} = {}^{area}\mathbf{T}_{wcs} \cdot \vec{p}^{\,i}_{wcs} \\
\vec{u}_{area} &= \begin{pmatrix} u_x \\ u_y \\ u_z \end{pmatrix} = {}^{area}\mathbf{T}_{wcs} \cdot \vec{u}_{wcs} \\
\vec{v}_{area} &= \begin{pmatrix} v_x \\ v_y \\ v_z \end{pmatrix} = {}^{area}\mathbf{T}_{wcs} \cdot \vec{v}_{wcs}
\end{aligned}$$

in das neue Koordinatensystem, so gilt:

$$\begin{array}{ccccc} 0 & \leq & p_x & \leq & u_x \quad \wedge \\ 0 & \leq & p_y & \leq & v_y \end{array}$$

falls der Sichtstrahl $\vec{s}$ die durch $\vec{a}$, $\vec{u}$ und $\vec{v}$ bestimmte Weltregion des Szenenmodells schneidet.

Wenn mindestens ein Sichtstrahl zu den Mittelpunkten der Einzelsuchräume $ssp_i \in SSP$ nicht durch eine Weltregion $wr_j \in SSP_s$ des Observierungsraumes des Szenenmodells unterbrochen wird, so liegt der Suchraum SSP in dem Sichtbereich der Kamera und ist somit sichtbar. Dementsprechend werden all die Kameras für den Einzug von Bildern ausgewählt, in denen SSP sichtbar ist.

3.4 Bildverarbeitung

Die in STABIL^{++} verwendeten Sensoren zur Interpretation der Szene sind CCD-Kameras. Der Interpretationsprozeß stützt sich daher zur Bestimmung der 3D Szenenmerkmale auf 2D Bildmerkmale. In diesem Kapitel zur Bildverarbeitung wird beschrieben, wie diese Bildmerkmale gewonnen werden.

Entsprechend dem in Kap. 3.1 dargestellten Prozeßablauf werden in dem Teilschritt der Bildgenerierung zunächst von den ausgewählten Kameras Bilder eingezogen und auf diesen eine Bildvorverarbeitung durchgeführt. In dem Teilschritt der Detektion werden anschließend aus den vorverarbeiteten Bildern entsprechend der Merkmale der Objektmodellteile die Bildmerkmale extrahiert. In den folgenden Abschnitten wird auf diese einzelnen Schritte eingegangen, wobei auf die im Kap. 2.5 aufgestellten Definitionen der Kamera cam und der Bilder img in den Glg. 2.11 und 2.12 zurückgegriffen wird.

Die in STABIL^{++} verwendete Bildverarbeitung stützt sich auf das Bildverarbeitungssystem HALCON[1]. Dies umfaßt das Erzeugen der Bilder, die Bildverarbeitungsoperatoren zur Segmentierung der Bilder, die Bestimmung von Maßzahlen auf den segmentierten Regionen, bis hin zur Darstellung und Überblendung der Ergebnisse in die Videobilder. Im Rahmen der Entwicklungen zu STABIL^{++} ist in HALCON die verwendete Kamerakalibrierung integriert worden, vgl. Anh. C. Desweiteren ist die verwendete adaptive Hintergrundschätzung im Rahmen dieser Arbeit und der verwendete adaptive Farbklassifikator als Operator in HALCON realisiert worden [Haf99].

3.4.1 Bildeinzug

Im Teilschritt zur Bildgenerierung wird zunächst überprüft, für welche im System integrierten Kameras der aktuelle Suchraum SPP sichtbart ist. Von den ausgewählten Kameras muß ein Videobild eingezogen werden. Hierbei ist zwischen verschiedenen Kameratypen zu unterscheiden.

Während bei den File- und Memory-Cameras jeweils nur die entsprechenden folgenden Bilder aus dem Speicher zu holen sind, muß bei den Live-Cameras das Bild von einer Digitalisierungskarte (*Framegrabber*) eingezogen werden. Insbesondere bei der Verwendung von mehreren Kameras im System muß, im Hinblick auf die Bestimmung der 3D Szenenmerkmale über einen Stereoansatz, auf eine Zeitgleichheit beim Einziehen der Bilder geachtet werden. Ist dies nicht gewährleistet, so kommt es, aufgrund der Bewegung der zu beobachtenden Objekte und den damit unterschiedlichen Positionen der zu extrahierenden Merkmale in den Bildern, zu Ungenauigkeiten bei der Bestimmung der Szenenmerkmale.

Entsprechend der Videonorm (z.B. PAL) erzeugt eine analoge Videokamera mit einer Bildwiederholrate von 25 Vollbildern / Sekunde alle 40 ms ein Bild.[2] Diese 40 ms werden benötigt, um die Signale zu übertragen. Daher muß beim Start der Aufzeichnung über eine Digitalisierungskarte jedesmal so lange gewartet werden, bis bei der Signalübertragung der Start eines neuen Bildes signalisiert wird. Diese Wartezeit beträgt im Mittel 20 ms. Wird jedoch von mehreren Kameras und somit von mehreren Digitalisierungskarten ein Bild eingezogen, so addiert

[1] HALCON ist ein Produkt der MVTec Software GmbH, München, http://www.mvtec.com. HALCON ist aus dem Bildverarbeitungssystem HORUS, das seit Mitte der 80er Jahre am Lehrstuhl von Prof. Radig an der Technischen Universität München entwickelt wurde, hervorgegangen [ELMG^{+}93, ES96, ES97]. Anm.: Bildverarbeitungssystem ist die geläufige Bezeichnung, exakter ist jedoch der Begriff Bildanalysesystem.

[2] Entsprechend der jeweiligen halben Netzfrequenz ergibt sich in Europa mit der PAL-Norm eine Bildfrequenz von 25 Hz und in den USA und Japan mit der NTSC-Norm 30 Hz.

sich jeweils die Wartezeit. Um dies zu vermeiden, werden die Videokameras synchronisiert. Mit der Synchronisation ist gewährleistet, daß alle im System integrierten Kameras zum gleichen Zeitpunkt mit dem Auslesen des CCD-Chips und mit der Signalübertragung eines Videobildes beginnen.

Um die Zeitgleichheit bei der Aufzeichnung zu garantieren, werden idealerweise zusätzlich noch synchronisierbare Digitalisierungskarten verwendet. Bei diesen wird mit einem Signal (*trigger signal*) der Bildeinzug angestoßen, die Bilder auf den Digitalisierungskarten zwischengespeichert und dann erst dem Interpretationsprozeß zur Verfügung gestellt. Stehen nur herkömmliche, unabhängige Digitalisierungskarten zur Verfügung, so muß gewährleistet werden, daß die Digitalisierung mit den kürzest möglichen zeitlichen Abständen erfolgt. Daher wird, entsprechend dem Alg. 3.2 auch die Bildvorverarbeitung für die eingezogenen Bilder erst dann ausgeführt, wenn alle Bilder im System vorliegen. Es wird daher beim Einziehen der Bilder lediglich für das erzeugte Bild img die Bildkanäle CAN und der Aufnahmezeitpunkt t gesetzt. Für die Bilder der Kameratypen File- und Memory-Camera ist jeweils bei der vor der Verarbeitung durchgeführten Aufnahme (*off-line* Verarbeitung) auf die Zeitgleichheit zu achten.

Die zum Aufnahmezeitpunkt geltenden Kameraparameter des verwendeten Kameramodells $camPar$ und $camPose$ werden den erzeugten Bildern ebenfalls beigefügt. Für stationäre Kameras sind diese Parameter zuvor durch den Prozeß der Kamerakalibrierung ermittelt worden. Bei der Verwendung von aktiven Kameras muß jedoch gewährleistet sein, daß zum Aufnahmezeitpunkt eine gültige Kalibrierung vorliegt.

3.4.2 Bildvorverarbeitung

Auf den eingezogenen Bildern wird unmittelbar nach dem Einzug mit den der entsprechenden Kamera $cam_i \in CAM$ bekannten Bildverarbeitungsoperatoren $ip(.) \in IP$ und Klassifikatoren $clc^{attr_j} \in CLC$ eine Vorverarbeitung ausgeführt. Diese Vorverarbeitung wird als sog. *low-level* Verarbeitung bezeichnet, da hier nur auf der Ebene der Bildpunkte eine Bildverbesserung oder Segmentierung des Bildes vorgenommen wird und generell noch kein Modellwissen verwendet wird. Eine Ausnahme bildet die Projektion der Suchräume, die den Kameras bekannt sind.

Die durch die Vorverarbeitung verbesserten Bilder und die durch die Segmentierung erzeugten Bildregionen mit entsprechenden Eigenschaften / Attributen dienen als Grundlage der Extraktion der Bildmerkmale im weiteren Interpretationsprozeß.

Bildverbesserung

Die einer Kamera zugehörigen Bildverarbeitungsoperatoren $IP = \{ip_1(.), \ldots, ip_n(.)\}$ zur Bildverbesserung sind z.B. Operatoren zur Kontrastverstärkung oder zur Beseitigung des *Interlace*-Effektes. Es können den Kameras jedoch weitere *low-level* Operatoren zur Bildverbesserung zugeordnet werden, wenn dies für die eigentliche Extraktion der Bildmerkmale notwendig ist.

Kontrastverstärkung: Hauptsächlich Bildverarbeitungsoperatoren, die anhand von Gradientenverfahren Bildkanten suchen und Operatoren zur Texturanalyse benötigen Bilder, deren Grauwerthistogramm eine gleichmäßige Ausnutzung des Grauwertspektrums zeigt. Es ist jedoch häufig durch ungleichmäßige oder schwankende Beleuchtung bei der Aufnahme bedingt, daß der obere (helle) und der untere (dunkle) Bereich des Spektrums nicht genutzt wird. Mit einer Kontrastverstärkung / Kontrastoptimierung läßt sich der Grauwertverlauf verändern. Hierbei wird zwischen dem Aufhellen, dem Abdunkeln und der Kontrasterhöhung der Bilder unterschieden. Dementsprechend werden hierzu unterschiedliche Kennlinien verwendet, [Ric95].

Es ist zu beachten, daß durch diese Verfahren der Informationsgehalt des Bildes nicht verbessert werden kann, es wird lediglich die Darstellung im Grauwertbereich verändert. Daher ist vorrangig für optimale Aufnahmebedingungen zu sorgen. Das bedeutet, daß die Szene möglichst gleichmäßig auszuleuchten ist und die Blende der Optik gerade soweit zu öffnen ist, daß keine Überstrahlungen auftreten.

Neben der Kontrastverstärkung durch Bildverarbeitungsoperatoren gibt es noch drei weitere Methoden zu einer optimalen Ausnutzung des Grauwertspektrums: Erstens können Objektive mit einer automatischen videosignal-gesteuerten Blende verwendet werden. Zweitens ist eine Signalanpassung/-verstärkung in allen modernen Kameras vorhanden. Drittens kann die Signalverstärkung auch noch in einigen Digitalisierungskarten vorgenommen werden. Die Signalverstärkung ist als *automatic gain control* (AGC) bekannt. Die beiden letzten Methoden bringen keinen weiteren Informationsgewinn und sind daher direkt mit den Bildverarbeitungsoperatoren zur Kontrastverstärkung zu vergleichen. Die automatische Blendensteuerung optimiert dagegen die auf den CCD-Chip der Kamera einfallende Lichtmenge.

Im Hinblick auf die in STABIL^{++} verwendeten Methoden zur Segmentierung der Bilder sei hier darauf hingewiesen, daß bei adaptiven Verfahren eine dynamische und sich somit in der Kennlinie veränderliche Kontrastverstärkung nicht verwendet werden kann. Daher ist auf eine Signalverstärkung und automatische Blendensteuerung zu verzichten. Das Gleiche gilt für einen automatischen Weißabgleich bei Farbkameras. Hierbei wird in allen drei Farbkanälen (Rot-, Grün-, Blau-Kanal) quasi eine Kontrastoptimierung vorgenommen, so daß selbst bei unterschiedlicher Farbe der Beleuchtung ungefähr ein mittlerer Farbwert des Bildes erhalten bleibt.

Beseitigung des Interlace-Effektes: Von analogen Videokameras werden mit einer Frequenz von 25 oder 30 Hz Vollbilder erzeugt.[3] Jedoch teilt sich ein Vollbild (*frame*) in zwei Halbbilder (*fields*) – jeweils aus den Zeilen mit geraden und ungeraden Zeilennummern. Entsprechend der Videonorm werden die beiden Halbbilder im Videosignal nacheinander übertragen. Dementsprechend werden bei herkömmlichen Kameras auch die Elemente einer Spalte des CCD-Chips zweier aufeinanderfolgender Zeilen mit einem Versatz von 40 oder 33.3 ms ausgelesen, [Sei99]. Das hat zur Folge, daß bei schnellen Objektbewegungen, besonders bei einer Bewegungsrichtung in Richtung der horizontalen Ausdehnung des CCD-Chips, scheinbar ein horizontaler Versatz zwischen zwei aufeinander folgenden Bildzeilen entsteht, der als *Interlace*-Effekt bezeichnet wird.

Man begegnet dem Interlace-Effekt, indem man z.B. durch eine Interpolation zwischen zwei aufeinanderfolgenden Zeilen diese quasi gegeneinander verschiebt. Eine weitere Möglichkeit besteht darin, generell nur jede zweite Zeile für die Bildverarbeitung zu verwenden oder eine Zeilendoppelung vorzunehmen, [Ric95]. Es ist bei dieser Veränderung der Bildinhalte darauf zu achten, inwieweit diese einen Einfluß auf die inneren Kameraparameter $camPar$ haben.

Auch hier gibt es Möglichkeiten, den Interlace-Effekt bei der Aufnahmeeinheit zu unterdrücken. Zum einen besteht bei den meisten Kameras die Möglichkeit, durch die Wahl einer kurzen Belichtungszeit (*shutter*) die Zeit zu verkürzen, mit der der CCD-Chip belichtet wird. Damit wird zudem noch erreicht, daß auch die "Verwischung" innerhalb einer Zeile verringert wird. Die Reduzierung der Belichtungszeit ist jedoch gleichzusetzen mit einer Verringerung der Beleuchtungsstärke; daher ist bei kurzen Belichtungszeiten für eine ausreichend gute Beleuchtung zu sorgen. Wird jedoch nur die Blende weiter geöffnet, so reduziert sich der Tiefenschärfebereich.

[3] Vgl. auch Anm. auf S. 76.

Die zweite Möglichkeit bieten *progressive-scan*-Kameras, die in zunehmendem Maße angeboten werden. Bei diesen Kameras wird der Interlace-Effekt gezielt dadurch umgangen, daß zwei aufeinander folgende Bildzeilen auch nacheinander ausgelesen werden. Dies ist insbesondere dann sinnvoll, wenn statt der analogen Videosignalübertragung und anschließender Digitalisierung in der Digitalisierungskarte des Rechners, die Digitalisierung direkt in der Kamera erfolgt und die digitalen Signale der Bildpunkte über eine Schnittstellenkarte in den Rechner gelangen. Man spricht dann von Digitalkameras.[4]

Segmentierung

Die Segmentierung der Videobilder anhand der Klassifikatoren $clc^{attr_j} \in CLC$, die der entsprechenden Kamera bekannt sind, erfolgt auf den korrigierten und verbesserten Bildern. Mit den Klassifikatoren werden die einzelnen Bildpunkte des Bildes zu Regionen mit gleichen Eigenschaften / Attributen zusammengefaßt. So werden in STABIL^{++} durch eine Vorder- / Hintergrundsegmentierung zunächst für die Bilder Regionen bestimmt, in denen aufgrund von Veränderungen in den Bildern einer Bildfolge Objekte zu erwarten sind. Diese Regionen werden entsprechend des Attributes "dem Vordergrund zugehörig" in $REG^{(fg)} = \{reg_1^{(fg)}, \ldots, reg_n^{(fg)}\}$ vereinigt.

Desweiteren werden anhand des 3D Suchraumes SSP, der der Kamera bekannt ist, entsprechend projizierte Bildregionen $REG^{(ssp)} = \{reg^{(ssp_1)}, \ldots, reg^{(ssp_n)}\}$ gebildet. Auch über diese Regionen wird das Bild auf Bereiche beschränkt, in denen Objekte zu erwarten sind, nur daß hier durch die vorhergesagten Suchräume auch Modellwissen verwendet wird. Die Einschränkung der Bilder auf die Regionen $REG^{(fg)}$ und $REG^{(ssp)}$ wird auch als Aufmerksamkeitssteuerung (*region of interrest*) bezeichnet. Die weiteren Bildverarbeitungsoperatoren werden daher nur noch in der Schnittmenge dieser Bildregionen angewendet, wodurch eine Beschleunigung der weiteren Verarbeitung erreicht wird.

Ein wichtiges Merkmal, das in STABIL^{++} Verwendung findet, ist die Farbe von Objektmodellteilen oder von Markierungen der Gelenke. Daher ist den Kameras ein adaptiver Farbklassifikator clc^{color} zur Bestimmung von Bildregionen mit gleicher Farbe bekannt, so daß dem Bild z.B. Regionen $REG^{(red)}$, $REG^{(blue)}$ oder $REG^{(skincolored)}$ zugeordnet werden können. Die angesprochenen drei Segmentierungen werden in den folgenden Abschnitten kurz erläutert.

Vorder- / Hintergrundsegmentierung: Zur Segmentierung des Bildes in eine Vorder- und Hintergrundregion wird in STABIL^{++} ein adaptiver Hintergrundschätzer verwendet.[5] Hierbei wird für jeden Bildpunkt ein Filter zur Schätzung des Grauwertes des Bildhintergrundes verwendet. Der Filter basiert auf der Theorie des in [Kal60] vorgestellten Kalmanfilters. Der eingesetzte Filter verwendet hierbei einen Zustandsvektor $\hat{g}$, in dem der eigentliche Grauwert und die zeitliche Änderung (erste Ableitung) des Grauwertes abgebildet ist. Entsprechend einer Modellvorstellung des Grauwertverhaltens in einer Systemmatrix A wird ein neuer Systemzustand $\tilde{g}$ vorhergesagt. Der vorhergesagte Wert wird mit dem gemessenen Grauwert g aus einem Kanal $can_i \in CAN$ des aktuellen Bildes img verglichen.[6] Ist die Abweichung klein, dann geht man

[4] Digitalkameras, bei denen anstelle eines analogen Videosignals direkt ein Feld von digitalen Bildpunkten abgegriffen werden kann, sind nicht mit Kameras zu verwechseln, die lediglich über eine digitale Ansteuerung oder eine Bildverbesserung mit digitalen Signalprozessoren verfügen.

[5] Eine Gegenüberstellung von verschiedenen weiteren Algorithmen zur Vorder- / Hintergrundsegmentierung findet sich in [TKBM99].

[6] Sind mehrere Farbkanäle vorhanden, so kann der Filter auf allen Kanälen angewendet werden und die Vorder- / Hintergrundentscheidung mit den Ergebnissen in allen Kanälen getroffen werden, vgl. [Sti96]. In vielen Anwendungen reicht es jedoch aus, einen Kanal zu verwenden oder zuvor aus drei Farbkanälen ein Grauwertbild

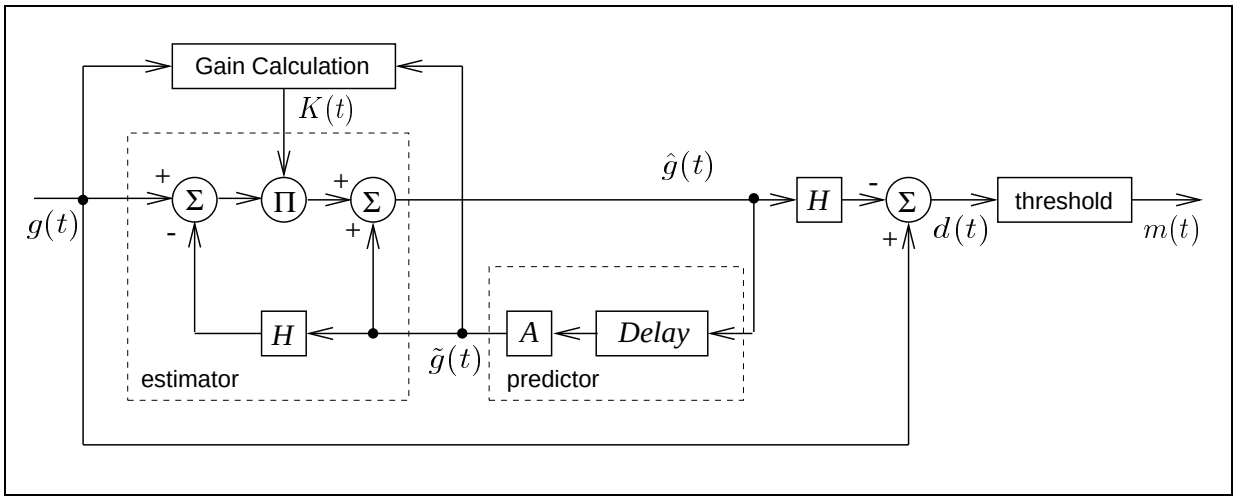


Abbildung 3.6: *Schematische Blockdarstellung des Kalmanfilters zur Schätzung der Grauwerte des Hintergrundbildes.*

davon aus, daß weiterhin in dem Bildpunkt Hintergrund abgebildet wird. Die kleine Änderung, die zum einen von nicht zu vermeidenden Kamerarauschen oder natürlichen Beleuchtungsänderungen verursacht sind, werden in den neuen Zustandsvektor adaptiert, d.h. es wird die Vorhersage mit einem großen Kalman-Gain K multipliziert und zur Vorhersage hinzuaddiert, vgl. Abb. 3.6.

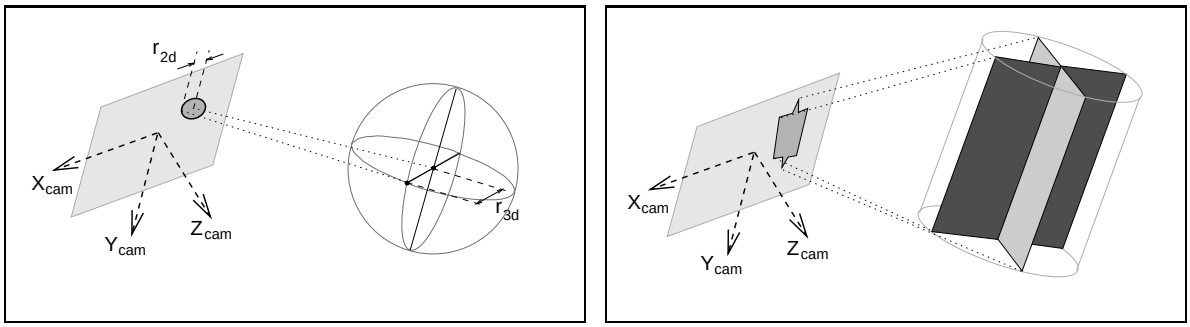
Ist die Abweichung zwischen Vorhersage und gemessenem Grauwert jedoch größer als ein zu bestimmender Schwellenwert (*threshold*), so geht man davon aus, daß in dem Bildpunkt nicht mehr der Hintergrund abgebildet wird, sondern der Hintergrund von einem Objekt im Vordergrund verdeckt wird. Große Grauwertänderungen, die somit, entsprechend der Modellvorstellung, zu Objekten im Vordergrund gehören, werden langsamer oder nicht in das geschätzte Hintergrundbild adaptiert. Hierzu wird dann ein anderes Kalman-Gain bestimmt und die Differenz zwischen Schätzwert $\hat{g}$ und Meßwert g entsprechend anders gewichtet. Bei der Bestimmung des zu verwendenden Kalman-Gain (*Gain Calculation*) wird somit zwischen einer Zuordnung des Bildpunktes zu Vorder- oder Hintergrund unterschieden.

Zur Bestimmung der Vordergrundregion wird die Differenz d zwischen dem Grauwert g des Bildpunktes im aktuellen Bild und dem geschätzten Grauwert $\hat{g}$ des Hintergrundes gebildet. Ist die Differenz größer als ein zu bestimmender Schwellenwert, so wird der Bildpunkt mit $m(t) = 1$ "als zur Vordergrundregion zugehörig" bezeichnet.[7] In Abb. 3.8 (a) ist die Vordergrundregion eingezeichnet, die sich durch die eine sich im Raum bewegende Person ergibt. Man erkennt zum einen, daß die Region des Objektes nicht komplett, d.h. vollflächig, als Vordergrund gekennzeichnet wurde. Dies begründet sich auf teilweise geringe Differenzen zwischen den Grauwerten eines Bildpunktes im geschätzten Hintergrundbild und im aktuellen Bild. Zum anderen sind im Bild noch weitere vereinzelte Bildpunkte als Vordergrund gekennzeichnet. Diese beruhen auf dem Rauschen im Kamerasignal. Um eine vollflächige Vordergrundregion des Objektes zu erhalten, werden mit morphologischen Operatoren vereinzelte Punkte eliminiert und Lücken geschlossen[8].

zu bestimmen, vgl. [Haf99].

[7] $m(t) \in [0:1]$.

[8] Hier werden die binären morphologischen Operationen "Opening" und "Closing" mit runden Masken verwendet, [Hab91].



(a) *(b)*

Abbildung 3.7: *Schematische Projektion (a) kugelförmiger und (b) zylindrischer Einzelsuchräume.*

Der hier verwendete Ansatz wurde ursprünglich in [KvB90] vorgestellt und im Rahmen des Projektes zu STABIL^{++} für die Anforderungen der Personendetektion und -verfolgung angepaßt. Die exakte Arbeitsweise der realisierten Hintergrundschätzung ist ausführlich in [RMK95] veröffentlicht worden. Ergänzende Untersuchungen, insbesondere zur Wahl der Parameter der Systemmatrix A und deren Abhängigkeit von sich ändernder Bildwiederholrate sind in [Sta99] zu finden. Desweiteren ist in [Ebe99] eine Erweiterung des zunächst nur für stationäre Kameras geltenden Ansatzes auf Kuppelkameras zu finden.[9] Dort werden einzelne Hintergrundbilder auf eine Halbkugel projiziert, so daß ein komplettes Hintergrundbild für den Schwenk- und Neigebereich der Kuppelkamera gebildet wird, vgl. auch [WM96].

Projektion der Suchräume: Um die Bilder von den Kameras auf Regionen einzuschränken, in denen Objektmodellinstanzen und initial Objektmodelle erwartet werden, werden die 3D Einzelsuchräume $ssp_i \in SSP$ der Objektmodellteile in die Bilder projiziert. Die Projektion der 3D Suchräume in 2D Regionen im Bild erfolgt anhand der Projektionsgleichungen 2.13 - 2.18.

Für die kugelförmigen Suchräume, die für die primären Merkmale in STABIL^{++} gebildet werden, ergibt sich eine einfache Projektion: Es muß für die entsprechende kreisförmige Region der Mittelpunkt und der Radius bestimmt werden, vgl. Abb. 3.7 (a). Hierzu wird der Mittelpunkt $\vec{p}^{\,0}_{wcs}$ des Suchraumes in das Kamerakoordinatensystem cam transformiert und ein weiterer Punkt $\vec{p}^{\,r}_{cam}$ im Abstand des Radius r_{3D} vom Mittelpunkt z.B. entsprechend:

$$\begin{aligned} \vec{p}^{\,r}_{cam} &= \vec{p}^{\,0}_{cam} + r_{(3d)} \cdot [1,0,0]^T \\ \text{mit} & \\ \vec{p}^{\,0}_{cam} &= {}^{cam}\mathbf{T}_{wcs} \cdot \vec{p}^{\,0}_{wcs} \end{aligned}$$

gebildet.[10] Der Radius r_{2D} der kreisförmigen Region ergibt sich aus der euklid'schen Distanz zwischen den Projektionen von $\vec{p}^{\,0}_{cam}$ und $\vec{p}^{\,r}_{cam}$.

Für die sekundären Merkmale werden die Suchräume entsprechend der geometrischen Primitiven der Objektmodellteile gebildet. Die kugelförmigen Suchräume für die elliptischen Körper vol_{ell} werden wie vorstehend angegeben projiziert. Für die Volumenkörper vol_{trCone}

[9] Es ist jedem Bildpunkt ein Filter und somit ein Hintergrundbildpunkt zugeordnet. Ändert sich der abgebildete Bildausschnitt aufgrund von Kamerabewegungen, so muß dieser von Veränderungen im Hintergund und von Veränderungen durch Objekte im Vordergrund unterschieden werden können.

[10] Hier ist ein Punkt auf dem Kugelrand gewählt worden, der in der Projektion (im Bild) rechts von dem Mittelpunkt zu liegen kommt, vgl. auch Abb. 3.7.

müssen zylindrische Suchräume projiziert werden. Diese Suchräume werden, im Hinblick auf eine algorithmisch einfache Projektion, durch zwei senkrecht zueinander stehende Flächen angenähert. Die Schnittlinie zwischen den beiden Flächen verläuft auf der Rotationsachse des zylindrischen Suchraumes, vgl. Abb. 3.7.

Es werden von jeder Fläche jeweils die vier Eckpunkte ins Bild projiziert. Die projizierten Punkte begrenzen eine viereckige Region. Beide einzelne Regionen werden schließlich zu einer gemeinsamen Region vereinigt, die dann die angenäherte Projektion des zylindrischen Suchraumes repräsentiert. Entsprechend der gewählten Approximation des Zylinders durch zwei Flächen, kann die Größe der projizierten Region je nach Orientierung der Flächen im Raum variieren. Läuft die optische Achse (Z_{cam}) parallel zu dem Normalenvektor auf einer der beiden Flächen, so hat die projizierte Region eine Breite entsprechend der Zylinderbreite. Dreht man nun die beiden Flächen um die Rotationsachse des Zylinders, so daß die optische Achse einen Winkel $\neq 0$ mit dem Normalenvektor bildet, so verkleinert sich die Breite der Region. Bei einem Winkel von $45°$ erreicht die Breite nur noch $\frac{1}{\sqrt{2}}$ der maximalen Breite. Entsprechend dieser Variabilität der projizierten Suchräume ist bei der Bestimmung der Suchraumgröße für die sekundären Merkmale ein größerer Durchmesser zu berücksichtigen. Es wird ebenso vernachlässigt, daß bei einer Blickrichtung der optischen Achse, die auf der Rotationsachse des Zylinders verläuft, die Projektionen der Flächen zu Linien werden. Dieser Fall kann bei der Verwendung der zylindrischen Suchräume für die sekundären Merkmale in STABIL^{++} vernachlässigt werden, da jeweils in Richtung der Rotationsachse der Suchraum des Vorgänger- oder Nachfolgeobjektmodellteils zu liegen kommt. Durch diese Suchräume wird dann die entsprechende Bildregion in $REG^{(ssp)}$ aufgenommen.

Bei der vorgestellten Projektion der kugelförmigen und zylindrischen Suchräume wird jeweils nur für die projizierten Punkte die 3D / 2D Abbildung konsequent entsprechend des Kameramodells durchgeführt. Die Erzeugung eines Kreises als Projektion der Kugel und die Erzeugung von Polygonen aus den projizierten Eckpunkten der beiden Flächen vernachlässigt die im Kameramodell berücksichtigte Verzerrung. Nachdem die zu projizierenden Körper schon eine Annäherung sind, kann, solang keine extrem weitwinkligen Objektive mit extremer Verzerrung verwendet werden, der durch die Vernachlässigung der Verzerrung entstehende Fehler in Kauf genommen werden.

Adaptive Farbklassifikation:[11] Wie Hafner in seiner Arbeit zu dem in STABIL^{++} verwendeten adaptiven Farbklassifikator [Haf99] angemerkt hat, hat Farbe als Objektmerkmal[12] den entscheidenden Vorteil einer vergleichsweise hohen Robustheit bei Änderungen von Umgebungsparametern, so daß auch bei Beleuchtungsänderungen eine stabile Objektverfolgung möglich ist. Daher läßt sich Farbe ideal als Attribut der Modellmerkmale zur Objektdetektion und -verfolgung in natürlicher Umgebung verwenden.[13]

Gegenüber den üblichen Segmentierungsalgorithmen, bei denen eine rein binäre Zuordnung von Bildpunkten zu Regionen im Bild vorgenommen wird, wird bei der Farbklassifikation eine Zuordnung der Bildpunkte zu Klassen in einem Merkmalsraum durchgeführt. Basierend auf der Annahme, daß Farben in kameragenerierten Bildern einer mehrdimensionalen Normalverteilung folgen, wird eine stochastische Klassifikation der Farben im sog. $I_1I_2I_3$-Farbraum nach Ohta [OKS80] vorgenommen, der sich aus dem *RGB*-Farbraum (*Rot-Grün-Blau*) wie folgt

[11] Vgl. zu diesem Abschnitt auch [HKM95], [HM96], [HM97] und die Ausführungen in [RMR$^+$99].

[12] Farbe ist ein Basisattribut der in STABIL^{++} eingeführten Modell-, Szenen- und Bildmerkmale.

[13] Vgl. Anwendung zur Personendetektion in Kap. 4.2, bei der dem Objektmodellteil des Kopfes eine "hautfarbene" 3D Ellipse zugeordnet ist und Kap. 4.3.1 zur Anwendung der Bewegungserfassung, bei der die Gelenke mit farbigen Bändern markiert sind.

transformieren läßt:

$$\begin{pmatrix} I_1 \\ I_2 \\ I_3 \end{pmatrix} = \begin{pmatrix} \frac{R+G+B}{3} \\ \frac{R-B}{2} \\ \frac{2G-R-B}{4} \end{pmatrix}$$

Mit dem $I_1I_2I_3$-Farbraum steht ein Farbraum zur Verfügung, in dem sich kompakte Farbklassen bilden, die sich somit separieren lassen. Ferner treten keine Singularitäten wie in den bekannten HSV (*hue saturation value*) und HSI (*hue saturation intensity*) Farbräumen auf. Der $I_1I_2I_3$-Farbraum ist zudem, im Gegensatz zu den *CIE* Farbräumen (*commission internationale de l'eclaire*), ein bezüglich einer effizienten Transformation linearer Farbraum. Weiterhin wird in dem ersten Kanal I_1 die Helligkeit kodiert, so daß die Farbinformation in den Kanälen I_2 und I_3 repräsentiert wird. Hiermit ist eine Datenreduktion erreicht worden, so daß die Farbklassen als 2D Normalverteilungen repräsentiert werden können. Entsprechend ist eine Farbklasse Ω_k mit einem zweidimensionalen Merkmalsvektor $\vec{c}$ durch die zweidimensionale Gauß'sche Dichtefunktion:

$$p(\vec{c}\,|\Omega_k) = p(\vec{c}\,|\vec{\mu}_k; K_k) = \frac{1}{2\pi\sqrt{\det K_k}}\, e^{-\frac{1}{2}\,(\vec{c}-\vec{\mu}_k)'\,K_k^{-1}\,(\vec{c}-\vec{\mu}_k)}$$

bestimmt. Hierbei ist $\vec{\mu}_k$ der Mittelwert der Farbklasse (Mittelpunkt der Farbklasse im I_2I_3-Raum) und K_k die Kovarianzmatrix, mit der die Ausdehnung und die Orientierung der Klasse Ω_k im I_2I_3-Raum angegeben ist. Je größer der Wert der Dichte $p(\vec{c}|\Omega_k)$ für einen Merkmalsvektor $\vec{c}$ ist, um so größer ist die Wahrscheinlichkeit, daß der Farbwert den $\vec{c}$ repräsentiert, der Klasse Ω_k angehört.

Die Segmentierung der Farbklassen basiert auf der Bayes'schen Entscheidungstheorie. Hierzu wird die Segmentierung von n Farben in n Farbklassen $\Omega_\lambda, \lambda = 1, \ldots, n$ und einer zusätzlichen Rückweisungsklasse Ω_0 anhand der Entscheidungsregel $\delta(\Omega_\lambda|\vec{c})$ durchgeführt. Aufgrund einer Vereinfachung der Gauß'schen Dichtefunktion $p(\vec{c}|\Omega_k)$ durch:

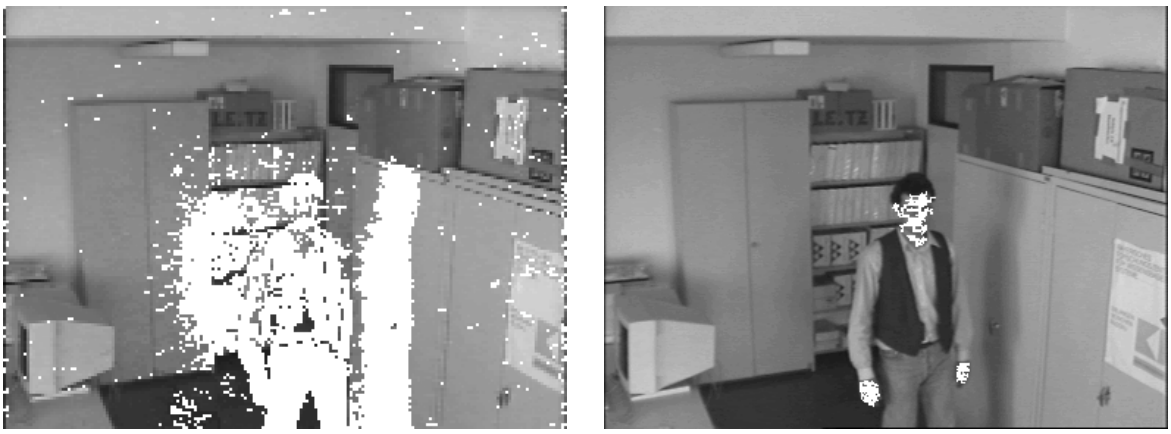
$$d(\vec{c}\,|\Omega_k) = \ln(\det K_k) + (\vec{c}-\vec{\mu}_k)'\,K_k^{-1}\,(\vec{c}-\vec{\mu}_k)$$

ergibt sich für die Entscheidungsregel $\delta(\Omega_\lambda|\vec{c})$:

$$\begin{array}{lcl} p_k\, d(\vec{c}\,|\Omega_k) & = & \max_{\lambda=1\ldots n}\, p_\lambda\, d(\vec{c}\,|\Omega_\lambda) \\ \text{mit} & & \\ \delta(\Omega_k|\vec{c}) & = & 1, \quad \text{falls} \quad p_k\, d(\vec{c}\,|\Omega_k) \geq \beta \cdot \sum_{j=1}^{n} p_j\, d(\vec{c}\,|\Omega_j) \\ \delta(\Omega_0|\vec{c}) & = & 1, \quad \text{sonst} \end{array}$$

Hierbei ist $p_\lambda\, d(\vec{c}\,|\Omega_\lambda), \lambda = 1, \ldots, n$ die a-priori Wahrscheinlichkeit jeweils der n Farbklassen.

Diese Entscheidung wird für jeden Bildpunkt durchgeführt; hierbei wird zunächst entsprechend der Farbraumtransformation der entsprechende Merkmalsvektor $\vec{c}$ gebildet. Für $\vec{c}$ wird dann die maximale gewichtete Klassenwahrscheinlichkeit $p_k\, d(\vec{c}\,|\Omega_k)$ bestimmt, d.h. es wird die Klasse bestimmt, in der $\vec{c}$ am wahrscheinlichsten liegt. Entsprechend der Bayes'schen Kostenfunktionen werden noch die Kosten der Falschklassifikation minimiert, indem in Abhängigkeit der Wahrscheinlichkeit entschieden wird, ob $\vec{c}$ der Klasse Ω_k oder der Rückweisungsklasse Ω_0 zugeordnet wird. Hierbei ist $\beta \in [0, 1]$ ein Rückweisungsparameter, mit dem der Grad der Rückweisung bestimmt wird. Je größer β wird, um so mehr Merkmalsvektoren werden, falls diese nur mit geringer Wahrscheinlichkeit einer Klasse zugeordnet werden können, zurückgewiesen.



(a) *(b)*

Abbildung 3.8: *(a) Anwendung der Vorder- / Hintergrundsegmentierung, und (b) Anwendung der Farbklassifikation für die Farbklasse "Hautfarbe"; die weißen Regionen $REG^{(fg)}$ und $REG^{(skincolored)}$ sind dem, hier als Grauwertbild dargestellten, Originalfarbbild überlagert.*

Damit die Bildpunkte gleicher Farbe auch bei sich verändernden Beleuchtungsverhältnissen der gleichen Farbklasse zugewiesen werden, stellt [Haf99] ein überwachtes Lernen zur Adaption der Klassenparameter vor. Hierbei wird ein iterativer Ansatz zum entscheidungsüberwachten Lernen verwendet, indem bei der Klassifikation in aufeinanderfolgenden Bildern einer Bildfolge die Veränderungen in den Mittelwerten $\vec{\mu}_k$, der Kovarianzmatrizen K_k, aber auch den Wahrscheinlichkeiten der Klassenzugehörigkeit für einen Merkmalsvektor $\vec{c}$ angepaßt werden. Hierbei verändert sich für die Klassen die Lage im I_2I_3-Raum, die Größe und die Form.

Welche Farben zu segmentieren sind, ergibt sich in STABIL^{++} aus den Attributen der verwendeten Modellmerkmale. Jedoch sind die Parameter der Farbklassen, wie der Mittelwert $\vec{\mu}_k$, die Kovarianzmatrix K_k und die a-priori Wahrscheinlichkeit p_k) von den einzelnen Kameras, der Signalübertragung und Digitalisierungskarte abhängig. Daher werden für jede Kamera in STABIL^{++} alle zu klassifizierenden Farbklassen initial einzeln angelernt.[14] Somit muß jeder Kamera $cam_i \in CAM$ ein Farbklassifikator als einer der Klassifikatoren $clc^{color} \in CLC_i$ zugewiesen werden, für den dann jeweils die Parameter der zu segmentierenden Farbklassen bekannt sein müssen. Jeder Kamera sind somit "Farben" $color_i \in COL$ bekannt. Die Farben in $COLOR$ variieren entsprechend der in den Anwendungen verwendeten primären Merkmale der Objektmodellteile.

Bei der Wahl der zu verwendenden Farben ist aufgrund einer robusten Separierbarkeit der Farbklassen darauf zu achten, daß die Farbklassen nach Möglichkeit im I_2I_3-Raum räumlich weit auseinander liegen. Hierauf hat die Farbtemperatur der Beleuchtung neben dem Weißabgleich der Kamera erheblichen Einfluß. Zudem ist darauf zu achten, daß die Szene gleichmäßig ausgeleuchtet ist und die Dynamik des Kamerasignals optimal ausgenutzt wird, denn bei Überstrahlung (zu hell) fallen die Farbklassen in der Mitte des I_2I_3-Raumes auf der sog. "Unbunt"-Geraden zusammen. Ist die Ausleuchtung zu gering oder die Blende zu weit geschlossen, dann wandern die Farbklassen im Farbraum zunächst nach außen und entarten; dies begründet sich

[14] Für die Trainingsphase werden in Testaufnahmen repräsentative Bildregionen manuell ausgewählt.

in der gewählten Farbraumtransformation. Denn ist aufgrund des niedrigen Signalpegels einer der Kanäle $= 0$, so kommt der transformierte Bildpunkt auf dem Rand des I_2I_3-Raumes zu liegen. Sinnvollerweise wird die Klassifikation erst ab einem bestimmten Signalpegel in allen drei Farbkanälen (*RGB*) zugelassen.[15]

In Abb. 3.8 (b) ist das Klassifikationsergebnis für die Farbe "Hautfarbe" dargestellt. Die Klassifikation ist nur auf den Bildbereichen $REG^{(fg)}$, die als Vordergrund segmentiert worden sind, durchgeführt worden, vgl. Abb. 3.8 (a). Durch diese Zusatzinformation wird zum einen die Klassifikation beschleunigt, zum anderen wird sie robuster. Denn "hautfarbene" Bildbereiche, die zum Hintergrund und somit nicht zu den zu detektierenden Objekten gehören werden nicht in die Farbklassifikation einbezogen. Für eine derartige Einschränkung des Segmentierungsbereiches eignet sich auch die Region des Suchbereiches $REG^{(ssp)}$.

3.4.3 Extraktion von Bildmerkmalen

Die eigentliche Extraktion der 2D Bildmerkmale $\mathbf{i}$ wird durch die Objektmodellteile selbst initiiert. Hierzu sind den primären Merkmalen $\mathbf{f}$ der Objektmodellteile, entsprechend den verschiedenen Attributen $attr^{\mathbf{m}}$ der zugehörigen Modellmerkmale $\mathbf{m}$, verschiedene Bildverarbeitungsoperatoren $ip(.) \in IP$ zugeordnet. Mit diesen Operatoren werden die 2D Bildmerkmale in den Videobildern entsprechend den zugehörigen Attributen $attr^{\mathbf{i}}$ extrahiert. Die Extraktion wird auf den, im Interpretationsschritt der Bildgenerierung eingezogenen und vorverarbeiteten Bildern, durchgeführt. Somit stehen hierzu die Segmentationsergebnisse in Form der Regionen $REG^{(fg)}$, $REG^{(ssp)}$ und $REG^{(color_1)} \ldots REG^{(color_n)}$ mit $color_k \in COLOR$ zur Verfügung. Vgl. hierzu auch die Zeilen 5 - 15 im Alg. 3.3 zur Detektion im Interpretationszyklus.

Aufgrund der in Kap. 4 beschriebenen Anwendungen beschränkt sich in den folgenden Abschnitten die Darstellung auf die Bildmerkmale der farbigen Ellipsen, kreisförmigen Landmarken und einer einfachen manuellen Markierung von Bildpunkten. Es ist in der Struktur von STABIL^{++} jedoch einfach möglich, die Anzahl der verschiedenen Bildmerkmale zu erhöhen. Es müssen hierzu die entsprechenden Bildverarbeitungsoperatoren $IP = \{ip_1(.), \ldots, ip_n(.)\}$ den primären Merkmalen der Objektmodellteile zugeordnet werden, wie in Kap. 2.4.2 erläutert.

Allen Bildmerkmalen ist gemeinsam, daß sie einen Bildpunkt $\vec{p}_{img}$ bestimmen, der die Position des Merkmals im Bild angibt. Für den im Interpretationsprozeß notwendigen Übergang von den 2D Bildmerkmalen zu 3D Szenenmerkmalen wird zu jedem Bildmerkmal ein 3D Sichtstrahl benötigt. Daher werden jedem Bildmerkmal zusätzlich zu $\vec{p}_{img}$ noch die Kameraparameter $camPar$ und $camPose$ mitgegeben. Die Kameraparameter werden von dem Bild img übernommen, auf dem die Bildverarbeitungsoperatoren $ip(img) \in IP$ zur Extraktion des Bildmerkmals ausgeführt werden.

Farbellipsen

Zu einem 3D Modellmerkmal eines farbigen Ellipsoiden muß in den Videobildern, entsprechend der Projektion ein 2D Bildmerkmal in Form einer Ellipse extrahiert werden. Diese Merkmale werden in den schon genannten Anwendungen zur Bestimmung des Objektmodellteiles des Kopfes als "hautfarbener" Ellipsoid und der mit farbigen Bändern markierten Gelenke ver-

[15] Für die Verwendung des Merkmals Farbe muß daher für ausreichende Beleuchtung gesorgt werden, denn sonst gilt für den Farbklassifikator "Nachts sind alle Katzen grau" – entsprechend des Sprichwortes.

wendet. Eine Farbellipse zeichnet sich zusätzlich zu dem Mittelpunkt $\vec{p}_{img}$ noch durch folgende Attribute aus:

- die Zugehörigkeit seiner Bildpunkte zu einer Farbklasse entsprechend einer Farbe $color_k$,
- der Größe der Fläche a seiner Region (Anzahl der Bildpunkte),
- zwei, die geometrische Form bestimmende, Halbmesser r_1, r_2 und
- einen Winkel φ, der die Orientierung der Ellipse angibt.

Diese Attribute werden bei der Extraktion bestimmt. Um die Zugehörigkeit zu einer Farbklasse zu gewährleisten, werden die weiteren Operatoren nur auf die Einzelregionen $reg_j^{color_k} \in REG^{color_k}$, die durch die Farbklassifikation aus der Bildvorverarbeitung bestimmt worden sind, angewendet. Das Basisattribut der farbigen Ellipsen ist die Zugehörigkeit zu einer Farbklasse. Alle weiteren Attribute sind Maßzahlen und lassen sich durch Vermessung der Ellipsen bestimmen, die durch geeignete Operatoren an die Regionen angepasst worden sind.

Landmarken

Neben den "natürlichen" Merkmalen eines Objektes können auch künstlich Markierungen vorgenommen werden. Man spricht dann von Landmarken. Hier kann man zum einen Farbmarkierungen verwenden, die dann im Videobild als Farbellipsen extrahiert werden. Zum anderen können schwarz-weiße kreisförmige Landmarken verwendet werden. Die Marken zeichnen sich dadurch aus, daß ein heller (weißer) Kreis von einem dunklen (schwarzer) Rand umgeben ist. Diese Art der Markierung beschränkt die möglichen Bewegungsrichtungen der Objekte, indem die Marken nur eine 2D Ausprägung haben und nahezu von vorne im Bild zu sehen sein müssen. Jedoch kann diese Art der Markierung auch auf Bildern mit nur einem Kanal, einem sog. Grauwertbild extrahiert werden und haben daher in bestimmten Anwendungen durchaus ihre Berechtigung.

Die kreisförmige Landmarke als Modellmerkmal muß im Videobild als helle kreisförmige Region extrahiert werden. Die kreisförmige Region zeichnet sich zusätzlich zu dem Mittelpunkt $\vec{p}_{img}$ durch folgende Attribute aus:

- "ist eine helle Region mit dunklem Rand",
- die Größe der Fläche a seiner Region / Anzahl der Bildpunkte,
- der Radius r der Region,
- einem Maß der Kompaktheit,
- einem Maß der Anisometrie und
- einem Maß der Zirkularität

aus.

Zur Bestimmung der Regionen wird auf dem Kanal can_1 des Grauwertbildes img ein dynamisches Schwellenwertverfahren zur Segmentierung angewendet. Das Schwellenwertverfahren verwendet als Referenz zur Bestimmung der Schwellen das geglättete tiefpassgefilterte Bild. Hierzu wird eine Glättungsmaske verwendet, die ungefähr der Größe der zu erwartenden Regionen entsprechen sollte. Das Basisattribut, das zur Gruppierung der Merkmale verwendet wird ist die Beschreibung "ist eine helle Region mit dunklem Rand", alle weiteren Attribute werden als Formmerkmale auf den so erhaltenen Regionen bestimmt.

Manuelle Auswahl

Als ein weiteres Merkmal kann zu Testzwecken oder zur manuellen Korrektur der automatischen Interpretation die Auswahl bestimmter Bildpunkte als Bildmerkmal verwendet werden. Hierbei wird mit dem Zeige- u. Eingabegerät[16] am Bildschirm im dargestellten Videobild ein Bildpunkt markiert. Das so “eingegebene” Bildmerkmal zeichnet sich nur durch seine Position $\vec{p}_{img}$ und durch das Basisattribut “ist eine manuelle Auswahl” aus. Es sind keine Maßzahlen als weitere Attribute zu bestimmen.

[16] Computer-Maus.

3.5 2D / 3D Übergang

Nach dem Schritt der Extraktion der 2D Bildmerkmale aus den Videobildern werden in STABIL^{++} im Interpretationszyklus entsprechend des Alg. 3.3 die 3D Szenenmerkmale bestimmt, um diese dann bei der Generierung der Hypothesen den 3D Modellmerkmalen zuzuordnen. Der 2D / 3D Übergang ist in Zeile 16 jedoch nur möglich, wenn hierzu kein Wissen des Objektmodells verwendet werden muß. Ist dies der Fall, so werden die 3D Szenenmerkmale erst bei der Zuordnung zu den Modellmerkmalen bestimmt. Hiermit unterscheiden sich zwei verschiedene in STABIL^{++} realisierte Ansätze zum 2D / 3D Übergang.

Dies ist zum einen ein monokularer Ansatz (Mono-Ansatz) und ein binokularer Ansatz (Stereoansatz). Diese beiden Ansätze werden in den folgenden Abschnitten erläutert, wobei auf die Definition der 2D Bildmerkmale und 3D Szenenmerkmale aus Kap. 2.4 zurückgegriffen wird. Zuvor wird noch erläutert, wie in STABIL^{++} zwischen beiden Ansätzen in einem Interpretationszyklus unterschieden wird und wie diese kombiniert verwendet werden können.

3.5.1 Unterscheidung zwischen Mono und Stereo

Sind 2D Bildmerkmale eines zugehörigen 3D Szenenmerkmals in mindestens zwei Kamerabildern extrahiert worden, so wird ein (Mehrfach-) Stereoansatz verwendet und die 3D Position des Szenenmerkmals wird durch den Schnittpunkt von zwei (oder mehr) Sichtstrahlen bestimmt. Dieser Schritt kann somit auch direkt nach der Extraktion der Bildmerkmale durchgeführt werden, so daß in Zeile 16 des Alg. 3.3 3D Punkte erzeugt werden.

Steht jedoch nur ein 2D Bildmerkmal aus einem Kamerabild zur Verfügung, so muß die 3D Position des zugehörigen Szenenmerkmals in dem sog. Mono-Ansatz geschätzt werden. Für diese Schätzung wird Modellwissen benötigt. Auf Modellwissen kann jedoch nur über ein 3D Modellmerkmal, das einem Objektmodellteil zugeordnet ist, zugegriffen werden. Eine Zuordnung von 3D Szenenmerkmalen und 3D Modellmerkmalen erfolgt erst bei der Generierung der Hypothesen. Daher wird der 2D / 3D Übergang, falls der Mono-Ansatz verwendet werden muß, während des Schrittes zur Generierung der Hypothesen ausgeführt (Zeile 19 im Alg. 3.3).

Mono- und Stereoansatz zum 2D / 3D Übergang sind in STABIL^{++} jedoch auch innerhalb eines Interpretationszyklus kombinierbar. Hierzu wird zunächst für alle 2D Bildmerkmale $\mathbf{i} \in \mathbf{I}$ versucht, mit dem Stereoansatz in $\mathbf{I}$ Paare (Tripel, Quadrupel, ...) von Bildmerkmalen zu finden, die in unterschiedlichen Kamerabildern den gleichen 3D Punkt abbilden. Für diese Paare (Tripel, Quadrupel, ...) wird jeweils ein 3D Szenenmerkmal $\mathbf{s}$ erzeugt. Für $\mathbf{s}$ ist die 3D Position mit $\vec{p}_{wcs}$ gesetzt und in $\mathbf{I}_{extr}$ sind die verwendeten 2D Bildmerkmale vermerkt worden.

Für alle 2D Bildmerkmale $\mathbf{i}_j \in \mathbf{I}$, die keinem 3D Szenenmerkmal $\mathbf{s}_k \in \mathbf{S}$ zugeordnet werden konnten, wird ein sog. *Pseudo*szenenmerkmal erzeugt. Diese Pseudoszenenmerkmale zeichnen sich dadurch aus, daß in $\mathbf{I}_{extr}$ nur ein, nicht mit dem Stereoansatz zugeordnetes Bildmerkmal, vermerkt ist, somit ist $|\mathbf{I}_{extr}| = 1$. Entsprechend ist für ein Pseudoszenenmerkmal noch keine gültige 3D Position in $\vec{p}_{wcs}$ gesetzt.

Werden bei der Generierung der Hypothesen die Szenenmerkmale $\mathbf{s}_k \in \mathbf{S}$ den 3D Modellmerkmalen zugeordnet, können diese 3D Szenenmerkmale direkt verwendet werden. Bei der Zuordnung von Pseudoszenenmerkmalen muß erst noch die 3D Position mit dem Mono-Ansatz bestimmt werden. Jedoch kann auf das hierzu notwendige 3D Modellwissen über das 3D Modellmerkmal zugegriffen werden, zu dem ein Pseudoszenenmerkmal zugeordnet werden soll.

Durch diese Unterscheidung von Szenenmerkmalen mit gesetzter 3D Position und Pseudoszenenmerkmalen ist es in STABIL^{++} möglich, die beiden Ansätze in einem Interpretationszyklus zu kombinieren. Dies macht jedoch nur Sinn, wenn die Schätzung im Mono-Ansatz

auf gesichertes 3D Wissen des Modells zurückgreifen kann. Ist dies anwendungsbedingt nicht möglich, so läßt sich die Schätzung mit dem Mono-Ansatz unterbinden. Damit werden alle Pseudoszenenmerkmale, und somit die in diesen vermerkten Bildmerkmale, nicht zur Generierung von Hypothesen verwendet. Im Gegensatz dazu ist es auch möglich, daß nur Pseudoszenenmerkmale zur Verfügung stehen, da entweder nur eine Kamera verwendet wird oder Teile des Suchbereiches SSP nicht in mindestens zwei Kameras sichtbar ist.

Zwischen dem Mono- und Stereoansatz wird auch noch bei der für die 3D Szenenmerkmale zu setzende Qualität q unterschieden. Mit diesem Qualitätsmaß wird ausgedrückt, wie sicher das 3D Merkmal detektiert worden ist und somit auch, wie sicher das zugehörige Objektmodell detektiert wurde. Daher wird q für die Bewertung der Hypothesen verwendet, wodurch dann einer Hypothese der Vorrang gegeben werden soll, die auf den genaueren Szenenmerkmalen aus dem Stereoansatz basiert ($q = 1$). Die Güte eines Szenenmerkmals, das mit dem Mono-Ansatz bestimmt worden ist, muß zwischen dem Maß für den Stereoansatz und dem Maß einer Schätzung liegen. Hierdurch kann bei der Auswahl der Hypothesen unterschieden werden. Es wird dort einer Hypothese, die (teilweise) auf Szenenmerkmalen basiert, die durch den Mono-Ansatz bestimmt worden sind, der Vorrang vor einer Hypothese eingeräumt, bei der die Szenenmerkmale (teilweise) auf Schätzungen durch die Bewegungsvorhersage beruhen. Entsprechend der zur Bestimmung der Qualität der Vorhersagen verwendeten Alterungsfunktion in Glg. 3.3 und einer anzunehmenden maximalen Bildwiederholrate von 25 Bildern / Sekunde (PAL-Norm) wird das Qualitätsmaß bei dem Mono-Ansatz auf $q = 0,975$ festgelegt.

3.5.2 Monokularer Ansatz

Der monokulare Ansatz (Mono-Ansatz) zur Bestimmung eines 3D Szenenmerkmals $\mathbf{s}$ aus einem 2D Bildmerkmal $\mathbf{i}$ stützt sich auf Heuristiken, die sich aus dem Modellwissen ergeben und kann daher erst durchgeführt werden, wenn bei der Generierung der Hypothesen versucht wird, einem Pseudoszenenmerkmal ein Modellmerkmal zuzuordnen. Durch den mit $\mathbf{i} \in \mathbf{I}_{extr}$ festgelegten Bildpunkt $\vec{p}_{img}$ und den dem Bildpunkt bekannten Kameraparametern $camPar$ und $camPose$ läßt sich ein Sichtstrahl $\vec{s}$ bestimmen. Dieser Sichtstrahl verläuft durch $\vec{p}_{img}$ in der Bildebene der Kamera und dem Ursprung des Kamerakoordinatensystems $[X_{cam}, Y_{cam}, Z_{cam}]$ und ist entsprechend den Glg. 2.19 - 2.21 bestimmt. Um von dem Sichtstrahl den 3D Punkt des Szenenmerkmals zu erhalten, wird die Länge des Sichtstrahls ermittelt, bei der dieser auf das Objektmodellteil oder dessen Merkmale trifft. Daher spricht man auch von dem Problem der Tiefenschätzung.

Für die in STABIL^{++} verwendeten Punktmerkmale gibt es hierzu mehrere Ansätze. Zum einen besteht die Möglichkeit, über eine fest vorgegebene Höhe des Objektmodellteiles oder der Merkmale den Sichtstrahl in der entsprechenden Höhe zu schneiden. Eine weitere Möglichkeit besteht darin, die Ausdehnung eines Bildmerkmals zu nutzen und zum mittleren Sichtstrahl noch weitere Sichtstrahlen an den Rändern der Bildregion zu verwenden. Dabei ergibt sich die Tiefe durch den Abstand, bei der die Größe des zugehörigen Modellmerkmals in das Bündel aus Sichtstrahlen paßt. In den folgenden Abschnitten werden die beiden verschiedenen Ansätze näher erläutert.

Es empfiehlt sich, je nach Anwendung auch mehrere Tiefenschätzungen zu kombinieren und aus den Ergebnissen einen Mittelwert zu bilden. Stehen mehr als zwei Ergebnisse zur Verfügung, kann zusätzlich anhand der Varianz eine Gewichtung vorgenommen und Ausreißer kontrolliert werden. Für die so ermittelten 3D Punkte ist anschließend noch zu prüfen, ob diese noch innerhalb des Sichtbereiches der Kamera zu liegen kommen. Zusätzlich wird für alle Punkte, die mit der Tiefenschätzung ermittelt wurden, überprüft, ob diese innerhalb des Obser-

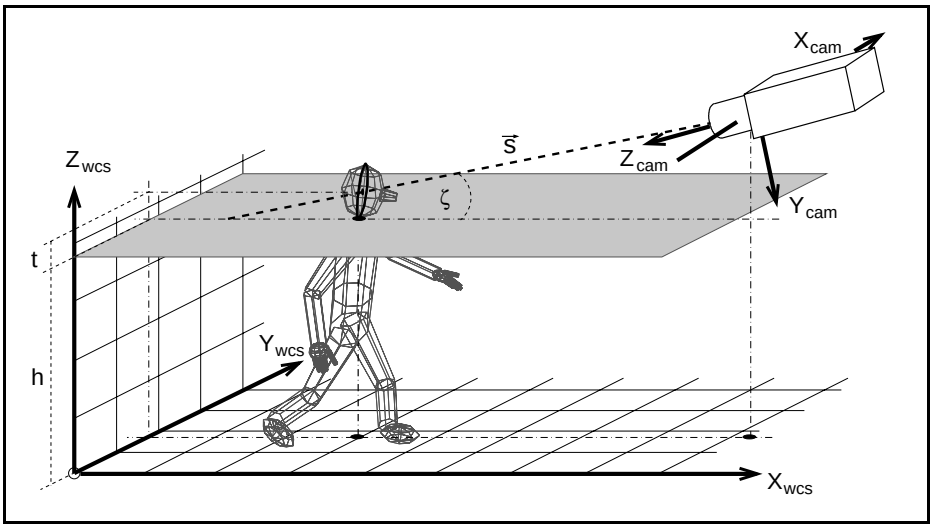


Abbildung 3.9: *Mono-Ansatz: Tiefenschätzung über die Höhe des Objektmodellteiles des Kopfes.*

vierungsraumes SSP_s des Szenenmodells liegen. Damit muß getestet werden, ob zwischen der Kamera und dem ermittelten 3D Punkt ein Sichthindernis liegt, vgl. die hierzu analoge Vorgehensweise bei der Bestimmung der Sichtbarkeit der Suchräume in Kap. 3.3.4.

Tiefenschätzung über die Höhe des Objektmodellteiles

Kann aufgrund der Anwendung davon ausgegangen werden, daß das Objektmodellteil in seiner möglichen Position im Raum beschränkt ist, so kann diese Heuristik zur Bestimmung der Tiefeninformation genutzt werden. Eine mögliche Heuristik ist eine fixe Höhe des Objektmodellteiles über der xy-Ebene des Weltkoordinatensystems.[1] Man geht also davon aus, daß der Ursprung des lokalen Koordinatensystems des Objektmodellteiles sich für den aktuellen Interpretationszyklus nur in dieser Ebene befinden kann. Die Höhe muß über das Objektmodellteil ermittelt werden, zu dessen Modellmerkmal $\mathbf{m}$ das aus dem Bildmerkmal $\mathbf{i}$ zu generierende Szenenmerkmal $\mathbf{s}$ zugeordnet werden soll. Hierzu wird, entsprechend der geometrischen Modellstruktur der Objektmodellinstanz, die Lage des Ursprungs des lokalen Koordinatensystems im Weltkoordinatensystem ermittelt.

In Abb. 3.9 ist z.B. die Detektion einer Person dargestellt. Hierbei ist für das Objektmodellteil $omp_{3.3}$ des Kopfes als Modellmerkmal eine 3D Ellipse verwendet worden. Diese Ellipse liegt in der xz-Ebene des lokalen Koordinatensystems des Objektmodellteils. Der Mittelpunkt der Ellipse liegt auf der z-Achse, wobei dieser um den Translationsvektor $\vec{t}$ des Modellmerkmals nach oben verschoben ist. Entsprechend den Basisattributen der Merkmale ist bei der Extraktion ein entsprechendes Bildmerkmal $\mathbf{i}$ ermittelt worden, so daß ein Sichtstrahl durch den Mittelpunkt der extrahierten Ellipse verläuft. Um die 3D Position der Ellipse zu ermitteln, muß der Abstand der Person zur Kamera bestimmt werden. Man erhält die Position ${}^{\circ}\vec{p}_{wcs}^{\,3.3}$ des Objektmodellteiles $omp_{3.3}$:

$$ {}^{\circ}\vec{p}_{wcs}^{\,3.3} = {}^{wcs}\,\mathbf{T}_{omp_{3.3}} \cdot [0,0,0]^T \tag{3.9} $$

[1] Es können hier auch beliebige andere Ebenen im Weltkoordinatensystem festgelegt werden, auf denen sich das Objektmodellteil bewegen kann.

Bei der Re-Detektion einer Objektmodellinstanz ist ${}^{\circ}\vec{p}_{wcs}^{\,3.3}$ die Position des lokalen Koordinatensystems für den Kopf aus dem letzten Interpretationszyklus. Wird jedoch ein initiales Objektmodell detektiert, so ist ${}^{\circ}\vec{p}_{wcs}^{\,3.3}$ durch die voreingestellte Größe des Modells und die Gelenkwinkel bestimmt. Zur Bestimmung des Mittelpunktes der 3D Ellipse des Objektmodellteils $omp_{3.3}$ wird ${}^{\circ}\vec{p}_{wcs}^{\,3.3}$ noch um den Translationsvektor $\vec{t}$ des Modellmerkmals korrigiert, man erhält:

$$ {}^{\circ}\vec{p}'^{\,3.3}_{wcs} = {}^{\circ}\vec{p}_{wcs}^{\,3.3} + {}^{wcs}\,\mathbf{T}_{omp_{2.3}} \cdot \vec{t} \tag{3.10} $$

Hierbei ist das Objektmodellteil $omp_{2.3}$ das Vorgängerobjektmodellteil entsprechend der verwendeten Standardmodellierung, vgl. Tab. 2.1. Es wird daher die Korrektur der 3D Position mit der Orientierung des lokalen Koordinatensystems des Objektmodellteiles des Halses vorgenommen.[2]

Der Punkt $\vec{p}_{wcs}$ des zu ermittelnden Szenenmerkmals ist somit durch den Schnittpunkt des Sichtstrahls $\vec{s}$ mit einer Ebene, die im Weltkoordinatensystem definiert ist, bestimmt. Es gilt:[3]

$$ \vec{p}_{wcs} = {}^{\circ}\vec{p}_{wcs}^{\,cam} + \frac{\vec{n}_{wcs} \cdot (\vec{a}_{wcs} - {}^{\circ}\vec{p}_{wcs}^{\,cam})}{\vec{n}_{wcs} \cdot \vec{s}_{wcs}} \cdot \vec{s}_{wcs} \tag{3.11} $$

Hierbei ist ${}^{\circ}\vec{p}_{wcs}^{\,cam} = {}^{wcs}\mathbf{T}_{cam} \cdot [0,0,0]^T$ der Ursprung des Kamerakoordinatensystems. Die zu schneidende Ebene wird durch einen Punkt $\vec{a}_{wcs}$ in der Ebene und durch den Normalenvektor $\vec{n}_{wcs}$ der Ebene angegeben. Für das in Abb. 3.9 gezeigte Beispiel gilt dann:

$$ \begin{aligned} \vec{a}_{wcs} &= [0,0,h']^T \\ \vec{n}_{wcs} &= [0,0,h']^T \\ \text{mit} & \\ h' &= [0,0,1] \cdot {}^{\circ}\vec{p}'^{\,3.3}_{wcs} \end{aligned} $$

Damit reduziert sich die Glg. 3.11 zu:

$$ \begin{aligned} \vec{p}_{wcs} &= {}^{\circ}\vec{p}_{wcs}^{\,cam} + \frac{h' - z_{cam}}{z_s} \cdot \vec{s}_{wcs} \\ \text{mit} & \\ {}^{\circ}\vec{p}_{wcs}^{\,cam} &= \begin{pmatrix} x_{cam} \\ y_{cam} \\ z_{cam} \end{pmatrix} \\ \vec{s}_{wcs} &= \begin{pmatrix} x_s \\ y_s \\ z_s \end{pmatrix} \end{aligned} $$

Die Genauigkeit dieser Tiefenschätzung hängt zum einen davon ab, ob das der Heuristik zugrunde gelegte Wissen zutrifft, zum anderen selbstverständlich von der Güte der Kamerakalibrierung. Ein weiterer Faktor ist der Winkel ζ , in dem der Sichtstrahl auf die Ebene trifft. Für das genannte Beispiel hängt ζ vom Neigewinkel θ der Kamera ab. Je kleiner ζ wird, um so geringer wird die Genauigkeit, mit der die Tiefe geschätzt werden kann. Für $\zeta = 0$ ist dann keine Aussage über die Tiefe mehr möglich.

Tiefenschätzung über die Größe des Objektmodellteiles

Sind für das Bildmerkmal Attribute seiner Form und Geometrie bestimmt worden,[4] dann können zusätzlich zu einem Sichtstrahl $\vec{s}_0$ im Mittelpunkt des Punktmerkmales noch Sichtstrah-

[2] Vgl. . hierzu auch die Verwendung von $\vec{t}$ bei der Bestimmung der Suchräume in Kap. 3.2.4.

[3] Anm.: Das eigentliche Ergebnis der Tiefenschätzung ist die Länge des Sichtstrahls und somit der Abstand von $\vec{p}_{wcs}$ zum Ursprung des Kamerakoordinatensystems: $||{}^{cam}\mathbf{T}_{wcs} \cdot \vec{p}_{wcs}||$.

[4] Z.B. r_1, r_2 und φ oder r bei den in Kap. 3.4.3 beschriebenen Bildmerkmalen der Farbellipsen oder Landmarken.

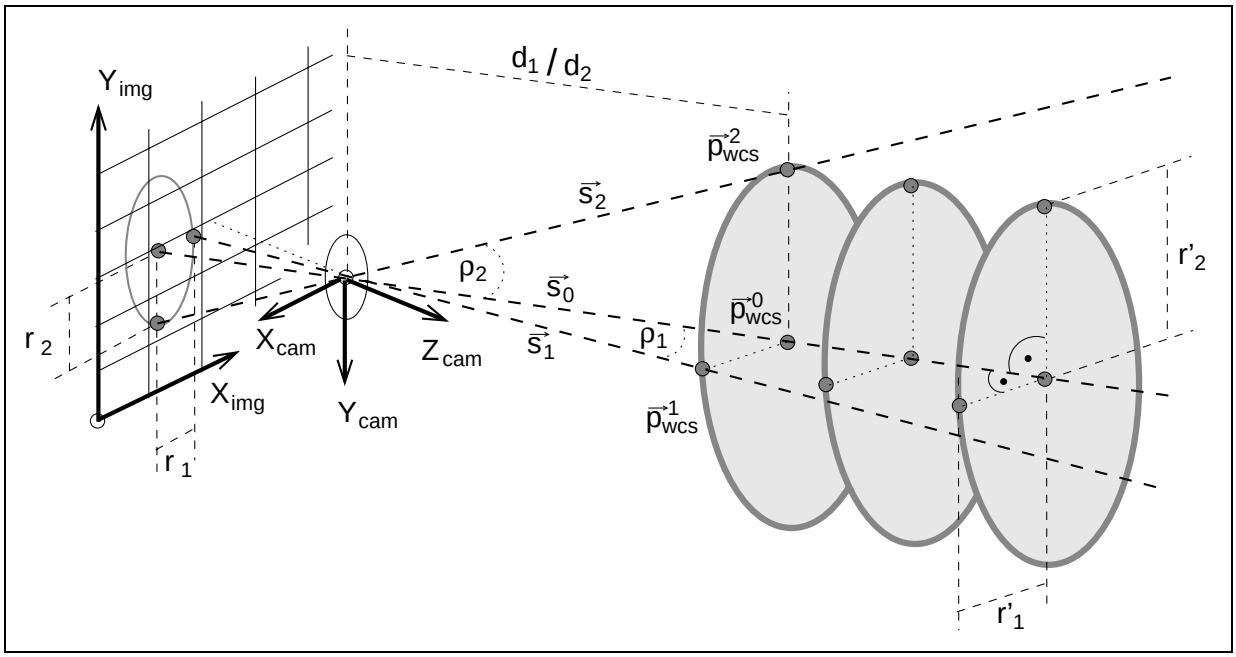


Abbildung 3.10: *Monokularer Ansatz: Tiefenschätzung über die Größe eines Objektmodellteiles; das Bildkoordinatensystem $[X_{img}, Y_{img}]$ und das Kamerakoordinatensystem $[X_{cam}, Y_{cam}]$ sind aufgrund des Kameramodells spiegelverkehrt orientiert.*

len an den Rändern der Fläche bestimmt werden. In Abb. 3.10 ist dies für eine Ellipse dargestellt. Es wird dort ein weiterer Sichtstrahl $\vec{s}_1$ durch einen Punkt in der Bildebene bestimmt, der im Abstand des Halbmessers r_1 und unter Berücksichtigung des Winkels φ, der die Lage der Ellipse angibt, rechts vom Mittelpunkt der Ellipse liegt.[5] Ein dritter Sichtstrahl $\vec{s}_2$ wird durch einen Punkt bestimmt, der im Abstand des zweiten Halbmessers r_2 auf dem Rand der Ellipse oberhalb des Mittelpunktes liegt.

Auf der anderen Seite ergibt sich aus dem zuzuordnenden Modellmerkmal $\mathbf{m}$ eine 3D Ellipse mit den beiden Halbmessern r'_1 und r'_2. Für das explizite Modellmerkmal der "hautfarbenen" 3D Ellipse für das Objektmodellteil des Kopfes sind diese die halbe Breite (x-Achsen Ausdehnung) des zugehörigen Volumenkörpers vol_{ell} und die halbe Höhe (z-Achsen Ausdehnung im lokalen Koordinatensystem). Man kann sich vorstellen, daß zur Bestimmung der Tiefeninformation diese 3D Ellipse solange entlang des mittleren Sichtstrahls $\vec{s}_0$ verschoben wird, bis die Sichtstrahlen $\vec{s}_1$ und $\vec{s}_2$ auf den Rand der Ellipse treffen.

Nachdem die Exzentrizität der extrahierten Ellipse im Bild und der 3D Ellipse nicht zwingend gleich sein müssen, wird für jedes Sichtstrahlpaar ($\vec{s}_0$ / $\vec{s}_1$ und $\vec{s}_0$ / $\vec{s}_2$) die Tiefeninformation getrennt bestimmt. Die Tiefeninformationen können dann wie o.a. gemittelt werden. Für das Sichtstrahlpaar $\vec{s}_0$ / $\vec{s}_1$ ergibt sich die Tiefeninformation d_1 als Länge des mittleren Sichtstrahls

[5] In Abb. 3.10 ist $\varphi = 0$.

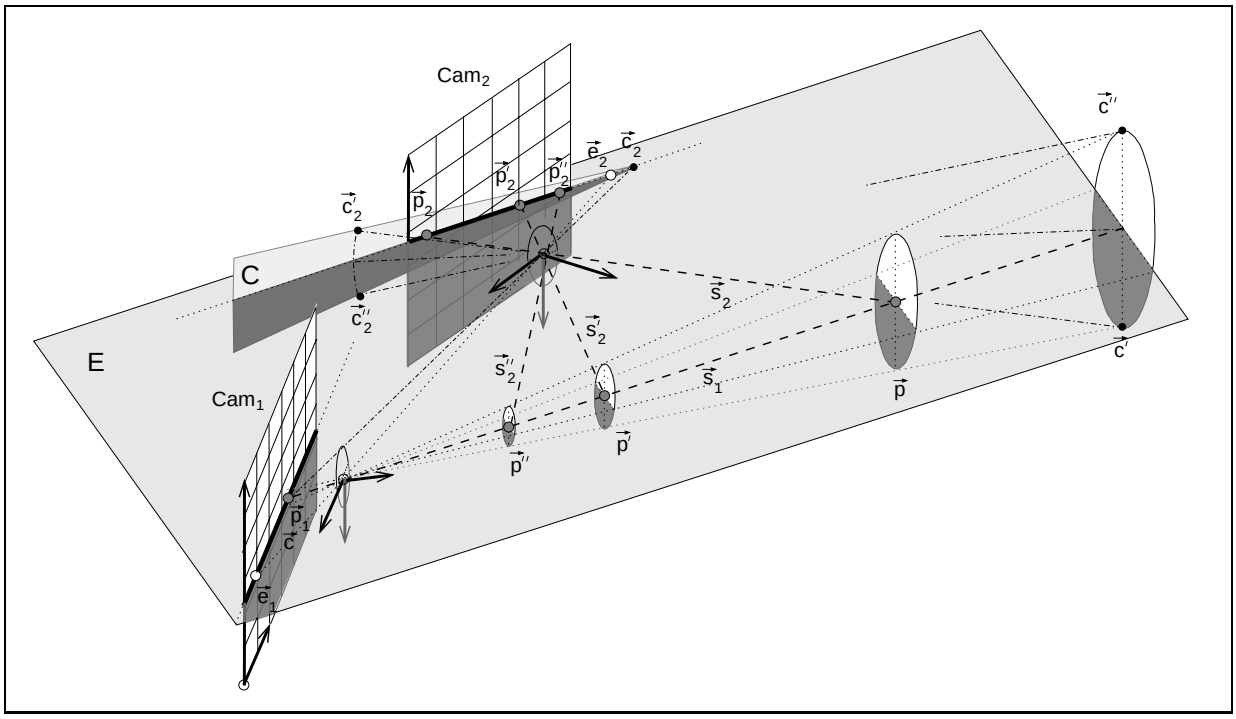


Abbildung 3.11: *Epipolar-Geometrie des binokularen Stereoaufbaus.*

$\vec{s}_0$ wie folgt:[6]

$$\begin{aligned} d_1 &= \frac{r_1'}{\tan \rho_1} \\ \text{mit} & \\ \rho_1 &= \cos^{-1} \frac{\vec{s}_0 \cdot \vec{s}_1}{\|\vec{s}_0\| \cdot \|\vec{s}_1\|} \end{aligned}$$

Der zu bestimmende 3D Punkt des Szenenmerkmals ergibt sich aus der Tiefe d_1 entsprechend:

$$\vec{p}_{wcs} = {}^{\circ}\vec{p}_{wcs}^{\,cam} + d_1 \cdot \frac{\vec{s}_0}{\|\vec{s}_0\|}$$

wobei mit ${}^{\circ}\vec{p}_{wcs}^{\,cam}$ der Ursprung des Kamerakoordinatensystems angegeben ist.

3.5.3 Binokularer Ansatz

Nach der Extraktion der Bildmerkmale $\mathbf{i} \in \mathbf{I}$ wird im Interpretationsprozeß versucht, in $\mathbf{I}$ Paare (Tripel, Quadrupel, ...) von Bildmerkmalen mit gleichen Basisattributen zu finden, die in Bildern unterschiedlicher Kameras extrahiert worden sind und die den gleichen Szenenpunkt abbilden. Dieser 3D Punkt bestimmt $\vec{p}_{wcs}$ des zugehörigen Szenenmerkmals. Werden hierbei Bildpunkte aus zwei Kameras einander zugeordnet, so spricht man von binokularem Stereo.

Für einen Stereo-Aufbau gilt die *Epipolar-Geometrie* der zwei beteiligten Kameras, vgl. [XZ96]. Diese Geometrie ist in Abb. 3.11 dargestellt. Dementsprechend ergibt sich für einen 2D Bildpunkt $\vec{p}_1$, der ein Bildmerkmal in der Bildebene der ersten Kamera repräsentiert, ein Sichtstrahl $\vec{s}_1$. Gesucht wird hierzu ein 2D Punkt $\vec{p}_2$, in der Bildebene der zweiten Kamera,

[6] Dies gilt analog für jedes weitere Sichtstrahlpaar.

der den 3D Szenenpunkt $\vec{p}$ über den Sichtstrahl $\vec{s}_2$ abbildet. Der 3D Szenenpunkt kann jedoch auf verschiedenen Positionen auf dem Sichtstrahl $\vec{s}_1$ liegen; so kann neben $\vec{p}$ auch $\vec{p'}$ und $\vec{p''}$ abgebildet werden. Mit den zugehörigen Sichtstrahlen $\vec{s'}_2$ und $\vec{s''}_2$ werden z.B. in der Bildebene der zweiten Kamera die weiteren 2D Punkte $\vec{p'}_2$ und $\vec{p''}_2$ abgebildet. Entsprechend der Epipolar-Geometrie liegen die Projektionen aller möglichen 3D Punkte des Sichtstrahls $\vec{s}_1$, so auch die Punkte $\vec{p}_2$, $\vec{p'}_2$ und $\vec{p''}_2$, auf der sog. *Epipolar-Linie* in der Bildebene der zweiten Kamera.

Diese Linie ist durch den Schnitt der Bildebene der zweiten Kamera mit der sog. *Epipolar-Ebene* E gegeben. Diese Ebene ist wiederum durch die Projektionszentren der beiden Kameras, die den Ursprüngen der Kamerakoordinatensysteme entsprechen und durch den Sichtstrahl $\vec{s}_s$ bestimmt. Somit liegen alle die Szenenpunkte, die mit dem Sichtstrahl $\vec{s}_1$ in dem Punkt $\vec{p}_1$ in der Bildebene der ersten Kamera abgebildet werden können, in E. Ebenso liegen in E alle möglichen Projektionen dieser Punkte auf die Bildebene der zweiten Kamera. Durch all diese Projektionen ist die Epipolar-Linie bestimmt, die dem Schnitt von E und der Bildebene entspricht. In Abb. 3.11 ist diese Linie als Gerade eingezeichnet, jedoch ergibt sich entsprechend der in dem verwendeten Kameramodell berücksichtigten Verzerrung eine gekrümmte Linie. Auch die Epipolar-Ebene E ist unter Berücksichtigung der Verzerrung eine gekrümmte Ebene.

Projiziert man das Projektionszentrum der ersten Kamera in die Bildebene der zweiten Kamera, so erhält man einen Punkt $\vec{e}_2$, der ebenfalls auf der Epipolar-Linie liegt. Dieser Punkt wird der *Epipol* genannt. In dem in Abb. 3.11 skizzierten Aufbau liegt $\vec{e}_2$ rechts neben der eingezeichneten Bildfläche der zweiten Kamera. Der zweite Epipol $\vec{e}_1$ des Aufbaus liegt in der Bildfläche der ersten Kamera.

Korrespondenzsuche

Entsprechend des sog. *Epipolar-Constraint* reduziert sich das Problem der Korrespondenzsuche somit von zwei Dimensionen auf eine. Es muß nicht mehr in der kompletten Bildebene der zweiten Kamera nach einem Bildpunkt gesucht werden, der dem Bildpunkt $\vec{p}_1$ zugeordnet werden kann, sondern nur noch auf der Epipolar-Linie. Jedoch muß aufgrund von Ungenauigkeiten bei der Kalibrierung und von Ungenauigkeiten bei der Bestimmung des Mittelpunktes der Bildmerkmale ausgegangen werden. Daher werden sich die Sichtstrahlen, die für die beiden Kameras den gleichen Szenenpunkt abbilden, nicht exakt schneiden. Es ist daher sinnvoll, statt nur auf der Epipolar-Linie, in einer *Korrespondenz-Region* C nach zu $\vec{p}_1$ korrespondierenden Bildmerkmalen zu suchen.

Die Region C liegt in der Bildebene der zweiten Kamera und ergibt sich durch eine zulässige Ungenauigkeit. Diese Ungenauigkeit ist wiederum durch einen zulässigen Ungenauigkeitskreis, der um den Bildpunkt $\vec{p}_1$ gelegt wird, bestimmt. Das Maß der Ungenauigkeit wird daher in Bildpunkten angegeben. Entsprechend der Umkehrung der Projektion ergibt sich für den Ungenauigkeitskreis ein 3D Kegel, vgl. Abb. 3.11. Aufgrund einer zugelassenen Ungenauigkeit müssen Sichtstrahlen der zweiten Kamera diesen Kegel schneiden, wenn die zugehörigen Punkte mit dem Punkt $\vec{p}_1$ ein zulässiges Stereopaar bilden. Es ist zu erkennen, daß mit größerem Abstand von der Kamera bei einem gleichen Ungenauigkeitsmaß die zulässige 3D Ungenauigkeit zunimmt.

Zur Bestimmung der zulässigen 3D Ungenauigkeit wird der Radius r_{3D} des Ungenauigkeitskegels in dem entsprechenden Abstand auf dem Sichtstrahl $\vec{s}_1$ bestimmt. So gilt z.B. an der

Position von $\vec{p}$ bei einem gegebenen Radius r_{2D} des Ungenauigkeitskreises:

$$\begin{array}{lcl} r_{3D} & = & \|\vec{p}_{cam_1}\| \cdot \tan\varphi \\ \text{mit} & & \\ \varphi & = & \cos^{-1} \frac{\vec{s}_1 \cdot \vec{s}_1^{\,u}}{\|\vec{s}_1\| \cdot \|\vec{s}_1^{\,u}\|} \end{array} \tag{3.12}$$

Hierbei entspricht $\vec{p}_{cam_1}$ dem Punkt $\vec{p}$, jedoch mit Bezug zum Kamerakoordinatensystem $[X_{cam}, Y_{cam}, Z_{cam}]$ der ersten Kamera und $\vec{s}_1^{\,u}$ dem Sichtstrahl eines Punktes $\vec{p}^{\,u}_{img_1}$ auf dem Ungenauigkeitskreis. $\vec{p}^{\,u}_{img_1}$ ist durch die Projektion von $\vec{p}_{cam_1}$ und r_{2d} bestimmt:[7]

$$\vec{p}^{\,u}_{img_1} = \vec{p}_{img_1} + \begin{pmatrix} 0 \\ r_{2D} \end{pmatrix} \tag{3.13}$$

Die durch den Ungenauigkeitskegel vorgegebene Korrespondenzregion C ist durch die Projektion der drei 3D Punkte $\vec{c}$, $\vec{c'}$ und $\vec{c''}$ in die Bildebene der zweiten Kamera gegeben ($\vec{c}_2$, $\vec{c'}_2$ und $\vec{c''}_2$). Hierbei gibt $\vec{c}$ die 3D Position von $\vec{p}_1$, der Projektion von $\vec{p}$ in die Bildebene der ersten Kamera, an. Es gilt für $\vec{c}$ im Weltkoordinatensystem:

$$\begin{array}{lcl} \vec{c}_{wcs} & = & {}^{wcs}\mathbf{T}_{cam_1} \cdot \vec{c}_{cam_1} \\ \text{mit} & & \\ \vec{c}_{cam_1} & = & \begin{pmatrix} p_x \\ p_y \\ -b_1 \end{pmatrix} \\ \vec{p}_{img_1} & = & \begin{pmatrix} p_x \\ p_y \end{pmatrix} \end{array}$$

wobei $\vec{p}_{img_1}$ dem Punkt $\vec{p}_1$ in Bildkoordinaten der ersten Kamera entspricht und b_1 die Kammerkonstante der ersten Kamera ist.

Zur Bestimmung der Punkte $\vec{c'}$ und $\vec{c''}$ wird die Epipolar-Linie in der ersten Kamera verwendet, die durch den Punkt $\vec{p}_1$ und den Epipol $\vec{e}_1$ bestimmt ist. Die zu $\vec{c'}$ und $\vec{c''}$ zugehörigen Projektionen $\vec{c'}_{img_1}$ und $\vec{c''}_{img_1}$ müssen dann senkrecht zu der Epipolar-Linie und jeweils ober- und unterhalb dieser auf dem Ungenauigkeitskreis liegen. Zur Bestimmung von $\vec{c'}$ und $\vec{c''}$ müssen die sich aus $\vec{c'}_{img_1}$ und $\vec{c''}_{img_1}$ ergebenden Sichtstrahlen in einer Tiefe geschnitten werden, bei der die Korrespondenz-Region C sicher bis zum Rand der Bildfläche der zweiten Kamera reicht. Für die Tiefe nimmt man daher einen genügend großen Wert an, der jedoch durch den Observierungsraum SSP_s des Szenenmodells begrenzt werden kann.

Bei der vorgestellten Bestimmung der Korrespondenz-Region C ist zum einen die im Kameramodell berücksichtigte Verzerrung κ und zum anderen die Unsicherheit in der Lage der Epipolar-Ebene unberücksichtigt geblieben. Entsprechend der Verzerrung sollten daher zur Bestimmung von C nicht nur die drei Eckpunkte $\vec{c}$, $\vec{c'}$ und $\vec{c''}$ herangezogen werden, sondern die Sichtstrahlen zu den Punkten $\vec{c'}_{img_1}$ und $\vec{c''}_{img_1}$ in unterschiedlicher Tiefe geschnitten werden. Die sich hieraus ergebenden 3D Punkte werden in die Bildebene der zweiten Kamera projiziert. Verbindet man alle so projizierten Punkte und auch $\vec{c}_2$ miteinander, so erhält man für C ein Polygon in der Bildebene der zweiten Kamera.

Nachdem die Lage der Epipolar-Ebene E durch die Position der Projektionszentren der beiden Kameras bestimmt ist, muß noch die Unsicherheit der Kalibrierung des Stereoaufbaus

[7] $\vec{p}_{img_1}$ ist als Projektion von $\vec{p}_{cam_1}$ mit dem in Abb. 3.11 als $\vec{p}_1$ bezeichneten 2D Punkt identisch, jedoch im Kamerakoordinatensystem $[X_{cam_1}, Y_{cam_1}, Z_{cam_1}]$ der ersten Kamera definiert.

eingehen, vgl. Anh. C.4. E ist zum einen durch den Sichtstrahl $\vec{s}_1$ bestimmt, zum anderen jedoch durch die Position des Brennpunktes der zweiten Kamera. Wird für diesen angenommen, daß seine Position real jeweils etwas ober- oder unterhalb von E liegen kann, so entarten die Epipolar-Linien zu Hyperbeln. Diese Unsicherheit wird jedoch in dem gewählten Unsicherheitskreis mit zugehörigem Unsicherheitskegel ausreichend berücksichtigt.

Auswahlkriterien

Um für eine gültige Stereozuordnung zu $\vec{p}_1$ eine Auswahl aus den Bildpunkten, die innerhalb der Korrespondenz-Region C liegen, treffen zu können, werden weitere Auswahlkriterien angesetzt. Ein Kriterium ist die relative Genauigkeit und die Sichtbarkeit des zugehörigen 3D Punktes. Hierzu muß jedoch zunächst der 3D Punkt bestimmt werden, der sich aus dem Bildpunkt $\vec{p}_1$ und einem Bildpunkt $\vec{p}_2$ aus der Region C ergibt.

Für diesen 3D Punkt wird angenommen, daß sich dieser in der Mitte auf der Strecke der kürzesten Distanz zwischen den beiden Sichtstrahlen befindet. Liegen auf den Sichtstrahlen an den Positionen des kürzesten Abstandes die beiden Punkte $\vec{a}_1$ und $\vec{a}_2$, so gilt für die Distanz $\vec{d}$:

$$\begin{aligned} \vec{d} &= \vec{a}_1 - \vec{a}_2 \\ &= ({}^{\circ}\vec{p}^{\,cam_1} + \lambda \cdot \vec{s}_1) - ({}^{\circ}\vec{p}^{\,cam_2} + \mu \cdot \vec{s}_2) \end{aligned}$$

Man bestimmt λ und μ, indem man $\|\vec{d}\|^2$ minimiert. Der zu setzende 3D Punkt des $\vec{p}_{wcs}$ des Szenenmerkmals ergibt sich dann aus $\vec{a}_1 + \frac{1}{2} \cdot \vec{d}$.

Für den so bestimmten Punkt wird nun getestet, ob sich dieser innerhalb des Observierungsraumes SSP_s des Szenenmodells befindet und ob sich ein Sichthindernis in Form einer Weltregion wr zwischen der Kamera und dem 3D Punkt befindet, vgl. die hierzu analoge Vorgehensweise bei der Bestimmung der Sichtbarkeit der Suchräume in Kap. 3.3.4.

Liegen in der Korrespondenz-Region C mehrere Bildpunkte, für die ein "sichtbarer" 3D Punkt bestimmt werden kann, so wird der Punkt als Korrespondenz ausgewählt, bei dem sich die zugehörigen Sichtstrahlen am nächsten kommen. Hierzu kann jedoch nicht die absolute Distanz $\|\vec{d}\|$ verwendet werden, sondern es muß eine relative Genauigkeit bezüglich des Unsicherheitskegels ermittelt werden. Damit wird in größerem Abstand zur Kamera auch ein größerer Abstand der Sichtstrahlen zugelassen. Entsprechend den Glg. 3.12 und 3.13 wird r_{3D} an der Position $\vec{a}_1$ auf dem Sichtstrahl $\vec{s}_1$ bestimmt und als Maß der zulässigen Ungenauigkeit verwendet. Nachdem die Korrespondenz-Region C jedoch die Projektion des Unsicherheitskegels in die Bildebene der zweiten Kamera ist, kann als relatives Maß auch die Breite der Region an der Position des zugeordneten Bildpunktes $\vec{p}_2$ verwendet werden.

Ist ein 3D Punkt für das Szenenmerkmal gefunden, so wird dieser in $\vec{p}_{wcs}$ vermerkt. Weiterhin werden die verwendeten Bildmerkmale der ersten und zweiten Kamera in $\mathbf{I}_{extr}$ aufgenommen.

Mehrere Kameras

In STABIL^{++} werden in dem Stereoansatz die Szenenmerkmale auch aus mehr als zwei Bildmerkmalen ermittelt, wenn in mehr als zwei Kameras Projektionen des Szenenmerkmals abgebildet sind. Hierzu werden die Bildmerkmale $\mathbf{I}$ entsprechend der zugehörigen Kameras gruppiert, so daß $\mathbf{I} = \mathbf{I}_{cam_1} \cup \ldots \cup \mathbf{I}_{cam_n}$ gilt. Zunächst werden entsprechend der oben angegebenen Schritte für die Bildmerkmale der ersten und zweiten Kamera aus $\mathbf{I}_{cam_1}$ und $\mathbf{I}_{cam_2}$ Korrespondenzen gesucht und Szenenmerkmale $\mathbf{s} \in \mathbf{S}_{1,2}$ erzeugt. Für alle Bildmerkmale, die man nicht

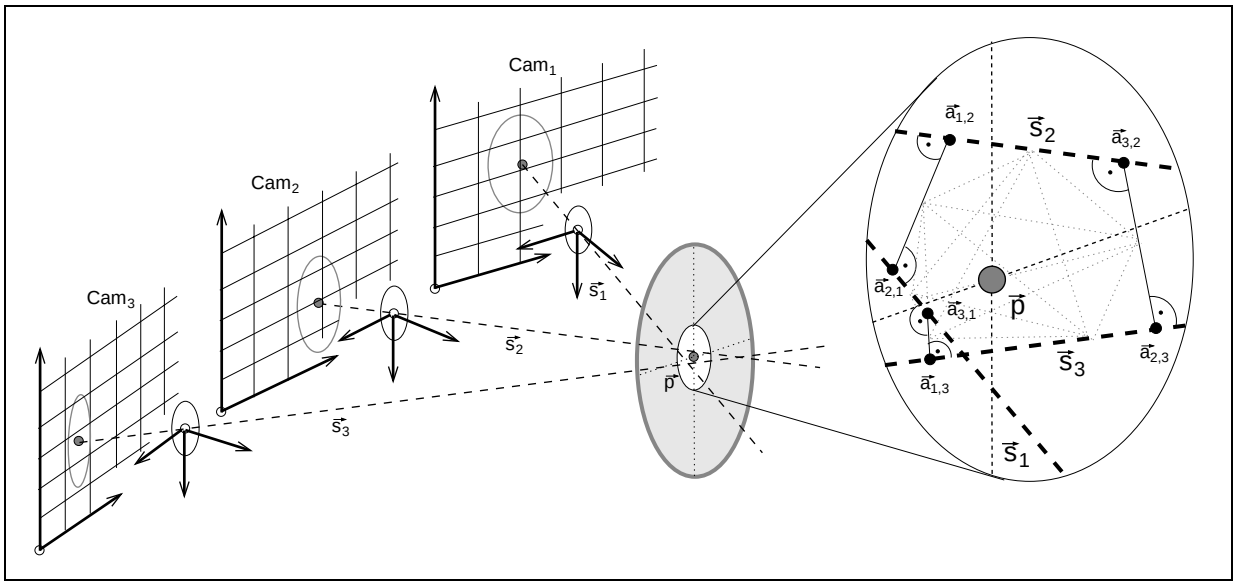


Abbildung 3.12: *Mehrfachstereo: Mittelung des 3D Punktes.*

zuordnen konnte, werden Pseudoszenenmerkmale erzeugt. In den Pseudoszenenmerkmalen ist keine 3D Position vermerkt und in $\mathbf{I}_{extr}$ ist lediglich ein Bildmerkmal eingetragen, vgl. auch Abschn. 3.5.1.

Für eine weitere dritte Kamera cam_3 wird dann versucht, für alle Szenenmerkmale $\mathbf{s}_j \in \mathbf{S}_{1,2}$ ein korrespondierendes Bildmerkmal in $\mathbf{I} = \mathbf{I}_{cam_3}$ zu finden. Hierzu werden die in $\mathbf{I}_{extr}$ der $\mathbf{s}_j$ eingetragenen Bildmerkmale verwendet. Für die Pseudoszenenmerkmale ergibt sich somit nur die Korrespondenzsuche für zwei Bildmerkmale, wie bei der ersten und zweiten Kamera. Ist ein 3D Punkt gefunden, so wird in dem entsprechenden Szenenmerkmal in $\mathbf{I}_{extr}$ noch das zugeordnete Bildmerkmal $\mathbf{i} \in \mathbf{I}_{cam_3}$ aufgenommen.

Wird für ein Szenenmerkmal $\mathbf{s}_k \in \mathbf{S}_{1,2}$, für das schon eine 3D Position gesetzt worden ist, eine Korrespondenz in $\mathbf{I}_{cam_3}$ gesucht, so wird die Korrespondenzsuche für jedes in $\mathbf{I}_{extr}$ vermerkte Bildmerkmal mit den Bildmerkmalen in $\mathbf{I}_{cam_3}$ durchgeführt. Wird keine weitere Korrespondenz gefunden, so kann $\mathbf{s}_k$ unverändert bleiben. Wird jedoch in $\mathbf{I}_{cam_3}$ ein weiteres korrespondierendes Bildmerkmal gefunden, so erhält man drei Sichtstrahlen $\vec{s}_1$, $\vec{s}_2$ und $\vec{s}_2$. Entsprechend der Darstellung in Abb. 3.12 ergeben sich für diese drei Sichtstrahlen jeweils sechs Punkte $\vec{a}_{1,2}$, $\vec{a}_{2,1}$, $\vec{a}_{1,3}$, $\vec{a}_{3,1}$, $\vec{a}_{2,3}$, und $\vec{a}_{3,2}$, die die Positionen auf den Sichtstrahlen markieren, die den kürzesten Abstand zwischen jeweils zwei Sichtstrahlen kennzeichnen. Der 3D Punkt $\vec{p}$ des Szenenmerkmals ergibt sich durch einen mittleren Punkt entsprechend:

$$\vec{p} = \frac{\sum_{i=1,\,j=1}^{n,\,n} a_{i,j}}{n}, \quad i \neq j, \quad n = \text{Anzahl der Kameras}$$

Das aus $\mathbf{I}_{cam_3}$ jetzt zusätzlich verwendete Bildmerkmal wird noch in $\mathbf{I}_{extr}$ des Szenenmerkmals eingetragen. Somit kann entsprechend der dargestellten Vorgehensweise für drei Kameras auch für jede weitere Kamera vorgegangen werden, so daß 3D Punkte aus beliebig vielen Stereo-Schnitten ermittelt werden können. Die Anzahl der zu verwendenden Kameras ist hierbei von der Anwendung abhängig.

3.6 Generierung von Hypothesen

3.6.1 Einführung

Im Interpretationszyklus wird mit der Generierung der Hypothesen die eigentliche Interpretation vorgenommen; es wird hierbei versucht, gültige Zuordnungen von 3D Szenenmerkmalen $\mathbf{s}_j \in \mathbf{S}$ zu den 3D Modellmerkmalen $\mathbf{m}_k \in \mathbf{M}_\nu$ der zu detektierenden Objektmodellinstanz obj_ν zu finden, vgl. Zeile 17, Alg. 3.3.[1] Wie im Kap. 3.5 beschrieben, werden in der Zeile 16 des Alg. die der Zuordnung zugrunde liegenden Szenenmerkmale erzeugt. Können jedoch keine 2D Bildmerkmale für einen Stereoansatz verwendet werden, so existieren in $\mathbf{S}$ Pseudoszenenmerkmale. Für diese Szenenmerkmale wird daher unmittelbar bei der Zuordnung zu einem Modellmerkmal die 3D Position über eine Tiefenschätzung ermittelt.

Ist es für ein Modellmerkmal nicht möglich, aus den im aktuellen Interpretationszyklus vorhandenen Bildmerkmalen ein 3D Szenenmerkmal zu ermitteln, dann wird die Position eines Szenenmerkmals vorhergesagt. Hierzu wird wie in dem Schritt zum 2D / 3D Übergang ein Pseudoszenenmerkmal generiert. Bei diesem ist jedoch kein Bildmerkmal in $\mathbf{I}_{extr}$ vermerkt, somit ist $|\mathbf{I}_{extr}| = 0$. Die 3D Position des Pseudoszenenmerkmals $\vec{p}_{wcs}$ wird entsprechend des vorhergesagten Suchraumes des zuzuordnenden primären Merkmals gesetzt. Als Qualitätsmaß q der Pseudoszenenmerkmale wird die Qualität der entsprechenden Positionsvorhersage übernommen, mit der auch die Größe des Suchraumes bestimmt wurde. Mit dem Qualitätsmaß wird sichergestellt, daß bei der Zuordnung von Szenenmerkmalen zu Modellmerkmalen die Szenenmerkmale bevorzugt verwendet werden, die aus Bildmerkmalen hervorgegangen sind.

Nachdem der Bezug eines Modellmerkmals $\mathbf{m}$ zu einem Objektmodellteil omp_μ in seinem primären Merkmal $\mathbf{f}_\mu$ modelliert ist, ist die Zuordnung der Szenenmerkmale zu den Modellmerkmalen bei der Generierung der Hypothesen auch mit einer Zuordnung zu den primären Merkmalen gleichzusetzen. Eine Zuordnung für das Objektmodellteil omp_μ wird als Assoziation $assoc_\mu$ von Szenenmerkmal $\mathbf{s}_j$ und primärem Merkmal $\mathbf{f}_\mu$ bezeichnet und ist als:

$$assoc_\mu = \langle \mathbf{s}_j, \mathbf{f}_\mu \rangle \tag{3.14}$$

definiert. Eine gültige Assoziation zeichnet sich dadurch aus, daß zum einen die Attribute $\{attr_1^{\mathbf{s}}, \ldots, attr_n^{\mathbf{s}}\}$ des Szenenmerkmals $\mathbf{s}$ mit den Attributen $\{attr_1^{\mathbf{m}}, \ldots, attr_k^{\mathbf{m}}\}$ des Modellmerkmals $\mathbf{m}$ übereinstimmen. Hierbei werden generell nur Assoziationen von Szenen- und Modellmerkmalen erzeugt, die das gleiche Basisattribut haben. Beispielsweise werden einem Modellmerkmal einer farbigen 3D Ellipse mit dem Attribut “blau” nur 3D Szenenmerkmale zugeordnet, die aus 2D Bildmerkmalen bestimmt wurden, die wiederum auf Bildregionen $reg_i \in REG^{blau}$ basieren. Zum anderen wird die Gültigkeit einer Assoziation durch Restriktionen / Bewertungsregeln $restr_k(.) \in RESTR_\mu$ bestimmt, die dem primären Merkmal $\mathbf{f}_\mu$ bekannt sind.

Kann für jedes Objektmodellteil $omp_i \in OMP_\nu$, dem ein primäres Merkmal zugeordnet ist, mindestens eine gültige Assoziation $assoc_i$ aufgestellt werden, so erhält man hierdurch mindestens eine sog. Hypothese $h_l \in H_\nu$ für die zu detektierende Objektmodellinstanz obj_ν. Eine Hypothese h ist als:

$$\begin{aligned} &h = \langle \{assoc_1, \ldots, assoc_n\}, q, obj \rangle \\ &\text{wobei} \\ &\forall\, assoc_i : restr_k(assoc_i) = \text{gültig}, \quad i = 1 \ldots n,\ k = 1 \ldots m_i \end{aligned} \tag{3.15}$$

[1] Die bei dem 2D / 3D Übergang bestimmten 3D Positionen der Szenenmerkmale müssen noch um den Translationsvektor $\vec{t}$ des zuzuordnenden Modellmerkmals korrigiert werden, vgl. Glg. 3.5 und 3.10.

definiert. Hierbei kann für jede Hypothese der Bezug zu der Objektmodellinstanz über obj hergestellt werden. Es ist weiterhin n die Anzahl der Objektmodellteile von obj mit primärem Merkmal, m_i jeweils die Anzahl der Restriktionen, für das der Assoziation $assoc_i$ zugehörige primäre Objektmodellteil und q ein Qualitätsmaß der Hypothese. Man bezeichnet eine komplette Hypothese auch als *match* (engl.) des Objektmodells auf die Szenenmerkmale. Dementsprechend wird der Prozeß der Zuordnung auch *matching* genannt.[2]

Mit der Generierung einer Hypothese $h \in H_\nu$ ist für die Objektmodellinstanz obj_ν nur ein Teil der Modellstruktur bestimmt. Dies begründet sich darauf, daß durch die Zuordnungen von 3D Szenenmerkmalen zu 3D Modell- / primären Merkmalen in den Assoziationen zunächst nur die innere Modellstruktur berücksichtigt werden kann. Darüber hinaus werden anhand der 3D Positionen der Szenenmerkmale die Ursprünge der lokalen Koordinatensysteme der Objektmodellteile festgelegt. Somit ist der Translationsanteil der geometrischen Modellstruktur bestimmt. Der Rotationsanteil wird erst in einem weiteren Schritt ermittelt und, basierend auf den Kompositionen von Objektmodellteilen, auf einer kompletten Hypothese ausgeführt. Auch die Berücksichtigung der äußeren Modellstruktur wird erst für eine komplette Hypothese durchgeführt und dann, wie die Rotationsanteile der geometrischen Struktur, erst im Schritt zur Bewertung der Hypothese verwendet, bei der die Qualität q bestimmt wird.

Diese Teilung der Interpretation in einen Schritt der Hypothesengenerierung und in die Bewertung der Hypothesen mit Auswahl einer Hypothese ist in der großen Anzahl von möglichen Assoziationen begründet. So werden zuerst nur für alle möglichen Assoziationen die Restriktionen auf der Ebene der Objektmodellteile, die Restriktionen $restr_i \in RESTR$ der primären Merkmale und die Restriktionen, die sich aus der hierarchischen, inneren Objektmodellstruktur ergeben, angewendet. Somit wird die Anzahl der Hypothesen drastisch eingeschränkt. Die vergleichsweise aufwendige Bestimmung der kompletten geometrischen Struktur und die Berücksichtigung der äußeren Objektmodellstruktur muß dann nur noch auf einer geringen Anzahl von Hypothesen, die aus gültigen Assoziationen aufgebaut ist, ausgeführt werden.

Der Suchraum des Korrespondenzproblems,[3] der sich für die aufzustellenden Hypothesen ergibt, ist zunächst erst einmal durch die Anzahl der vom Objektmodell verwendeten primären Merkmale und der Anzahl der gefundenen Szenenmerkmale abhängig. Die Suchraumgröße läßt sich jedoch durch die Wahl von primären Merkmalen mit unterschiedlichen Attributen $attr_i^{\mathbf{m}}$ der zugehörigen Modellmerkmale reduzieren.

Die Suche nach gültigen Assoziationen wird in STABIL^{++} durch einen modellgetriebenen Interpretationsbaum gesteuert, [GLP84], [GLP87], [Mun96] und [Lan98]. Der Interpretationsbaum ist ein Suchbaum, in dem in jedem Knoten eine mögliche Zuordnung von 3D Szenenmerkmalen zu 3D Modellmerkmalen / primären Merkmalen vermerkt ist. Die Suche in dem Baum wird durch die Restriktionen der primären Merkmale und duch die Restriktionen gesteuert, die sich aus der hierarchischen, inneren Objektmodellstruktur ergeben.

Im folgenden Abschnitt sind zunächst die bei der Suche im Interpretationsbaum anzuwendenden Restriktionen beschrieben. Der Aufbau des eigentlichen Interpretationsbaumes schließt sich daran an. In zwei hierauf folgenden Abschnitten wird betrachtet, wie die Suche bei fehlenden oder falschen Szenenmerkmalen durchgeführt wird und wie der Aufbau der hierarchischen, inneren Objektmodellstruktur die Komplexität der Suche beeinflußt.

[2] Der Prozeß der Zuordnung ist auch als Struktur-Vergleichsverfahren bekannt.

[3] Die Verwendung des Begriffs “Suchraum” kann hier zu Verwirrung führen: Mit der Suchraumgröße des Korrespondenzproblems wird ein Maß für seine Komplexität angegeben. Dies ist nicht mit den 3D Suchräumen SSP, die für die Lage der Objektmodellteile vorhergesagt werden, zu verwechseln, vgl. Kap. 3.2.

$restr^{(originQ)}(.)$	**origin q**uality	Merkmalsebene
$restr^{(insideSsp)}(.)$	**inside s**earch **sp**ace	Merkmalsebene
$restr^{(areaFnd)}(.)$	**area fou**nd	Merkmalsebene
$restr^{(exentr)}(.)$	**exzentr**icity	Merkmalsebene
$restr^{(parentD)}(.)$	**parent d**istance	Modellebene
$restr^{(siblingD)}(.)$	**sibling d**istance	Modellebene

Tabelle 3.3: *Restriktionen zur Beschränkung der Größe des Interpretationsbaumes.*

3.6.2 Restriktionen

Zur Einschränkung der exponentiellen Komplexität der Suche nach gültigen Assoziationen wird im Interpretationsbaum für jede neu aufgestellte Assoziation die Zulässigkeit geprüft. Somit wird nicht erst eine komplett aufgestellte Hypothese überprüft, sondern frühzeitig eine ungültige Zuordnung erkannt. Für diese Prüfungen werden sog. Restriktionen $restr(.)$ (engl. *constraint*) verwendet. Man unterscheidet hierbei die Restriktionen nach der Anzahl der für die Entscheidung notwendigen Assoziationen. So werden lokal in einem Knoten des Interpretationsbaumes die Restriktionen nur auf einer Assoziation angewendet, man spricht von einer *unären* Restriktion (engl. *unary-constraint*). Restriktionen, die eine Aussage über die Zuordnungen in zwei Assoziationen machen und somit zwei Assoziationen in Relation zueinander betrachten, werden als *binäre* Restriktionen (engl. *binary-constraint*) bezeichnet. Werden drei Assoziationen in Relation zueinander betrachtet, so spricht man von *tertiären* Restriktionen.[4] In [Gri90a] wird weiterhin eine Erweiterung der Interpretation um 3D Kanten, zylindrische Merkmale und 3D Oberflächen beschrieben, daher wird dort noch zwischen 2D und 3D Restriktionen unterschieden.

In STABIL^{++} wird die Interpretation ebenfalls mit 3D Merkmalen durchgeführt, jedoch wird hier zwischen Restriktionen unterschieden, die zum einen auf der Merkmalsebene und zum anderen auf der Modellebene angewendet werden. Restriktionen der Merkmalsebene sind grundsätzlich unär, denn es wird mit diesen nur eine Aussage über das zuzuordnende Szenenmerkmal getroffen. Die mehrstelligen Restriktionen betrachten immer mehrere Assoziationen und berücksichtigen damit mehrere Objektmodellteile aus der inneren Objektmodellstruktur. Diese werden daher als Restriktionen auf der Modellebene bezeichnet. In der Tab. 3.3 ist eine Übersicht über die verschiedenen Restriktionen $restr^{type}(.)$ gegeben, wobei eine Erläuterung der sich aus den zugehörigen englischen Begriffen ergebenden Bezeichnungen $type$ aufgeführt ist.

Restriktionen auf Merkmalsebene

Die Restriktionen auf der Merkmalsebene teilen sich in generelle Restriktionen und Restriktionen, die explizit für ein primäres Merkmal mit $RESTR$ angegeben sind.

Mit den generellen Restriktionen auf der Merkmalsebene wird überprüft, ob die mit der Assoziation $assoc_\mu$ durchgeführte Zuordnung des 3D Szenenmerkmals $\mathbf{s}$ zu dem 3D Modellmerkmal $\mathbf{m}$ des primären Merkmals $\mathbf{f}_\mu$ des Objektmodellteiles omp_μ zulässig ist. Für alle Zuordnungen wird geprüft, ob die Güte / Qualität des Szenenmerkmals ausreichend ist und ob der zugehörige 3D Punkt innerhalb des Suchraumes liegt.

[4] Man spricht auch von einstelligen (unären) und mehrstelligen (binären und tertiären) Restriktionen, vgl. [Gri90b] und [Mun96].

Entsprechend der unterschiedlichen Attribute $attr^{\mathbf{m}}$ der Modellmerkmale können für verschiedene primäre Merkmale noch weitere Restriktionen überprüft werden, die bei dem primären Merkmal explizit angegeben sind. Im folgenden werden diese Restriktionen am Beispiel einer farbigen 3D Ellipse als primäres Merkmal erläutert. Hierbei werden die Fläche und die Exzentrizität der Ellipse überprüft.

Es sei hier nochmal erwähnt, daß die Basisattribute eines primären Merkmals, wie z.B. die Farbe einer Ellipse, eigentlich auch noch mit einer Restriktion überprüft werden müssen. Jedoch werden beim Aufbau des modellgetriebenen Interpretationsbaumes nur solche Szenenmerkmale $\mathbf{s} \in \mathbf{S}$ ausgewählt, bei denen das Basisattribut mit dem des Modellmerkmals übereinstimmt.

Güte des 3D Punktes: Mit $restr^{(originQ)}(.)$ wird für das Szenenmerkmal $\mathbf{s}$, das mit der Assoziation $assoc$ dem Modellmerkmal $\mathbf{m}$ des primären Merkmals $\mathbf{f}$ zugeordnet worden ist, die Qualität q seines 3D Punktes überprüft. Die Qualität des 3D Punktes wird entweder beim 2D / 3D Übergang gesetzt oder ist durch die Qualität der Vorhersage bestimmt, falls ein Pseudoszenenmerkmal zugeordnet wird. Somit gilt:

$$restr^{(originQ)}(assoc) = \text{wahr} \Leftrightarrow \text{wenn } q \geq q_{min}$$

hierbei ist q_{min} das gleiche Mindestmaß, wie es zur Bestimmung der 3D Suchräume verwendet wird.

Position im Suchraum: Mit $restr^{(insideSsp)}(.)$ wird überprüft, ob der 3D Punkt $\vec{p}_{wcs}$ des Szenenmerkmals $\mathbf{s}$ in einem der kugelförmigen Suchräume $\{ssp_1 \ldots ssp_n\}$ liegt, der für das primäre Merkmal bestimmt worden ist, das mit der Assoziation $assoc$ dem Szenenmerkmal $\mathbf{s}$ zugeordnet ist. Es gilt:

$$restr^{(insideSsp)}(assoc) = \text{wahr} \Leftrightarrow \text{wenn } \bigvee_{i=1}^{n} \|\vec{p}_{wcs} - \vec{p}_{wcs}^{\,ssp_i}\| \leq r_i = \text{wahr}$$

Hierbei ist r_i der Radius und $\vec{p}_{wcs}^{\,ssp_i}$ der Mittelpunkt des Suchraumes ssp_i. Aufgrund der Verwendung der gleichen Positionsvorhersage zur Bestimmung der Suchräume und der 3D Position der Pseudoszenenmerkmale, ist für diese $restr^{(insideSsp)}(.)$ immer erfüllt. Für Szenenmerkmale, die aus Bildmerkmalen extrahiert worden sind, ist $restr^{(insideSsp)}(.)$ nur für die Szenenmerkmale erfüllt, die innerhalb des Suchraumes des betrachteten Objektmodellteiles liegen.[5]

Gefundene Fläche: Ist für ein Bildmerkmal als Attribut eine Fläche a ermittelt worden, so kann die Restriktion $restr^{(areaFnd)}(.)$ angewendet werden. Mit dieser wird überprüft, ob die 2D Bildmerkmale $\mathbf{i}_k \in \mathbf{I}_{extr}$ des Szenenmerkmals $\mathbf{s}$ der Assoziation $assoc$ annähernd die Größe des projizierten primären Merkmals $\mathbf{f}_\mu$ der Assoziation $assoc$ haben. Zur Projektion eines impliziten primären Merkmals wird das Objektmodellteil omp_μ hypothetisch an die 3D Position $\vec{p}_{wcs}$ des Szenenmerkmals gesetzt und der Volumenkörper in jeweils die Bilder projiziert, aus denen die Bildmerkmale extrahiert worden sind. Für ein explizites primäres Merkmal wird die Position $\vec{p}_{wcs}$ für den Mittelpunkt des Merkmals angenommen und diese dann projiziert. Die für die Projektion notwendigen Kameraparameter $camPar$ und $camPose$ sind bei den Bildmerkmalen

[5] Anm.: Bei der Extraktion der Bildmerkmale werden die Suchräume aller primären Merkmale der Objektmodellteile zusammengefaßt projiziert, vgl. Kap. 3.4.2. Somit werden auch Szenenmerkmale mit passendem Basisattribut bestimmt, die nicht innerhalb des Einzelsuchraums des betrachteten Objektmodellteils liegen.

vermerkt. Es gilt daher:[6]

$$restr^{(areaFnd)}(assoc) = \text{wahr} \Leftrightarrow \text{wenn} \bigwedge_{j=1}^{n} area_{min} \leq \frac{a_i}{a^{(proj)}} \leq area_{max} = \text{wahr}$$

Hierbei ist n die Anzahl der Bildmerkmale $\mathbf{i}_k \in \mathbf{I}_{extr}$. Mit der Bildregiongröße $a^{(proj)}$ ist die Fläche des projizierten primären Merkmals $\mathbf{f}$ bezeichnet.[7] Mit $area_{min}$ und $area_{max}$ ist angegeben, um wieviel die projizierte Fläche minimal und maximal von der Fläche der extrahierten Bildmerkmale abweichen darf.

Bei der Anwendung der Restriktion $restr^{(areaFnd)}(.)$ ist jedoch zu berücksichtigen, ob die Projektion innerhalb der Bildfläche zu liegen kommt. Ist nur ein Teil innerhalb des Sichtbereiches der entsprechenden Kamera, so kann $restr^{(areaFnd)}(.)$ nur mit abgeschwächten Grenzwerten $area_{min}$ und $area_{max}$ angewendet werden.[8]

Exzentrizität: Sind für die zugehörigen Bildmerkmale des primären Merkmals einer 3D Ellipse als Attribut die beiden Halbmesser r_1 und r_2 bestimmt worden, so läßt sich mit dem Verhältnis von r_1 zu r_2 die Exzentrizität $exz^{\mathbf{i}}$ der extrahierten Bildregion bestimmen. Mit der Restriktion $restr^{(exzentr)}(.)$ läßt sich die Exzentrizität der Bildregionen, die den Bildmerkmalen $\mathbf{i}_k \in \mathbf{I}_{extr}$ des Szenenmerkmals $\mathbf{s}$ der Assoziation $assoc$ zugehören, mit der Exzentrizität des projizierten primären Merkmals $\mathbf{f}_\mu$ der Assoziation $assoc$ vergleichen. Auf die projizierte Bildregion wird, wie bei der Extraktion der Bildmerkmale, eine Ellipsenanpassung durchgeführt, über die man die Exzentrizität $exz^{\mathbf{f}}$ erhält. Es gilt daher:

$$restr^{(exzentr)}(assoc) = \text{wahr} \Leftrightarrow \text{wenn} \bigwedge_{j=1}^{n} exz_{min} \leq \frac{exz_j^{\mathbf{i}}}{exz^{\mathbf{f}}} \leq exz_{max} = \text{wahr}$$

Hierbei ist n die Anzahl der Bildmerkmale $\mathbf{i}_k \in \mathbf{I}_{extr}$. Mit exz_{min} und exz_{max} ist angegeben, um wieviel die Exzentrizität der projizierten Ellipse minimal und maximal von der Exzentrizität der Ellipse der extrahierten Bildmerkmale abweichen darf. Auch hier können Ausreißer über die Betrachtung des Mittelwertes des Verhältnisses der Exzentrizitäten unterdrückt werden.

Restriktionen auf Modellebene

Mit den Restriktionen auf Modellebene wird basierend auf der hierarchischen, inneren Objektmodellstruktur und dem Translationsanteil der geometrischen Struktur geprüft, ob mit den 3D Positionen der Assoziationen die 3D Abstände der Gelenke / Knoten zwischen den Objektmodellteilen des Objektmodells eingehalten sind. Diese Restriktion kann angewendet werden, da die zu detektierenden Objekte durch ein Modell repräsentiert werden, das sich aus einzelnen starren Objektmodellteilen zusammensetzt. Bei der vorgestellten Modellierung des menschlichen Körpers, die sich an der Knochenstruktur orientiert, kann man daher von einer “Knochenlängen”-Restriktion sprechen. Nachdem die Objektmodelle in der Regel nicht auf ein spezifisches Objekt, z.B. eine individuelle zu detektierende Person, angemessen sind, werden hier

[6] Für die bei der Projektion der Objektmodellteile oder der expliziten Merkmale notwendige Koordinatentransformation ist bisher nur die 3D Position ermittelt worden. Der Rotationsanteil kann sich entweder auf die im letzten Interpretationsschritt bestimmte Lage abstützen oder kann bei Rotationssymmetrie unberücksichtigt bleiben.

[7] Für die Projektion werden die Kameraparameter $camPar_j$ und $camPose_j$ des Bildmerkmals $\mathbf{i}_j$ verwendet.

[8] Ausreißer lassen sich durch zeitliche Mittelwertbildung unterdrücken (exponentielles Fenster).

jedoch keine festen Grenzen überprüft. Zudem wird eine Objektmodellinstanz auch an das detektierte Objekt angepaßt.

Bei den Restriktionen, die die 3D Abstände der Gelenke / Knoten überprüfen, werden in STABIL^{++} zwei Varianten unterschieden. Bei der ersten, der sog. "Vater-Abstand"-Restriktion, wird die "Knochenlänge" zum jeweiligen Vorgängerobjektmodellteil überprüft. Bei der zweiten, der sog. "Geschwister-Abstand"-Restriktion, wird geprüft, ob eine bestimmte Länge der Knochenstruktur zwischen Objektmodellteilen mit dem gleichen Vorgängerobjektmodellteil gegeben ist.

Vater-Abstand: Die Restriktion $restr^{(parentD)}(.)$ ist eine binäre Restriktion, denn diese wird auf zwei Assoziationen angewendet. Für ein Objektmodellteil omp_μ kann diese Restriktion daher nur angewendet werden, falls schon eine Assoziation für sein Vorgängerobjektmodellteil omp_ν besteht. Dies ist aufgrund des Aufbaus des verwendeten Interpretationsbaumes gewährleistet. Nur für das erste Objektmodellteil $omp_{0.1}$ der inneren Objektmodellstruktur kann diese Restriktion keine Anwendung finden. Für $restr^{(parentD)}(.)$ gilt:

$$restr^{(parentD)}(assoc_\mu, assoc_\nu) = \text{wahr} \Leftrightarrow \text{wenn } |d_{\mathbf{s}} - d_{\mathbf{m}}| \leq \Delta_{max} = \text{wahr}$$

wobei

$$\begin{array}{rcl} d_{\mathbf{s}} & = & \|{}^{\mathbf{s}}\vec{p}^{\,\mu}_{wcs} - {}^{\mathbf{s}}\vec{p}^{\,\nu}_{wcs}\| \\ d_{\mathbf{m}} & = & \|{}^{\circ}\vec{p}^{\,\mu}_{wcs} - {}^{\circ}\vec{p}^{\,\nu}_{wcs}\| \\ obj_\mu \in OBJ_\nu & & \end{array} \tag{3.16}$$

Hierbei ist d_s die 3D Distanz der beiden 3D Punkte der zugeordneten Szenenmerkmale und d_m die 3D Distanz auf der Seite der Modellmerkmale. d_m bestimmt sich daher aus den Positionen der Ursprünge der lokalen Koordinatensysteme der beiden Objektmodellteile. Mit Δ_{max} wird angegeben, um wieviel sich die beiden Distanzen unterscheiden dürfen, damit die Assoziation $assoc_\mu$ noch akzeptiert wird. Δ_{max} wird als die maximale 3D Zuordnungstoleranz (engl. *3d matching tolerance*) bezeichnet.[9]

Geschwister-Abstand: Die Überprüfung des Geschwisterabstandes mit der Restriktion $restr^{(siblingD)}(.)$ ist der Restriktion $restr^{(parentD)}(.)$ ähnlich. Es handelt sich hierbei jedoch um eine tertiäre Restriktion, bei der die 3D Distanzen zwischen zwei Objektmodellteilen omp_μ und omp_ξ verglichen, die in der hierarchischen, inneren Objektmodellstruktur das gleiche Vorgängerobjektmodellteil omp_ν haben. Es müssen daher für alle drei Objektmodellteile Assoziationen bestehen, damit $restr^{(siblingD)}(.)$ angewendet werden kann. $restr^{(siblingD)}(.)$ ist somit eine tertiäre Restriktion, für die gilt:

$$restr^{(siblingD)}(assoc_\mu, assoc_\nu, assoc_\xi) = \text{wahr} \Leftrightarrow \text{wenn } |d_{\mathbf{s}} - d_{\mathbf{m}}| \leq \Delta_{max} = \text{wahr}$$

wobei

$$\begin{array}{rcl} d_{\mathbf{s}} & = & \|{}^{\mathbf{s}}\vec{p}^{\,\mu}_{wcs} - {}^{\mathbf{s}}\vec{p}^{\,\xi}_{wcs}\| \\ d_{\mathbf{m}} & = & \|{}^{\circ}\vec{p}^{\,\mu}_{wcs} - {}^{\circ}\vec{p}^{\,\xi}_{wcs}\| \\ obj_\mu \in OBJ_\nu \wedge obj_\xi \in OBJ_\nu & & \end{array} \tag{3.17}$$

[9] Ist das zu detektierende Objekt in seinen Ausmaßen exakt bekannt und vermessen, so können für das initiale Objektmodell die Größe der einzelnen Objektmodellteile entsprechend gesetzt werden. Somit läßt sich eine kleine Zuordnungstoleranz wählen.

Hierbei ist d_s die 3D Distanz der beiden 3D Punkte der zugeordneten Szenenmerkmale und d_m die 3D Distanz auf der Seite der Modellmerkmale. d_m bestimmt sich daher aus den Positionen der Ursprünge der lokalen Koordinatensysteme der beiden Objektmodellteile. Mit Δ_{max} wird angegeben, um wieviel sich die beiden Distanzen unterscheiden dürfen, damit die Assoziation $assoc_\mu$ noch akzeptiert wird.

3.6.3 Aufbau des Interpretationsbaumes

Der zur Suche nach gültigen Assoziationen und damit auch zur Suche nach Hypothesen verwendete modellgetriebene Interpretationsbaum wird in STABIL^{++} implizit durch die Objektmodellstruktur aufgebaut. In diesem Abschnitt wird der Aufbau des und die Suche im Interpretationsbaum zunächst allgemein an dem Algorithmus und dann anhand einer Beispielinterpretation erläutert.

Algorithmus

Aufgrund der Baumstruktur des Interpretationsbaumes bietet sich eine Darstellung in Form eines rekursiven Algorithmus an, vgl. Alg. 3.6. Auf der Grundlage der Szenenmerkmale $\mathbf{S} = \mathbf{S}^{(1)} \cup \ldots \cup \mathbf{S}^{(m)}$, von denen eine Gruppierung nach den Basisattributen $(1) \ldots (m)$ bekannt ist, werden in dem Algorithmus Hypothesen $\{h_1, \ldots, h_k\}$ für die Objektmodellinstanz obj gesucht. Hierzu wird auf eine Liste von primären Merkmalen $\mathbf{F} = \{\mathbf{f}_1, \ldots, \mathbf{f}_n\}$ der Objektmodellinstanz obj zurückgegriffen, die schon im Teilschritt der Detektion im Alg. 3.3 in den Zeilen 5 - 10 rekursiv erzeugt wurden. In dieser Liste stehen die primären Merkmale der einzelnen Objektmodellteile von obj entsprechend der hierarchischen, inneren Objektmodellstruktur. Hierbei wird mit dem ausgezeichneten Wurzelelement $omp_{0.1}$ begonnen. Falls vorhanden, wird das primäre Merkmal in $\mathbf{F}$ aufgenommen und anschließend wird dies für alle Nachfolgeobjektmodellteile durchgeführt. Entsprechend der Abb. 2.1 (a) werden jeweils die "Äste" der inneren Objektmodellstruktur von oben nach unten und dann von links nach rechts durchlaufen. Man erhält daher für $\mathbf{F} = \{\mathbf{f}_{0.1}, \mathbf{f}_{1.1}, \mathbf{f}_{2.1}, \mathbf{f}_{3.1}, \mathbf{f}_{1.2}, \mathbf{f}_{2.2}, \ldots, \mathbf{f}_{2.5}\}$.[10]

Mit der so entstandenen Liste $\mathbf{F}$ der primären Merkmale und den zugehörigen Modellmerkmalen wird die Suche im Interpretationsbaum gesteuert. Daher wird auch davon gesprochen, daß der Interpretationsbaum implizit durch die hierarchische, innere Objektmodellstruktur bestimmt ist.[11] Zu Beginn der Generierung der Hypothesen wird auf der obersten Ebene, der Wurzelebene mit der Nummer 0, des Interpretationsbaumes versucht, für das erste primäre Merkmal $\mathbf{f}_1$ aus $\mathbf{F}$ eine Assoziation aufzustellen. Hierzu wird aus der Liste der Szenenmerkmale $\mathbf{S}$ ein Szenenmerkmal $\mathbf{s}$ mit dem gleichen Basisattribut $attr$ gesucht. Ist ein solches vorhanden, kann eine Assoziation aus $\mathbf{s}$ und $\mathbf{f}$ gebildet werden. Zuvor wird noch überprüft, ob es sich bei $\mathbf{s}$ um ein Pseudoszenenmerkmal handelt. Für ein Pseudoszenenmerkmal wird noch die Tiefenschätzung durchgeführt. Die so aufgestellte Assoziation ist der erste Knoten des Interpretationsbaumes in der Ebene 1.

Für diese Assoziation wird die Gültigkeit anhand der Restriktionen überprüft.[12] Ist die Asso-

[10] Aufgrund der gewählten Numerierung wird die erste Ziffer solange erhöht, bis ein Objektmodellteil ohne Nachfolger erreicht worden ist, anschließend wird die zweite Ziffer um eins erhöht und mit der ersten Ziffer wieder bei eins begonnen.

[11] Anm.: Die Baumstruktur der hierarchischen, inneren Objektmodellstruktur ist jedoch nicht mit der Baumstruktur des Interpretationsbaumes identisch.

[12] Für die erste Assoziation können nur die merkmalsbezogenen unären Restriktionen $restr_l(.) \in RESTR$ überprüft werden (Zeile 14). Für die mehrstelligen Restriktionen (Zeilen 16 und 19) müssen zuvor die notwendigen Assoziationen erstellt worden sein.

Eingabe: obj *// Objektmodellinstanz*
Eingabe: $\mathbf{F} = \{\mathbf{f}_1, \ldots, \mathbf{f}_n\}$ *// Liste von primären Merkmalen,* ***s. Alg. 3.3***
Eingabe: $\mathbf{S} = \mathbf{S}^{(1)} \cup \ldots \mathbf{S}^{(m)}$ *// Liste von Szenenmerkmalen, nach Basisattributen gruppiert*
Berechne: $H \Leftarrow$ Hypothesen $\{h_1, \ldots, h_k\}$ für obj

```
 1: A ⇐ ∅    // Liste von Assoziationen
 2: f ⇐ f_1 ∈ F    // hole oberstes primäres Merkmal aus F, Annahme: f ist omp_μ zugeordnet und
    omp_ν ist sein Vorgängerobjektmodellteil in obj
 3: F ⇐ F \ {f}    // nehme f aus der Liste der primären Merkmale heraus
 4: attr ⇐ Basisattribut von f
 5: S_tmp ⇐ S ∩ S^(attr)    // alle noch vorhandenen Szenenmerkmale mit dem Attribut attr
 6: S_sicher ⇐ S_tmp
 7: für alle s_i ∈ S_tmp wiederhole
 8:    S ⇐ S \ {s_i}    // nehme s_i aus der Liste der Szenenmerkmale heraus
 9:    b ⇐ Anzahl der Bildmerkmale in I_extr von s_i
10:    falls b < 2 dann
11:       s_i ⇐ Bestimme p⃗_wcs über monokularen Ansatz mit Wissen von f    // s. Kap. 3.5.2
12:    •
13:    assoc_μ ⇐ Erzeuge Assoziation aus s_i und f
14:    ok ⇐ restr_l(assoc_μ), ∀ restr_l(.) ∈ RESTR_μ
15:    falls assoc_ν ∈ A dann
16:       ok ⇐ ok ∧ restr^(parentD)(assoc_μ, assoc_ν)
17:       für alle (omp_j ∈ OMP_μ) ∧ (i ≠ μ) wiederhole
18:          falls (assoc_ν ∈ A) ∧ (assoc_j ∈ A) dann
19:             ok ⇐ ok ∧ restr^(siblingD)(assoc_μ, assoc_ν, assoc_j)
20:          •
21:       •
22:    •
23:    falls ok dann
24:       A ⇐ A ∪ assoc_μ
25:       falls F = ∅ dann
26:          h ⇐ neue Hypothese aus A erstellen
27:          H ⇐ H ∪ {h}
28:          A ⇐ A \ {assoc_μ}    // letzte Assoziation wieder entfernen
29:       sonst
30:          f ⇐ f_1 ∈ F    // Parameter für Rekursion
31:          F ⇐ F \ {f}    // nehme f aus der Liste der primären Merkmale heraus; Parameter für
             Rekursion
32:          attr ⇐ Basisattribut von f
33:          S_tmp ⇐ S ∩ S^(attr)    // Parameter für Rekursion
34:          rekursiv Zeilen 7 - 39    // Ebene nach unten
35:       •
36:    •
37: •
38: S ⇐ S ∪ S_sicher    // Freigeben der Szenenmerkmale mit Attribut attr
39: F ⇐ F ∪ {f}    // Ebene nach oben
```

Algorithmus 3.6: *Interpretationsbaum: Suche nach Hypothesen für das Objektmodell* obj.

ziation gültig, so kann für das zweite primäre Merkmal aus $\mathbf{F}$ versucht werden eine Assoziation zu finden. Man erreicht somit die nächste Stufe des Interpretationsbaumes (Ebene 2), vgl. Abb. 3.15 für eine Beispielinterpretation. Für das zweite primäre Merkmal wird wie für das erste verfahren. Hat das zweite primäre Merkmal ein anderes Basisattribut, so wird in der entsprechenden Gruppierung von $\mathbf{S}$ das erste Szenenmerkmal für die Assoziation verwendet. Hat das zweite primäre Merkmal jedoch das gleiche Basisattribut, so kann für eine Assoziation nur noch das zweite Szenenmerkmal aus der entsprechenden Gruppierung von $\mathbf{S}$ verwendet werden, denn das erste Szenenmerkmal ist von der ersten Assoziation bereits belegt. Im Rahmen der kompletten Traversierung des Interpretationsbaumes werden auch noch die anderen Kombinationen, die ein primäres Merkmal $\mathbf{f} \in \mathbf{F}$ mit jedem Szenenmerkmal $\mathbf{s} \in \mathbf{S}^{(attr)}$, bilden kann, entsprechend des Basisattributs $attr$ erzeugt und überprüft.

Sind für alle primären Merkmale gültige Assoziationen gebildet worden, so hat man die unterste Ebene des Interpretationsbaumes erreicht und eine Hypothese gefunden (Zeile 25). Die Hypothese h wird aus den gültigen Assoziationen gebildet, die beim aktuellen Stand der Traversierung für alle primären Merkmale $\mathbf{f}_i \in \mathbf{F}$ gefunden wurden. Die Güte q der Hypothese wird noch nicht gesetzt. In der Darstellung des Interpretationsbaumes zeichnet sich eine Hypothese durch einen kompletten Pfad von der Ebene 0 bis zur untersten Ebene aus.[13] In der Abb. 3.15 sind die Pfade der Hypothesen dicker eingezeichnet. Um noch die anderen Kombinationsmöglichkeiten von primären Merkmalen zu Szenenmerkmalen testen zu können, wird in der Zeile 28 die letzte Assoziation aus der Liste der Assoziationen wieder entfernt. Anschließend werden weitere Assoziationen mit den anderen Szenenmerkmalen $\mathbf{s}_i \in \mathbf{S}_{tmp}$ aufgestellt. Sind keine weiteren Szenenmerkmale mit dem passenden Basisattribut vorhanden, dann geht man in der Struktur des Interpretationsbaumes um eine Stufe nach oben. Hierzu werden in den Zeilen 38 und 39 alle in der aktuellen Ebene verwendeten Szenenmerkmale und das primäre Merkmal der aktuellen Ebene wieder freigegeben.

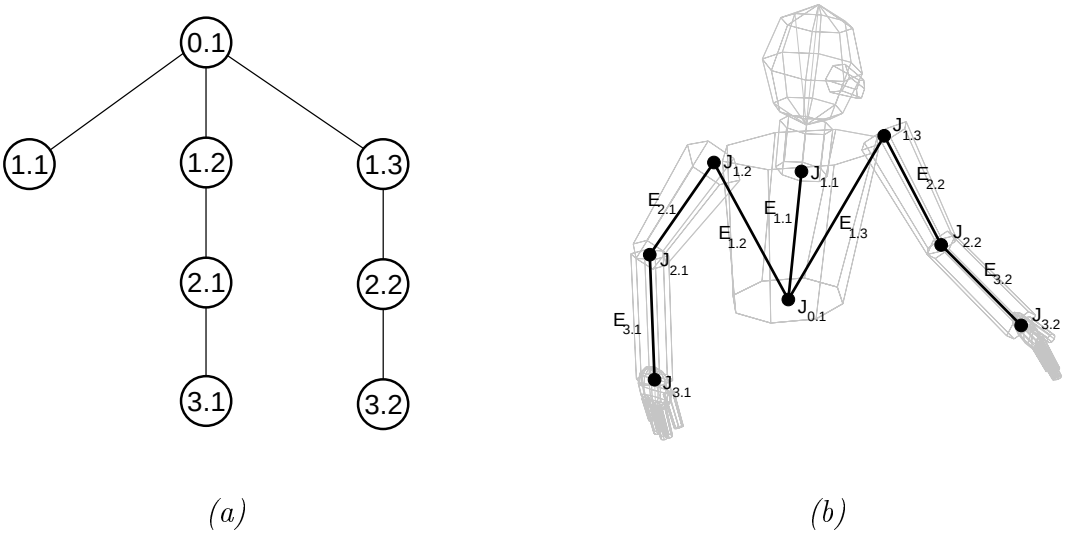
Wenn auf einer Ebene keine gültige Assoziation gefunden werden konnte, dann werden ebenfalls Szenenmerkmale und primäres Merkmal wieder freigegeben. Hierdurch erreicht man alle möglichen Kombinationen von Szenenmerkmalen und primären Merkmalen mit gleichem Basisattribut. Es sind alle Hypothesen aufgestellt, wenn man bei der Traversierung wieder auf der obersten Ebene angekommen ist und keine weiteren Szenenmerkmale mit dem Basisattribut des ersten primären Merkmals aus $\mathbf{F}$ zur Verfügung stehen.

Durch die Verwendung der Restriktionen wird eine komplette Traversierung des Interpretationsbaumes verhindert und nicht alle möglichen Assoziationen aufgebaut. Eine Bestimmung der Größe des Interpretationsbaumes und somit eine Angabe zur Komplexität der Suche nach Hypothesen ist in dem folgenden Abschnitt für die Beispielinterpretation angegeben.

Beispielinterpretation

Zur Darstellung des Aufbaus des Interpretationsbaumes anhand einer Beispielinterpretation wird eine Modellierung des menschlichen Körpers verwendet. Hierzu ist, gegenüber der im Kap. 2.3.3 eingeführten Standardmodellierung des menschlichen Körpers mit 16 Objektmodellteilen, ein Objektmodell verwendet worden, das nur aus acht Objektmodellteilen besteht, für die jeweils auch ein primäres Merkmal definiert ist. Hierbei sind Objektmodellteile für den Rumpf, den Hals, die beiden Oberarme, die beiden Unterarme und die beiden Hände verwendet worden, daher ergibt sich eine innere Objektmodellstruktur entsprechend Abb. 3.13. Aus dieser

[13] Die Knoten auf der untersten Ebene werden als Blätter bezeichnet.



(a) *(b)*

Abbildung 3.13: *Objektmodell für die Beispielinterpretation: (a) Baum der hierarchischen, inneren Struktur der Objektmodellteile und (b) Projektion der Struktur auf 3D Modell.*

Rumpf	0.1	grün	Hals	1.1	grün
rechter Oberarm	1.2	cyan	linker Oberarm	1.3	gelb
rechter Unterarm	2.1	gelb	linker Unterarm	2.2	cyan
rechte Hand	3.1	cyan	linke Hand	3.2	gelb

Tabelle 3.4: *Beispielinterpretation: Numerierung der Objektmodellteile und Farbattribut der primären Merkmale.*

inneren Objektmodellstruktur ergibt sich die Liste $\mathbf{F}$ der primären Merkmale:

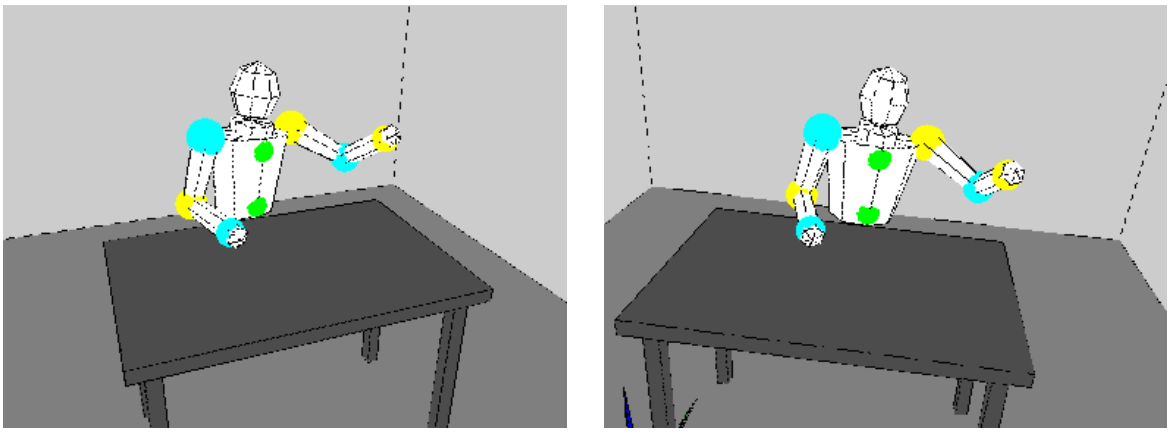
$$\mathbf{F} = \{\mathbf{f}_{0.1}, \mathbf{f}_{1.1}, \mathbf{f}_{1.2}, \mathbf{f}_{2.1}, \mathbf{f}_{3.1}, \mathbf{f}_{1.3}, \mathbf{f}_{2.2}, \mathbf{f}_{3.2}\}$$

Die geometrische Struktur entspricht der Struktur in der Standardmodellierung, vgl. Abb. 2.2. Es sind zudem für die Objektmodellteile die gleichen Volumenkörper gewählt worden. Zusätzlich sind hier jedoch den Objektmodellteilen primäre Merkmale in Form von farbigen Markierungen zugewiesen worden. In der Tab. 3.4 sind für die einzelnen Objektmodellteile die Numerierung und das Basisattribut der Farbe des zugeordneten Modell- / primären Merkmals aufgelistet. Gruppiert man die Modellmerkmale des Objektmodells entsprechend des Basisattributs der Farbe, so erhält man folgende drei Mengen von Modellmerkmalen:

$$\begin{aligned} \mathbf{M}^{(grün)} &= \{\mathbf{m}_{0.1}^{(grün)}, \mathbf{m}_{1.1}^{(grün)}\} \\ \mathbf{M}^{(gelb)} &= \{\mathbf{m}_{1.3}^{(gelb)}, \mathbf{m}_{2.1}^{(gelb)}, \mathbf{m}_{3.2}^{(gelb)}\} \\ \mathbf{M}^{(cyan)} &= \{\mathbf{m}_{1.2}^{(cyan)}, \mathbf{m}_{2.2}^{(cyan)}, \mathbf{m}_{3.1}^{(cyan)}\} \end{aligned} \tag{3.18}$$

Hierbei geben die Indizes die Nummern der Objektmodellteile an, zu derem primären Merkmal das Modellmerkmal zugehört.

Die Beispielinterpretation basiert auf Videobildern von zwei Kameras, so daß der Stereoansatz für die Bestimmung der 3D Szenenmerkmale verwendet wurde. Abb. 3.14 zeigt das erste Bildpaar einer Bildsequenz. Im folgenden ist der Aufbau des Interpretationsbaumes zunächst



(a) *(b)*

Abbildung 3.14: *Beispielinterpretation: Erstes Paar künstlich erzeugter Aufnahmen; (a) von der linken Kamera, (b) von der rechten Kamera; Die Farben der Markierungen der Tab. 3.4 zu entnehmen.*

für die initiale Detektion des Objektmodells auf diesem Bildpaar dargestellt. Daran schließt sich das Beispiel für die Re-Detektion an, wobei in dem verwendeten zweiten Bildpaar die aufgezeichnete Person den linken Arm nach unten gesenkt hat.

Initiale Detektion: Bei der initialen Detektion entspricht der Suchraum dem initialen Suchraum des Szenenmodells SSP^0, der mit einem genügend großen Radius gewählt worden ist, so daß die Bildverarbeitung auf dem kompletten Bild durchgeführt wird. Aufgrund des homogenen Hintergrundes und der homogenen Oberfläche des Modells in den künstlich erzeugten Aufnahmen,[14] wird mit dem Farbklassifikator in der Bildvorverarbeitung in jedem Bild exakt zwei grüne und jeweils drei gelbe und cyan-farbene Regionen segmentiert. Mit dem Stereoansatz sind aus den 2D Bildmerkmalen der beiden Bilder 3D Szenenmerkmale bestimmt worden, wobei man wiederum exakt zwei Szenenmerkmale mit dem Basisattribut der Farbe "grün" und jeweils drei mit dem Basisattribut der Farbe "gelb" und "cyan" erhält. Es können daher folgende drei Mengen von Szenenmerkmalen:

$$\begin{aligned} \mathbf{S}^{(grün)} &= \{\mathbf{s}_1^{(grün)}, \mathbf{s}_2^{(grün)}\} \\ \mathbf{S}^{(gelb)} &= \{\mathbf{s}_1^{(gelb)}, \mathbf{s}_2^{(gelb)}, \mathbf{s}_3^{(gelb)}\} \\ \mathbf{S}^{(cyan)} &= \{\mathbf{s}_1^{(cyan)}, \mathbf{s}_2^{(cyan)}, \mathbf{s}_3^{(cyan)}\} \end{aligned} \tag{3.19}$$

gebildet werden. Die Szenenmerkmale einer Menge unterscheiden sich jeweils in den 3D Positionen $\vec{p}_{wcs}$.

Von den im Abschn. 3.6.2 vorgestellten Restriktionen zur Überprüfung der Zulässigkeit der Assoziationen, werden in der hier dargestellten initialen Detektion nur die beiden Restriktionen $restr^{(parentD)}(.)$ und $restr^{(siblingD)}(.)$ verwendet. Die Restriktionen auf der Merkmalsebene greifen hier nicht: die 3D Punkte der Szenenmerkmale sind alle durch den Stereoansatz ermittelt worden, so daß die Qualität $q = 1$ ist, damit ist $restr^{(originQ)}(.)$ immer erfüllt. Durch die initiale Detektion sind sehr große Suchräume verwendet worden, so daß die 3D Positionen in diesen zu

[14] Diese Aufnahmen sind durch das System STABIL^{++} erstellt worden. Hierbei handelt es sich um eine 3D Darstellung des Objektmodells und dem Inventar / Weltregionen des Szenenmodells in einem OpenGL-Fenster. Die primären Merkmale der Objektmodellteile sind als farbige Kugeln dargestellt.

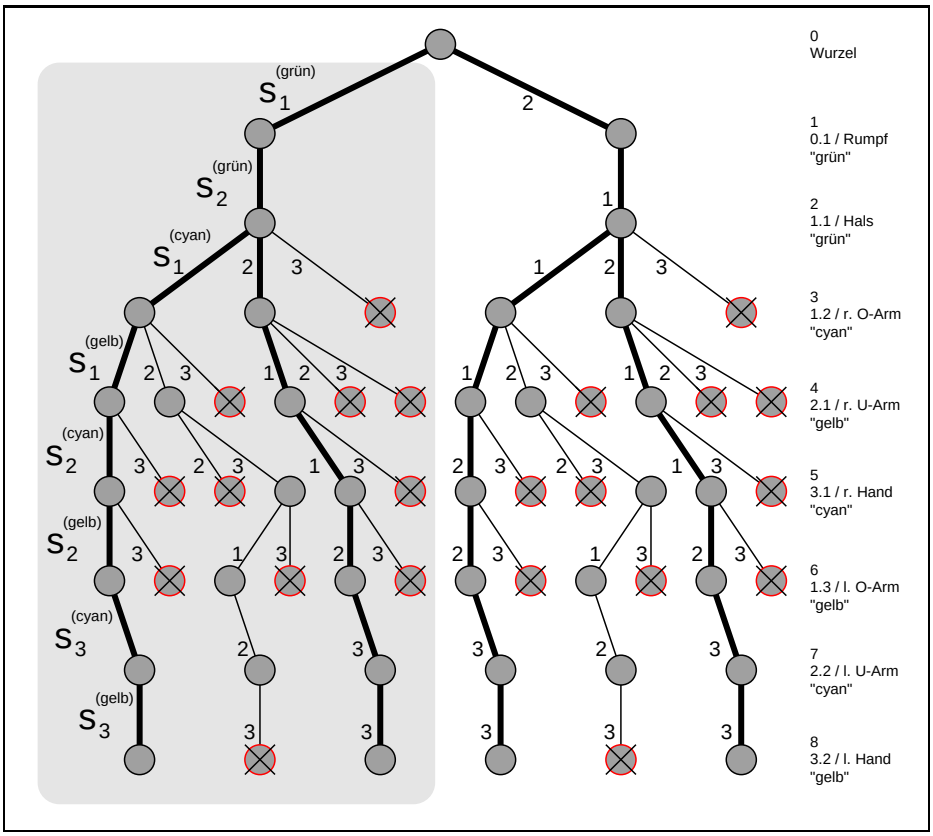


Abbildung 3.15: *Beispielinterpretation: Interpretationsbaum für initiale Detektion.*

liegen kommen, daher ist auch $restr^{(insideSsp)}(.)$ für alle Assoziationen erfüllt. Nachdem für die verwendeten primären Merkmale, außer dem Basisattribut der Farbe, keine weiteren Attribute definiert wurden, werden auch für die Szenenmerkmale keine weiteren Maßzahlen bestimmt; daher können die Restriktionen $restr^{(areaFnd)}(.)$ und $restr^{(exzentr)}(.)$ hier nicht angewendet werden.

In Abb. 3.15 ist der Interpretationsbaum für die initiale Detektion dargestellt. Zunächst wird in der Wurzelebene 0 für das erste primäre Merkmal $\mathbf{f}_{0.1} \in \mathbf{F}$ entsprechend des Basisattributes "grün" das Szenenmerkmal $\mathbf{s}_1^{(grün)} \in S(grün)$ ausgewählt. Nachdem es sich bei den zu verwendenden Restriktionen um mehrstellige Restriktionen handelt, können diese auf dieser Ebene des Interpretationsbaumes noch nicht angewendet werden. Daher ist die aufgestellte Assoziation eine gültige Zuordnung eines Szenenmerkmals zu dem primären Merkmal des Objektmodellteils des Rumpfes. Man erhält somit einen Knoten in der Ebene 1 des Interpretationsbaumes. Die Ebene ist entsprechend des Objektmodellteiles $omp_{0.1}$ gekennzeichnet.

Als nächstes soll für das primäre Merkmal $\mathbf{f}_{1.1}$ des Objektmodellteiles des Halses eine Assoziation aufgestellt werden. Nachdem $\mathbf{f}_{1.1}$ das gleiche Basisattribut wie $\mathbf{f}_{0.1}$ hat, steht in diesem Pfad nur noch das Szenenmerkmal $\mathbf{s}_2^{(grün)} \in S(grün)$ zur Verfügung. An den Kanten, die die Knoten des Interpretationsbaumes verbinden, sind jeweils die zugeordneten Szenenmerkmale vermerkt. Für diese Assoziation ist die Restriktion $restr^{(parentD)}(.)$ erfüllt und es wird ein Knoten auf der Ebene 2 des Interpretationsbaumes erzeugt. Die tertiäre Restriktion $restr^{(siblingD)}$ kann erst ab Ebene 3 des Interpretationsbaumes angewendet werden.

Nun wird für das dritte primäre Merkmal $\mathbf{f}_{1.2} \in \mathbf{F}$, das dem Objektmodellteil des rechten Oberarms zugehört, versucht, eine Assoziation zu finden. Entsprechend des Basisattributs

$cyan$ wird das Szenenmerkmal $\mathbf{s}_1^{(cyan)} \in \mathbf{S}(cyan)$ verwendet. Auch für diese Zuordnung ist die Restriktion $restr^{(parentD)}(.)$ erfüllt. Nachdem das Objektmodellteil $omp_{1.1}$ für den Hals das gleiche Vorgängerobjektmodellteil $omp_{0.1}$ wie das aktuell betrachtete Objektmodellteil $omp_{1.2}$ hat und schon Assoziationen für $\mathbf{f}_{0.1}$ und $\mathbf{f}_{1.1}$ aufgestellt worden sind, kann die Restriktion $restr^{(siblingD)}$ für die aktuelle Assoziation angewendet werden. Auch diese ist erfüllt, so daß ein Knoten in der Ebene 3 des Interpretationsbaumes eingetragen wird.

Es werden entsprechend des Alg. 3.6 alle möglichen Assoziationen aufgestellt und im Interpretationsbaum vermerkt. Die Knoten, die ungültige Assoziationen darstellen, sind durchgestrichen dargestellt. Man kann in der Abb. 3.15 erkennen, daß die Anzahl der Nachfolger in den tieferen Ebenen abnimmt. So können in Ebene 7 und 8 keine Szenenmerkmale mehr ausgewählt werden, sondern es muß das verbliebene letzte Szenenmerkmal mit dem entsprechenden Basisattribut verwendet werden. Weiterhin sind die Pfade, die von der Ebene 0 bis zur untersten Ebene 8 führen und bei denen alle Assoziationen gültig sind, mit dicken Kanten gekennzeichnet. Diese Pfade kennzeichnen damit die Hypothesen. In der dargestellten Beispielinterpretation sind vier Hypothesen gefunden worden.

Um eine bessere Vorstellung von der Zuordnung von 3D Szenenmerkmalen zu den 3D Modellmerkmalen der primären Merkmale zu bekommen, ist für den grau hinterlegten Teil des in Abb. 3.15 dargestellten Interpretationsbaumes in der Abb. 3.16 der entsprechende Ausschnitt des Interpretationsbaumes vergrößert dargestellt. In jedem Knoten ist das Bild der linken Kamera dargestellt. Diesen Bildern sind Projektionen der 3D Punkte der zugeordneten Szenenmerkmale als Kreise überlagert. Hierbei sind in einem Knoten nur die Kreise dargestellt, die den Szenenmerkmalen der Assoziationen entsprechen, die bis zu dem Knoten auf dem Weg von der Wurzelebene aufgestellt wurden. Die einzelnen projizierten Punkte sind entsprechend der inneren Objektmodellstruktur verbunden, so daß ein Eindruck der Lage der hypothetisch zugeordneten inneren Objektmodellstruktur entsteht.[15] Alle Knoten, die gültige Assoziationen repräsentieren, sind dunkler und dicker umrandet dargestellt. Bei den Knoten, die ungültige Assoziationen repräsentieren, ist die Projektion des 3D Punktes der zugehörigen Assoziation durchgestrichen dargestellt. An der Kennzeichnung rechts neben den Knoten ist zu erkennen, welche Restriktion nicht erfüllt werden konnte. 'P' steht für $restr^{(parentD)}(.)$ und 'S' für $restr^{(siblingD)}(.)$.[16]

Durch die Verwendung der Restriktionen wird die Anzahl der aufzustellenden Assoziationen reduziert, denn es werden in dem Interpretationsbaum nur die Pfade weiter verfolgt, die gültige Assoziationen beinhalten. Ohne die Anwendung der Restriktionen wäre der Suchraum für die Korrespondenzsuche wesentlich größer. Tab. 3.5 stellt für die vorgestellte Beispielinterpretation einige Maßzahlen vor, wobei die Angaben in Zeilen für jede Ebene des Interpretationsbaumes dargestellt sind. In der Spalte "Knoten" ist die Anzahl der tatsächlich aufgestellten Assoziationen eingetragen. Mit der Spalte "Nachf." ist vermerkt, wieviele Szenenmerkmale auf der folgenden Ebene noch zur Auswahl stehen. Die Anzahl der möglichen Nachfolgeknoten ist von der Anzahl der Szenenmerkmale eines Basisattributes abhängig und reduziert sich mit größer werdender Ebenennummer.

In der mit "max." gekennzeichneten Spalte ist die maximale Anzahl von Knoten vermerkt, die für eine Ebene aufgestellt werden kann, wenn keine Restriktionen greifen. Diese maximale Anzahl berechnet sich aus der Anzahl der Knoten in der vorhergehenden Ebene und der Anzahl der Nachfolger der vorhergehenden Ebene. Die maximale Anzahl der Knoten in der untersten Ebene gibt somit die maximale Anzahl von möglichen Hypothesen an. In der Beispielinterpre-

[15] Für die Modellierung des menschlichen Körpers entsteht zudem der Eindruck einer zugeordneten Knochenstruktur.

[16] Es ist zu beachten, daß die Restriktionen entsprechend Alg. 3.6 nacheinander überprüft werden, so daß mit der Kennzeichnung nur jeweils die erste Restriktion bezeichnet ist, die nicht erfüllt wurde.

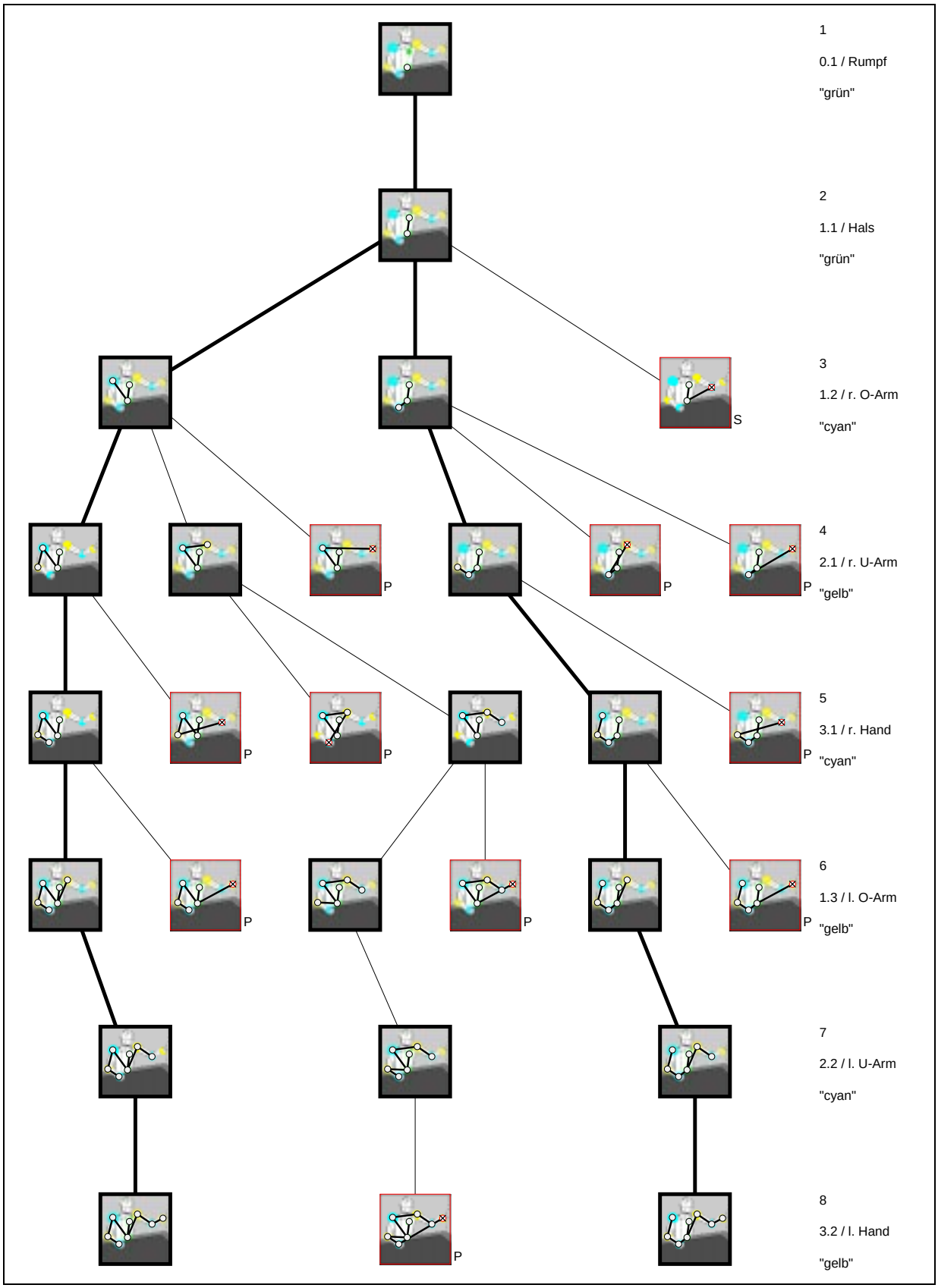


Abbildung 3.16: *Ausschnitt des Interpretationsbaums zur Beispielinterpretation mit Projektionen der zugeordneten 3D Szenenmerkmale, vgl. auch Abb. D.1.*

Ebene	Name	Attribut	Knoten	Nachf.	max.	reduz. %	gültig
0	Wurzel		1	2	1	0.000	1
1	Rumpf	grün	2	1	2	0.000	2
2	Hals	grün	2	3	2	0.000	2
3	r. O-Arm	cyan	6	3	6	0.000	4
4	r. U-Arm	gelb	12	2	18	33.333	6
5	r. Hand	cyan	12	2	36	66.667	6
6	l. O-Arm	gelb	12	1	72	83.333	6
7	l. U-Arm	cyan	6	1	72	91.667	6
8	l. Hand	gelb	6	0	72	91.667	4
Summe			59		281	79.004	37

Tabelle 3.5: *Beispielinterpretation: Maßzahlen zur Größe des Interpretationsbaumes bei der initialen Detektion.*

tation ist daher durch den Einsatz von den Restriktionen die Anzahl der Hypothesen von 72 auf 4 reduziert worden. Die Gesamtanzahl aller möglichen Assoziationen ist von 281 auf 59 aufgestellte Assoziationen reduziert worden, wobei hiervon 37 gültige Assoziationen sind. Mit der Spalte “reduz. %” ist angegeben, um wieviel pro Ebene die Anzahl der maximal möglichen Assoziationen reduziert wurde. Weitere Angaben zur Komplexität der Korrespondenzsuche im Interpretationsbaum werden am Ende dieses Abschnittes gemacht.

Re-Detektion: Bei der Re-Detektion der bei der initialen Detektion detektierten Objektmodellinstanz wird, im Gegensatz zur initialen Detektion, für jedes einzelne Objektmodellteil explizit ein 3D Suchraum bestimmt, vgl. Zeilen 6 - 9, Alg. 3.1. Da es sich hier um die erste Re-Detektion handelt, werden hier kugelförmige Suchräume bestimmt. Die Suchräume liegen um die Positionen der bei der initialen Detektion zugeordneten Szenenmerkmale, denn es können noch keine Vorhersagepositionen extrapoliert werden.

Durch diese Suchräume werden zum einen die Bildbereiche eingeschränkt, in denen die Bildverarbeitungsoperatoren angewendet werden. Die Suchräume sind hier so positioniert, daß alle Marken des Modells gefunden werden. Damit erhält man wieder drei Mengen von Szenenmerkmalen, vgl. Glg. 3.19. Zum anderen kann aufgrund der 3D Suchräume für die einzelnen Merkmale der Objektmodellteile beim Aufbau des Interpretationsbaumes die Restriktion $restr^{(insideSsp)}(.)$ zur Anwendung kommen.

Abb. 3.17 zeigt den Interpretationsbaum der Re-Detektion für die Beispielinterpretation. Zusätzlich zur Darstellung des Interpretationsbaumes der initialen Detektion in Abb. 3.16 sind hier die projizierten 3D Suchräume für die Merkmale der einzelnen Objektmodellteile eingezeichnet. Nachdem pro Ebene des Baumes für ein unterschiedliches Objektmodellteil die Zuordnung vorgenommen wird, unterscheiden sich die eingezeichneten 3D Suchräume entsprechend pro Ebene des Baumes.

Man erkennt, daß allein durch die Restriktion $restr^{(insideSsp)}(.)$ der Suchraum für die Zuordnungen soweit eingeschränkt werden konnte, daß sich in dem Beispiel anstelle der 72 maximal möglichen Hypothesen nur noch eine ergibt. Bei allen anderen Zuordnungen der 3D Szenenmerkmale zu 3D Modellmerkmalen mit dem entsprechenden Basisattribut lag die Position $\vec{p}_{wcs}$ des Szenenmerkmals außerhalb des Suchraumes. Neben den Knoten der ungültigen Assoziationen ist daher hier die Kennzeichnung ‘I’ entsprechend der Restriktion $restr^{(insideSsp)}(.)$ vermerkt.

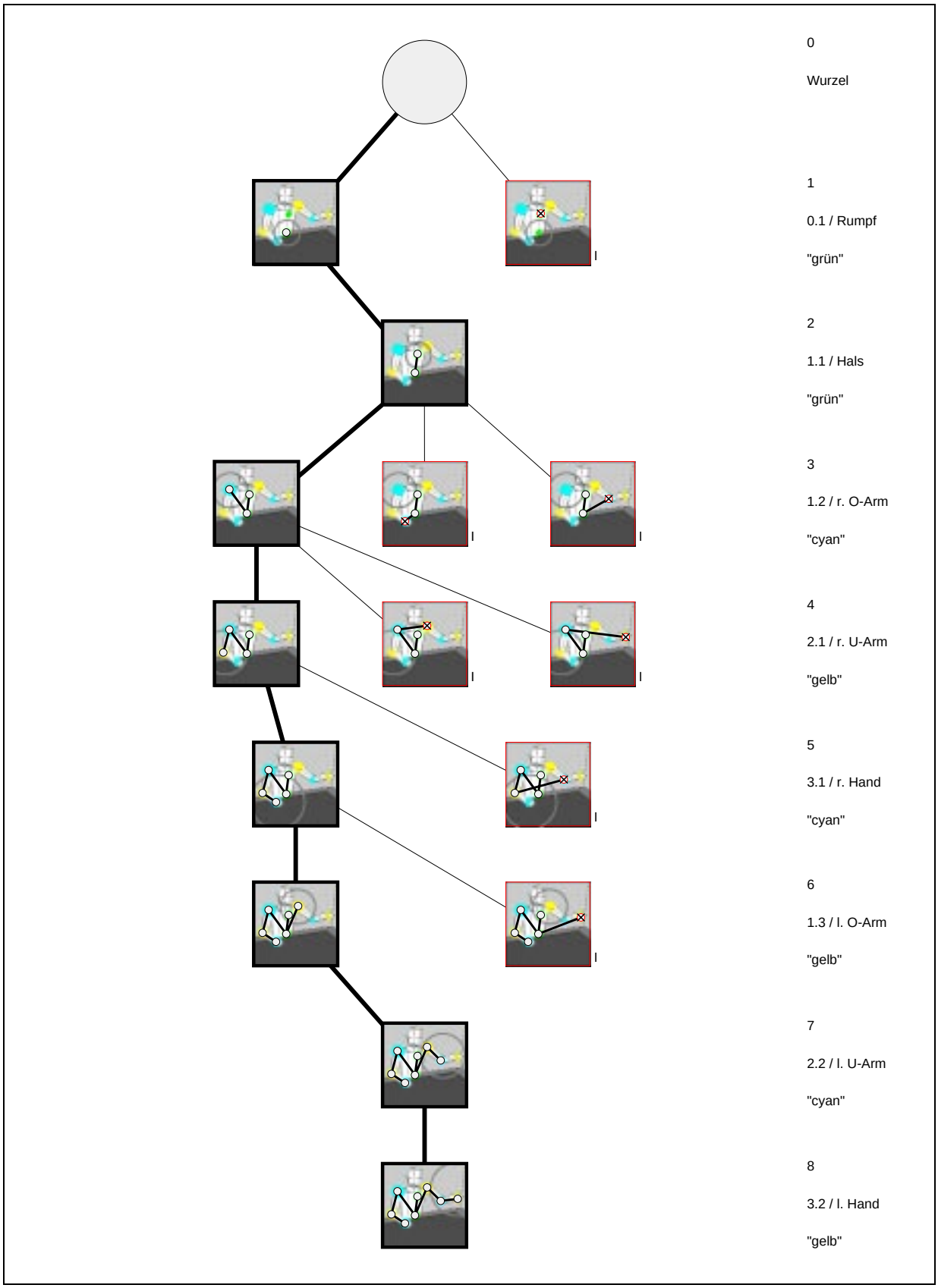


Abbildung 3.17: *Beispielinterpretation: Interpretationsbaum für Re-Detektion mit Projektionen der zugeordneten 3D Szenenmerkmale.*

Ebene	Name	Attribut	Knoten	Nachf.	max.	reduz. %	gültig
0	Wurzel		1	2	1	0.000	1
1	Rumpf	grün	2	1	2	0.000	1
2	Hals	grün	1	3	2	50.000	1
3	r. O-Arm	cyan	3	3	6	50.000	1
4	r. U-Arm	gelb	3	2	18	83.333	1
5	r. Hand	cyan	2	2	36	94.444	1
6	l. O-Arm	gelb	2	1	72	97.222	1
7	l. U-Arm	cyan	1	1	72	98.611	1
8	l. Hand	gelb	1	0	72	98.611	1
Summe			16		281	94.306	9

Tabelle 3.6: *Beispielinterpretation: Maßzahlen zur Größe des Interpretationsbaumes bei der Re-Detektion.*

In der Tab. 3.6 sind für den Interpretationsbaum der Re-Detektion die weiteren Maßzahlen aufgelistet. Es ist dort zu erkennen, daß von 281 möglichen Knoten nur noch 16 erzeugt werden, von denen 9 gültige Assoziationen darstellen.

3.6.4 Fehlende / falsche Szenenmerkmale

In dem vorhergehenden Abschnitt ist der Aufbau des Interpretationsbaumes beschrieben worden, bei dem einer gleichen Anzahl von Szenenmerkmalen eine gleiche Anzahl von Modellmerkmalen mit jeweils passendem Basisattribut gegenüberstanden. Zudem konnten alle Restriktionen erfüllt werden, so daß mindestens eine komplette Zuordnung der Szenenmerkmale zu den Modellmerkmalen vorgenommen werden konnte. In diesem Abschnitt soll nun betrachtet werden, wie der Aufbau des Interpretationsbaumes abgeändert werden muß, damit bei fehlenden oder zusätzlichen, aber nicht zuordnenbaren Szenenmerkmalen Hypothesen generiert werden können.

Mit dem bisher vorgestellten Verfahren zum Aufbau des Interpretationsbaumes kann ein Ast nicht weiter verfolgt werden, wenn die Restriktionen für die aufgestellten Assoziationen nicht erfüllt sind und wenn keine weiteren Szenenmerkmale mit dem entsprechenden Basisattribut zur Verfügung stehen. In einem solchen Ast kann damit keine Hypothese generiert werden. Geht man davon aus, daß in diesen Ästen jedoch alle weiteren Zuordnungen für die Objektmodellteile auf den tieferen Ebenen korrekt sind, verwirft man mit der bisherigen Vorgehensweise eine fast komplette Hypothese. Falls ein Szenenmerkmal, aufgrund von z.B. Eigen- oder Fremdverdeckung, bei der Verfolgung eines Objekts zeitweilig nicht zu detektieren ist, dann müßte die komplette Objektmodellinstanz verworfen werden.

Auch in [Gri90a] ist diese Problematik angesprochen und daher sog. *null character features* eingeführt worden, für die keine realen Daten vorhanden sind.[17] Auch in STABIL^{++} werden Szenenmerkmale verwendet, bei denen die 3D Position $\vec{p}_{wcs}$ unbestimmt bleibt. Diese Szenenmerkmale werden, in Analogie zu einem nicht existierenden Element einer Liste, als *nil*-Szenenmerkmale $\mathbf{s}_{nil}$ bezeichnet.[18] Mit den nil-Szenenmerkmalen hat man den Vorteil, daß auch bei fehlenden Szenenmerkmalen Hypothesen generiert werden können, jedoch bleiben für eins

[17] Grimson spricht von Merkmalen mit sog. *spurious data*.
[18] nil = *not in list* (engl.).

oder mehrere Objektmodellteile die 3D Positionen unbestimmt.

In STABIL^{++} werden über die Bestimmung der Suchbereiche 3D Positionsvorhersagen erstellt. Verwendet man die Vorhersagen als Schätzung einer 3D Position, so erhält man für ein Objektmodellteil ein *geschätztes* Szenenmerkmal $\hat{\mathbf{s}}$. Aufgrund der Unsicherheiten, denen die Schätzungen unterliegen, wird mit Einführung der geschätzten Szenenmerkmale nicht auf die nil-Szenenmerkmale verzichtet. Dies liegt darin begründet, daß ein geschätztes Szenenmerkmal bei schlechter Qualitätsbewertung ebenfalls verworfen werden kann oder die Schätzung ungenau ist, womit die entsprechenden Assoziationen durch die Restriktionen schlecht bewertet oder verworfen werden. Es werden daher beide nicht realen Szenenmerkmale immer paarweise verwendet. Hypothesen, die sich nur in der Assoziation eines Objektmodellteils durch die Verwendung von nil-Szenenmerkmal und geschätztem Szenenmerkmal unterscheiden, sind bei der Hypothesenbewertung unterschiedlich zu behandeln. Es wird hierbei der Hypothese mit dem geschätzten Szenenmerkmal der Vorrang gegeben, sofern die geschätzte Position sinnvoll ist, ansonsten wird die Hypothese mit dem nil-Szenenmerkmal verwendet.

Diese nicht realen Szenenmerkmale werden ausschließlich für die Re-Detektion verwendet. Dies begründet sich zum einen darin, daß eine Positionsvorhersage / explizite Suchraumbestimmung nur für eine Objektmodellinstanz angefertigt wird, jedoch nicht für das initiale Objektmodell. Desweiteren können mit diesen Szenenmerkmalen weniger Restriktionen mit entsprechender Aussagekraft angewendet werden, womit der Suchraum für die Korrespondenzfindung wesentlich größer wird. Dies betrifft zunächst alle Restriktionen auf der Merkmalsebene, die alle weiteren Attribute, außer den Basisattributen überprüfen. Für diese Restriktionen fehlen die 2D Bildmerkmale, die für reale Szenenmerkmale extrahiert werden und in $\mathbf{I}_{extr}$ bekannt sind.

Durch die fehlende 3D Position bei den nil-Restriktionen können weiterhin die Restriktionen $restr^{(originQ)}(.)$ und $restr^{(insideSsp)}(.)$ nicht angewendet werden. Bei den geschätzten Szenenmerkmalen $\hat{\mathbf{s}}$ liegt die 3D Position aufgrund der Bestimmung des Suchraumes mittels der Positionsvorhersage immer innerhalb des Suchraumes, so daß hier $restr^{(insideSsp)}(.)$ immer erfüllt ist. Somit kann von den Restriktionen auf Merkmalsebene nur die Restriktion $restr^{(originQ)}(.)$ für die geschätzten Szenenmerkmale $\hat{s}$ angewendet werden. Als Qualitätsmaß wird hier die Qualität der als Schätzung verwendeten 3D Positionsvorhersage verwendet, vgl. Kap. 3.2.3.

Nachdem für nil-Szenenmerkmale keine 3D Position bestimmt ist, wird durch diese auch die Anwendung der Restriktionen auf der Modellebene ($restr^{(parentD)}(.)$, $restr^{(siblingD)}$) eingeschränkt. So kann für die Assoziation mit einem nil-Szenenmerkmal nicht der Abstand zu dem zugehörigen Vorgängerobjektmodellteil bestimmt werden. Ferner können alle Restriktionen, die den Geschwister-Abstand überprüfen, nicht angewendet werden, falls hierbei die Assoziation mit dem nil-Szenenmerkmal durch die Restriktionen zu anderen Assoziationen in Relation gesetzt wird.

Die Verwendung der nil-Szenenmerkmale und der geschätzten Szenenmerkmale wird im folgenden anhand der im vorhergehenden Abschnitt dargestellten Re-Detektion der Beispielinterpretation erläutert. Zunächst wird die Interpretation bei einem fehlenden Szenenmerkmal beschrieben, für die einzelne ‘leere’ und ‘geschätzte’ Knoten im Interpretationsbaum erzeugt werden. Daran schließt sich die Betrachtung von komplett ‘leeren’ / ‘geschätzten’ Ebenen im Interpretationsbaum an. Schließlich wird noch ein komplett ‘geschätzter’ Baum betrachtet.

‘Leere’ / ‘geschätzte’ Knoten

Stehen für die Zuordnung der Modellmerkmale $\mathbf{m}_i^{(attr)} \in \mathbf{M}^{(attr)}$, $i = 1 \ldots n$ mit dem Basisattribut $attr$ weniger Szenenmerkmale $\mathbf{s}_j^{(attr)} \in \mathbf{S}^{(attr)}$, $j = 1 \ldots k$ zur Verfügung, d.h. $n < k$, so werden $n - k$ nil-Szenenmerkmale und geschätzte Szenenmerkmale erzeugt und $\mathbf{S}^{(attr)}$ hinzu-

gefügt:

$$\mathbf{S}^{(attr)} = \mathbf{S}^{(attr)} \cup \{\mathbf{s}_{nil_1} \ldots \mathbf{s}_{nil_m}\} \cup \{\hat{\mathbf{s}}_1, \ldots, \hat{\mathbf{s}}_m\},\ m = n - k$$

Dieses Hinzufügen der nicht realen Szenenmerkmale wird für alle verschiedenen Basisattribute der Modellmerkmale der zu re-detektierenden Objektmodellinstanz durchgeführt. Für den Aufbau des Interpretationsbaumes entsprechend Alg. 3.6, sind dann die Listen der primären Merkmale $\mathbf{F}$ und der Szenenmerkmale gleich mächtig. Aus den nil-Szenenmerkmalen können direkt sog. 'leere' Knoten erzeugt werden. Für die geschätzten Szenenmerkmale muß zur Erzeugung eines sog. 'geschätzten' Knotens des Interpretationsbaumes die 3D Position $\vec{p}_{wcs}$ gesetzt werden. Hierzu muß Alg. 3.6 zwischen Zeile 11 und 12 ergänzt werden. Dort muß $\vec{p}_{wcs}$ mit der Vorhersage $\tilde{\vec{p}}$ für das Objektmodellteil, dem das primäre Merkmal $\mathbf{f}$ zugeordnet ist, gesetzt werden.

Zur Erläuterung des Einsatzes von 'leeren' und 'geschätzten' Knoten im Interpretationsbaum sind bei der Re-Detektion der Objektmodellinstanz der Beispielinterpretation die 2D Bildmerkmale des Objektmodellteiles des linken Unterarms $omp_{2.2}$ in dem zu verwendenden Bildpaar nicht sichtbar.[19] Dementsprechend konnten nur zwei 3D Szenenmerkmale mit dem Basisattribut der Farbe "cyan" erzeugt werden. Es wird daher die Menge der Szenenmerkmale mit dem Basisattribut der Farbe "cyan" um ein nil-Szenenmerkmal und ein geschätztes Szenenmerkmal erweitert. Daher gilt für die Menge der Szenenmerkmale aus Glg. 3.19:

$$\begin{aligned} \mathbf{S}^{(grün)} &= \{\mathbf{s}_1^{(grün)}, \mathbf{s}_2^{(grün)}\} \\ \mathbf{S}^{(gelb)} &= \{\mathbf{s}_1^{(gelb)}, \mathbf{s}_2^{(gelb)}, \mathbf{s}_3^{(gelb)}\} \\ \mathbf{S}^{(cyan)} &= \{\mathbf{s}_1^{(cyan)}, \mathbf{s}_2^{(cyan)}, \mathbf{s}_{nil}^{(cyan)}, \hat{\mathbf{s}}^{(cyan)}\} \end{aligned} \tag{3.20}$$

Der Aufbau des Interpretationsbaumes verläuft zunächst wie in der Darstellung in Abb. 3.17, bis beim ersten Erreichen der Ebene 6 keine weiteren realen Szenenmerkmale für eine Assoziation mit dem primären Merkmal des Objektmodellteils $omp_{2.2}$ des linken Unterarms auf der Ebene 7 zur Verfügung stehen. Es wird daher zunächst eine Assoziation mit dem nil-Szenenmerkmal erzeugt. Für diese Assoziation kann keine Restriktion angewendet werden, so daß diese gültig ist. Es wird daher anschließend in der Ebene 8 eine Assoziation für das Objektmodellteil der linken Hand mit dem Szenenmerkmal $\mathbf{s}_3^{(gelb)}$ erzeugt. Für diese Assoziation der Ebene 8 kann die Restriktion $restr^{(parentD)}(.)$ nicht angewendet werden, da für das Vorgängerobjektmodellteil des zugehörigen Objektmodellteils die Assoziation mit dem nil-Szenenmerkmal erstellt wurde. Nachdem alle weiteren Restriktionen erfüllt sind und die unterste Ebene des Interpretationsbaumes erreicht worden ist, ist eine Hypothese gefunden, vgl. Abb. 3.18.[20]

In der Abb. 3.19 (a) sind die Zuordnungen dieser Hypothese dargestellt. Hierbei sind in einem Ausschnitt des Bildes der linken Kamera die projizierten 3D Punkte der zugeordneten Szenenmerkmale dargestellt. Anhand der Verbindungen zwischen den Punkten entsprechend der hierarchischen, inneren Objektmodellstruktur ist die hypothetische zugeordnete innere Objektmodellstruktur dargestellt. Durch die Verwendung des nil-Szenenmerkmals für die Assoziation des Objektmodellteils $omp_{2.2}$ des linken Unterarms, hat die Projektion für das Objektmodellteil $omp_{3.2}$ der linken Hand in der Abb. 3.19 (a) keine Verbindung zum restlichen Objektmodell.

[19] Bei der Generierung der verwendeten künstlichen Aufnahmen ist für die Markierung am Objektmodellteil des linken Unterarms die Farbe "weiß" verwendet worden.

[20] Die Projektionen der nil-Szenenmerkmale $\mathbf{s}_{nil}$ sind weiß und mit 'gestrichelter' Umrandung dargestellt, die Projektionen der geschätzten Szenenmerkmale $\hat{\mathbf{s}}^{(cyan)}$ sind dunkler als die nil-Szenenmerkmale und heller als die realen Szenenmerkmale dargestellt. Zudem sind diese mit einer 'gepunkteten' Umrandung dargestellt. Die Kanten zwischen den Knoten sind entsprechend gekennzeichnet.

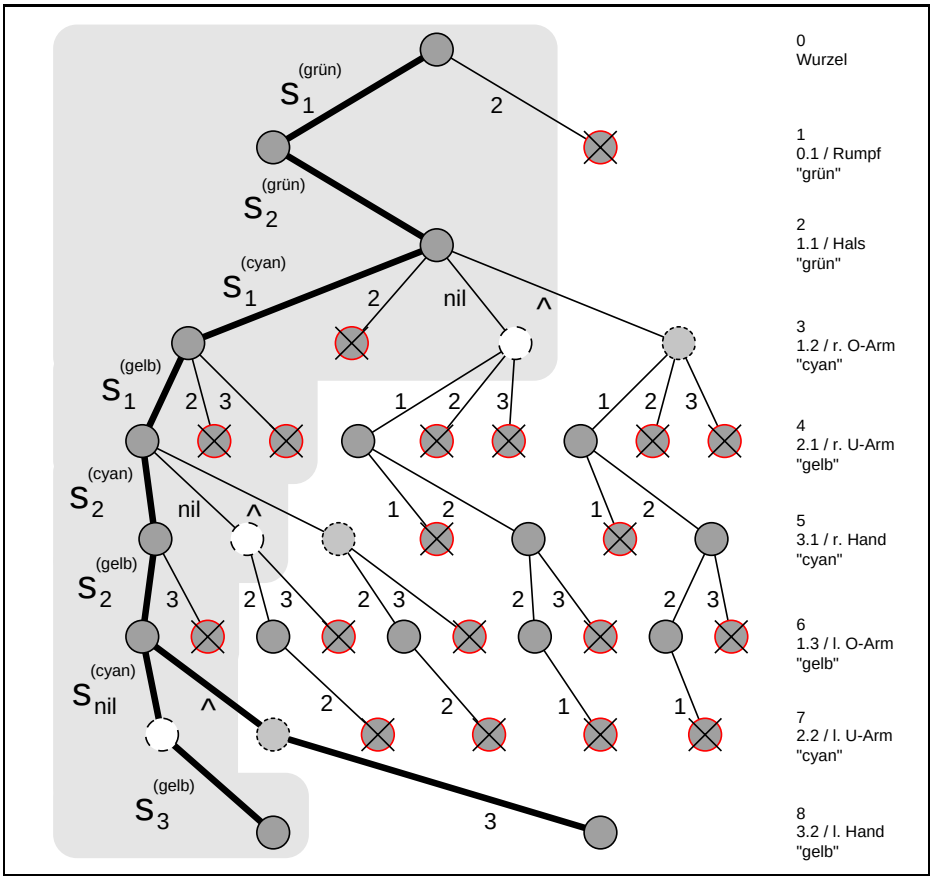
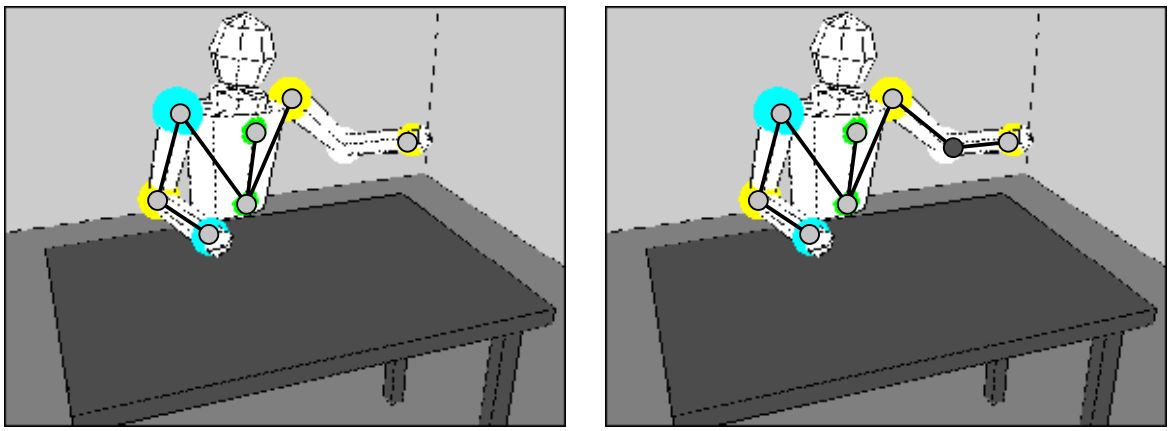


Abbildung 3.18: *Interpretationsbaum für Re-Detektion der Beispielinterpretation bei fehlendem Szenenmerkmal.*



(a) *(b)*

Abbildung 3.19: *Hypothesen bei fehlendem Szenenmerkmal: (a) Verwendung von 'leerer' Assoziation mit nil-Szenenmerkmal* $\mathbf{s}_{nil}$ *und (b) 'geschätzter' Assoziation mit Szenenmerkmal* $\hat{\mathbf{s}}$*; jeweils einem Ausschnitt des Bildes der linken Kamera überlagert.*

Ebene	Name	Attribut	Knoten	Nachf.	max.	reduz. %	gültig
0	Wurzel		1	2	1	0.000	1
1	Rumpf	grün	2	1	2	0.000	1
2	Hals	grün	1	4	2	50.000	1
3	r. O-Arm	cyan	4	3	8	50.000	3
4	r. U-Arm	gelb	9	3	24	62.500	3
5	r. Hand	cyan	7	2	72	90.278	5
6	l. O-Arm	gelb	10	2	144	93.056	5
7	l. U-Arm	cyan	6	1	288	97.917	2
8	l. Hand	gelb	2	0	288	99.306	2
Summe			42		829	94.934	23

Tabelle 3.7: *Maßzahlen zur Größe des Interpretationsbaumes für Re-Detektion der Beispielinterpretation bei fehlendem Szenenmerkmal.*

Auf der Ebene 7 des Interpretationsbaumes wird im weiteren noch eine Assoziation mit dem geschätzten Szenenmerkmal $\hat{\mathbf{s}}^{(cyan)}$ erzeugt. Hierbei wird die 3D Position $\vec{p}_{wcs}$ von $\hat{s}^{(cyan)}$ mit der für das primäre Merkmal von $omp_{2.2}$ vorhergesagten 3D Position $\tilde{\vec{p}}$ als Schätzung gesetzt. Gegenüber der Assoziation mit dem nil-Szenenmerkmal, kann hier und in der folgenden Assoziation in der Ebene 8 die Restriktion $restr^{(parentD)}(.)$ angewendet werden. Nachdem diese und die weiteren Restriktionen für diese Assoziation auf der Ebene 7 und für die folgende Restriktion erfüllt sind, ergibt sich aus dem Pfad mit diesen Assoziationen eine weitere Hypothese.

Die Zuordnungen dieser Hypothese sind in der Abb. 3.19 (b) dargestellt. Im Vergleich zur Abb. 3.19 (a) ist hier eine Projektion für ein dem primären Merkmal des Objektmodell $omp_{2.2}$ zugeordneten Szenenmerkmal eingezeichnet. Für die Projektion ist hierbei die geschätzte Position verwendet worden; zur Unterscheidung ist der Punkt dunkler als die Punkte der projizierten Positionen der realen Szenenmerkmale dargestellt.

Entsprechend der Struktur des Interpretationsbaumes werden auch auf allen anderen Ebenen, in denen die Zuordnungen für primäre Merkmale mit dem Basisattribut die Farbe “cyan” vorgenommen wird, die beiden nicht realen Szenenmerkmale zugeordnet. Das nil-Szenenmerkmal kann wieder direkt verwendet werden, wobei für das geschätzte Szenenmerkmal noch die 3D Position gesetzt werden muß. Auf der Ebene 5 wird hierzu die Vorhersage für das Objektmodellteil $obj_{3.1}$ der rechten Hand und auf der Ebene 3 die Vorhersage für das Objektmodellteil $obj_{1.2}$ des rechten Oberarms verwendet.

Nachdem die geschätzten Positionen für die Objektmodellteile auf den Ebenen 3, 5 und 7 des Interpretationsbaumes innerhalb der Toleranzgrenzen der Restriktionen auf der Modellebene $restr^{(parentD)}(.)$ und $restr^{(siblingD)}(.)$ liegen, repräsentieren alle ‘leeren’ und ‘geschätzte’ Knoten des Interpretationsbaumes der Beispielinterpretation gültige Assoziationen, so daß sich die Anzahl der zu erzeugenden Knoten erhöht. Für einen Vergleich der Größe der Interpretationsbäume ist in der Abb. 3.18 der Teil des Interpretationsbaumes grau hinterlegt, der der Größe des Interpretationsbaumes der Re-Detektion entspricht, bei der nur reale Szenenmerkmale verwendet wurden, vgl. Abb. 3.17.

In Tab. 3.7 sind die weiteren Maßzahlen für den Interpretationsbaum bei fehlendem Szenenmerkmal für das Objektmodellteil $omp_{2.2}$ des linken Unterarms dargestellt. Im Vergleich zur Tab. 3.6 sieht man, daß hier anstelle von 72 möglichen Hypothesen 288 Hypothesen möglich sind. Die Gesamtanzahl von möglichen Knoten ist von 281 auf 829 gestiegen. Durch die An-

wendung der Restriktionen konnte jedoch die Anzahl der Hypothesen auf zwei begrenzt werden, wobei 42 Assoziationen, gegenüber 16 Assoziationen bei der Re-Detektion mit realen Szenenmerkmalen, aufgestellt wurden.

'Leere' / 'geschätzte' Ebene

Ist die Liste der primären Merkmale $\mathbf{F}$ und die Liste der Szenenmerkmale $\mathbf{S}$, die im Interpretationsbaum verwendet werden, gleich mächtig, so werden zunächst keine nil- und geschätzten Szenenmerkmale in $\mathbf{S}$ hinzugefügt. Fehlt in $\mathbf{S}$ jedoch ein passendes Szenenmerkmal für ein primäres Merkmal und ist somit ein nicht zu verwendendes Szenenmerkmal bestimmt worden, so schlägt die Generierung der Hypothesen zunächst fehl. Die längsten Pfade des Interpretationsbaums enden dann auf der Ebene, in der, aufgrund des fehlenden Szenenmerkmals keine gültige Assoziation mehr erzeugt werden konnte. Wird im Alg. 3.6 bei der Re-Detektion einer Objektmodellinstanz das Ende erreicht, ohne daß die maximal unterste Ebene / größt mögliche Tiefe erreicht wurde, so konnte dem primären Merkmal eines Objektmodellteils kein Szenenmerkmal zugewiesen werden. Mit der untersten erreichten Ebene ist das Objektmodellteil bekannt, in dem die Interpretation abgebrochen ist.

Für dieses Objektmodellteil müssen daher nicht reale Szenenmerkmale zugelassen werden. Es wird die Menge der Szenenmerkmale mit dem entsprechenden Basisattribut $attr$ um die Szenenmerkmale $\mathbf{s}_{nil}^{(attr)}$ und $\hat{\mathbf{s}}^{(attr)}$ erweitert. Anschließend müßte man auf der höchsten Ebene, in der Merkmale mit dem entsprechenden Basisattribut zugeordnet werden, mit dem Aufbau des Interpretationsbaumes erneut beginnen. Damit würden alle möglichen Assoziationen von primären Merkmalen zu den realen Modellmerkmalen und den zusätzlich hinzugefügten, nicht realen Szenenmerkmalen mit dem entsprechenden Basisattribut, aufgebaut und getestet. Die Gesamtanzahl der aufzustellenden Assoziationen steigt somit an, vgl. Beispiel für die fehlenden Szenenmerkmale im letzten Abschnitt.

Dieser zusätzliche Aufwand kann jedoch eingeschränkt werden. Hierzu wird eine Zuordnung der hinzugefügten, nicht realen Szenenmerkmale auf das primäre Merkmale des Objektmodellteils beschränkt, das der Ebene entspricht, die der Interpretationsbaum bisher maximal erreichen konnte. Man spricht daher von einer 'leeren' / 'geschätzten' Ebene im Interpretationsbaum. In den Ebenen, die oberhalb dieser Ebene liegen und in denen Merkmale mit dem gleichen Basisattribut verwendet werden, kann auf die Zuordnung der nicht realen Szenenmerkmale verzichtet werden. Dies begründet sich darin, daß eine weitere, bisher noch nicht aufgestellte, Assoziation nur dann gültig ist, wenn die 3D Position der Assoziation sehr nahe an der 3D Position einer bereits aufgestellten Assoziation liegt. Dies ist wiederum durch die verwendeten Restriktionen $restr^{(insideSsp)}(.)$, $restr^{(parentD)}(.)$ und $restr^{(siblingD)}(.)$ begründet. Man würde somit sehr ähnliche Hypothesen erhalten, die sich nur innerhalb der 3D Position des lokalen Koordinatensystems eines Objektmodells unterscheiden. Dieser Unterschied liegt innerhalb der Grenzwerte, die bei den Restriktionen angewendet werden. Bei der Bewertung dieser Hypothesen kann ebenfalls fast kein Qualitätsunterschied festgestellt werden, so daß, zugunsten einer geringeren Komplexität des Interpretationsbaumes, diese geringe Ungenauigkeit in der 3D Position eines Objektmodellteils in Kauf genommen wird. In den Ebenen, die nach dem Einziehen der 'leeren' / 'geschätzten' Ebene nun unterhalb dieser Ebene erreicht werden können, können diese Szenenmerkmale generell nicht verwendet werden, da sie in der eingezogenen Ebene schon zugeordnet wurden.

Die Verwendung der 'leeren' / 'geschätzten' Ebene soll auch hier wieder an der eingeführten Beispielinterpretation erläutert werden. Hierzu sind, wie bei der Darstellung der 'fehlenden' Szenenmerkmale, in dem zu verwendenden Bildpaar die 2D Bildmerkmale des Ob-

Ebene	Name	Attribut	Knoten	Nachf.	max.	reduz. %	gültig
0	Wurzel		1	2	1	0.000	1
1	Rumpf	grün	2	1	2	0.000	1
2	Hals	grün	1	3	2	50.000	1
3	r. O-Arm	cyan	3	3	6	50.000	1
4	r. U-Arm	gelb	3	2	18	83.333	1
5	r. Hand	cyan	2	2	36	94.444	2
6	l. O-Arm	gelb	4	1	72	94.444	2
7 (I)	l. U-Arm	cyan	2	0	72	97.222	0
7 (II)	l. U-Arm	cyan	4	1	144	97.222	4
8	l. Hand	gelb	4	0	144	97.222	4
Summe			24 + 2		425	94.353	17

Tabelle 3.8: *Falsches Szenenmerkmal: Maßzahlen zur Größe des Interpretationsbaumes, wobei für die Ebene 7 in einem zweiten Durchlauf jeweils 'leere' und 'geschätzte' Knoten verwendet werden.*

jektmodellteiles des linken Unterarms $omp_{2.2}$ nicht sichtbar. Es stehen damit für die Zuordnung der primären Merkmale mit dem Basisattribut der Farbe "cyan" die Szenenmerkmale $\mathbf{S}^{(cyan)} = \{\mathbf{s}_1^{(cyan)}, \mathbf{s}_2^{(cyan)}\}$ zur Verfügung, vgl. Glg. 3.20. Zusätzlich können in dem Bildpaar innerhalb des projizierten 3D Suchraumes des Objektmodellteils $omp_{3.1}$ der rechten Hand noch weitere Bildmerkmale mit dem Attribut der Farbe "cyan" extrahiert werden. Somit stehen drei Szenenmerkmale mit dem Basisattribut der Farbe "cyan" zur Verfügung. Man erhält also ein zusätzliches drittes Szenenmerkmal $\mathbf{s}'^{(cyan)}_3 \in \mathbf{S}^{(cyan)}$. Aufgrund der Restriktionen kann für das zusätzliche Szenenmerkmal $\mathbf{s}'^{(cyan)}_3$ auf der Ebene 5 des Interpretationsbaumes eine weitere gültige Zuordnung erzeugt werden, vgl. Abb. 3.20.

Verfolgt man diese beiden Äste weiter, so wird für die Ebene 7 versucht, das jeweils andere Szenenmerkmal, das innerhalb des Suchraumes des Objektmodellteils $omp_{3.1}$ der rechten Hand liegt, für das Objektmodellteil $omp_{2.2}$ des linken Unterarms zuzuordnen. Die entsprechenden 3D Positionen der Szenenmerkmale liegen jedoch nicht innerhalb des Suchraumes des Objektmodellteils $omp_{2.2}$ des linken Unterarms, so daß die Assoziationen mit der Restriktion $restr^{(insideSsp)}(.)$ verworfen werden. Es wird damit der Aufbau des Interpretationsbaumes auf der Ebene 7 abgebrochen. Für diese Ebene wird, entsprechend der o.a. Ausführungen die Verwendung von 'leeren' und 'geschätzten" Knoten erzwungen. In der Abb. 3.21 sind die untersten drei Ebenen für den zweiten Durchlauf des Interpretationsbaumes abgebildet, wobei die 'gestrichelten' Knoten in der Ebene 7 Assoziationen mit nil-Szenenmerkmalen und die 'gepunkteten' Knoten Assoziationen mit geschätzten Szenenmerkmalen darstellen.

Man kann erkennen, daß bei der Verwendung der 'leeren' / 'geschätzten' Ebene mit dem Erreichen der Ebene 8 nun vier Hypothesen gefunden werden. Die Anzahl der maximal möglichen Hypothesen hat sich durch das Einziehen der 'leeren' / 'geschätzten' Ebene auf 144 erhöht, dies entspricht einer Verdoppelung gegenüber der Re-Detektion, bei der nur reale Szenenmerkmale verwendet werden, vgl. Tab. 3.6. Weitere Maßzahlen für die Re-Detektion der Beispielinterpretation mit 'leerer' / 'geschätzter' Ebene sind in Tab. 3.8 aufgelistet. Dort sind für die Ebene 7 zwei Zeilen eingetragen, wobei die zweite Zeile der 'leeren' / 'geschätzten' Ebene entspricht. Die Gesamtanzahl der aufgestellten Knoten beinhaltet daher zusätzlich die zwei Knoten für die ungültigen Assoziationen, die beim ersten Durchlauf für die Ebene 7 aufgestellt wurden.

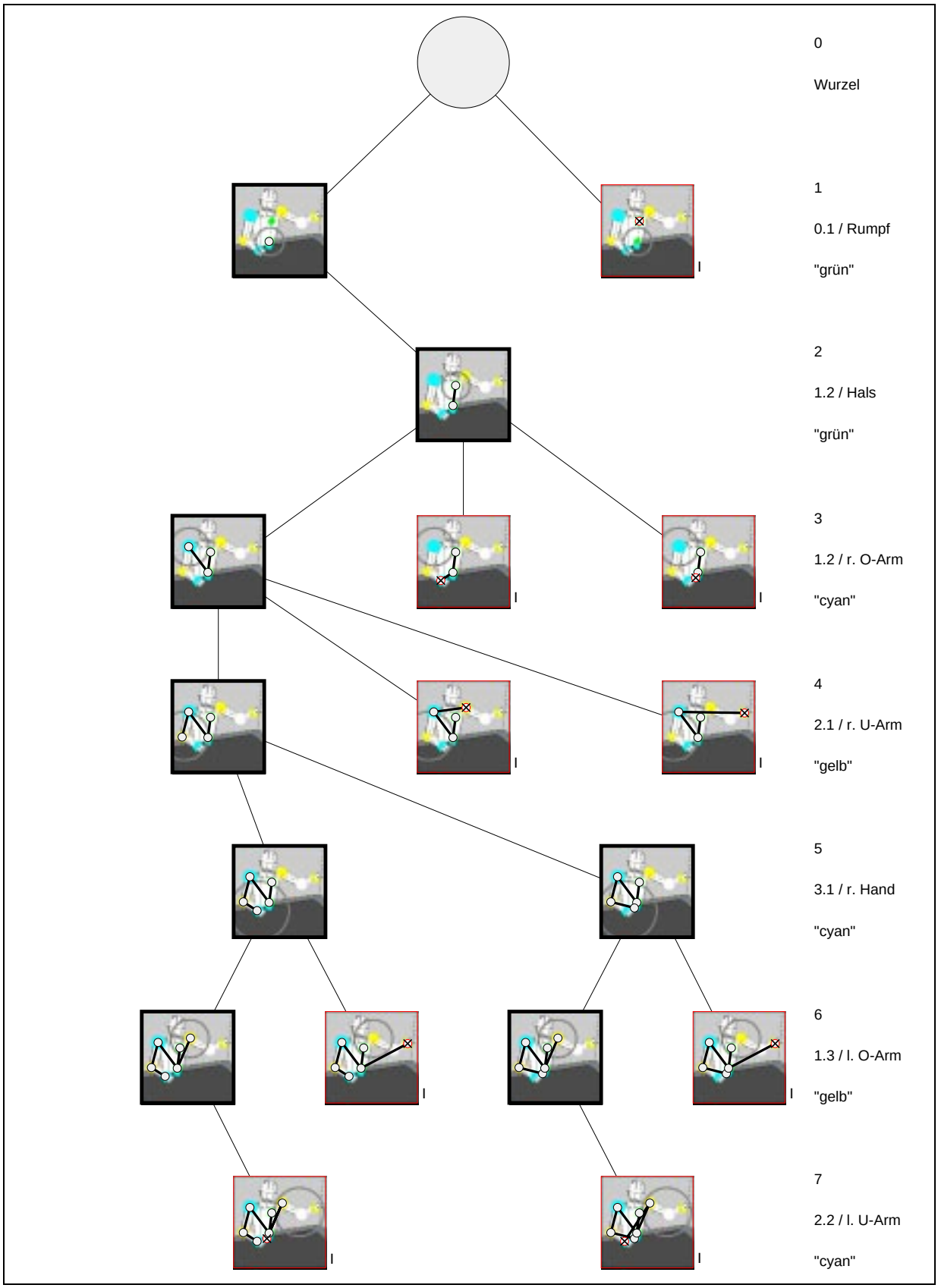


Abbildung 3.20: *Falsches Szenenmerkmal: Abbruch der Interpretation in Ebene 7 des Interpretationsbaumes.*

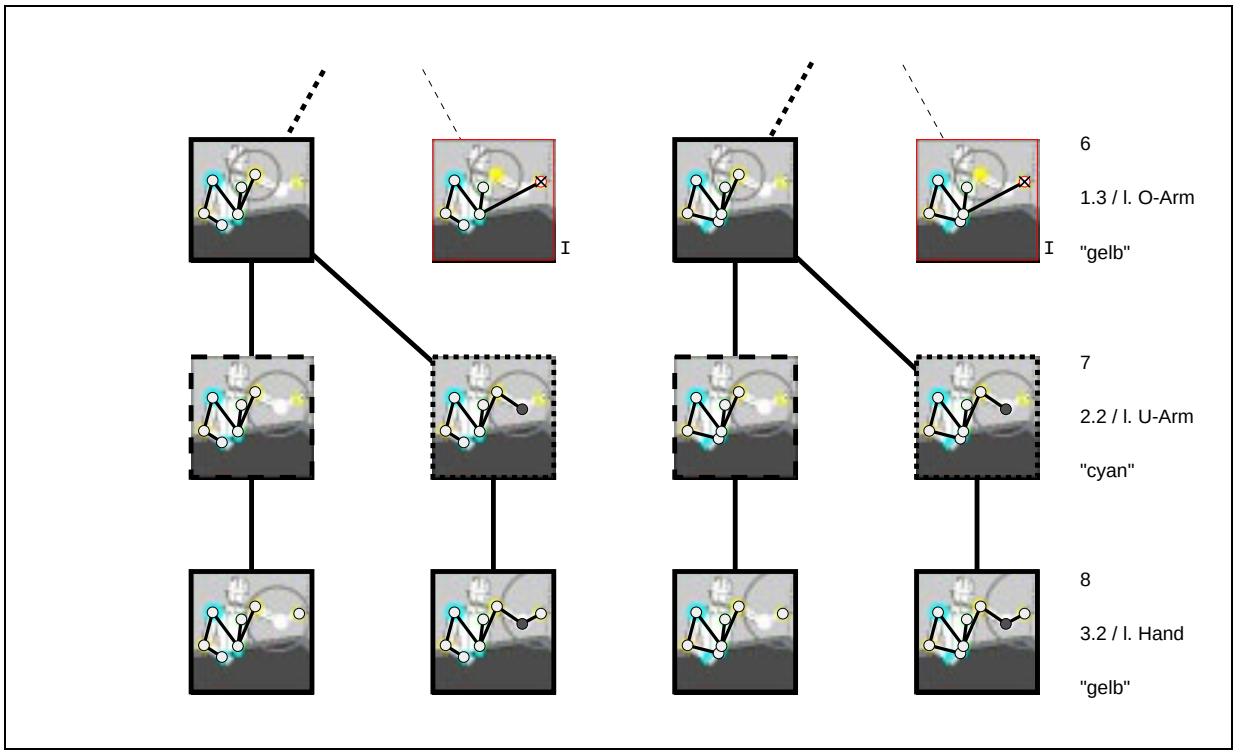


Abbildung 3.21: *Falsches Szenenmerkmal: Wiederaufsetzen der Interpretation mit 'leerer' / 'geschätzter' Ebene 7 im Interpretationsbaum.*

'Geschätzter' Baum

Kann bei der Re-Detektion einer Objektmodellinstanz auch unter Verwendung von 'leeren' / 'geschätzten' Ebenen keine Hypothese gefunden werden, so müßte man die komplette Objektmodellinstanz verwerfen. Geht man jedoch davon aus, daß nur aufgrund von zeitweilig verdeckten und / oder falsch detektierten Merkmalen keine Hypothese ermittelt werden konnte, so kann man für alle Objektmodellteile Schätzungen verwenden. Somit erhält man einen Interpretationsbaum, in dem nur ein Ast mit jeweils einer 'geschätzten' Assoziation für jedes primäre Merkmal enthalten ist. Man spricht daher von einem 'geschätzten' Baum.

Aufgrund der Restriktion $restr^{(originQ)}(.)$ wird eine Objektmodellinstanz jedoch nicht in jedem Fall als komplette Schätzung immer wieder verwendet. Entsprechend der Ausführungen zur Positionsvorhersage in Kap. 3.2.3, wird das Qualitätsmaß eines vorhergesagten Punktes bei wiederholter Vorhersage schlechter. Verwendet man diesen vorhergesagten Punkt als Schätzung für ein Szenenmerkmal, so wird dieses nur so lange akzeptiert, wie die Restriktion $restr^{(originQ)}(.)$ erfüllt ist.

Bei der Bewertung der Hypothesen wird die Qualität der 3D Positionen der den primären Merkmalen zugeordneten Szenenmerkmale ebenfalls noch berücksichtigt. Es werden daher die Hypothesen bevorzugt, bei denen die meisten realen Szenenmerkmale verwendet wurden.

3.6.5 Größe des Interpretationsbaumes

Der Aufwand zur Suche von Korrespondenzen zwischen Szenenmerkmalen und Modellmerkmalen wird durch den Aufwand der Traversierung des Interpretationsbaumes bestimmt. Damit ist die Komplexität der Korrespondenzsuche von der Größe des aufzustellenden Interpretati-

onsbaumes abhängig. Dies gilt in STABIL^{++} insbesondere deshalb, da in dem Interpretationsbaum eine erschöpfende Suche nach Hypothesen durchgeführt wird. Dies steht im Gegensatz zu der Darstellung der "Kontrolle der Suchraumexplosion" in [Gri90a] und der "Suchstrategie" in [Lan98]. Dort wird vorgeschlagen, die Suche im Interpretationsbaum auf eine maximale Anzahl von Verzweigungen und / oder maximale Anzahl von Hypothesen zu beschränken. Hierzu muß sichergestellt werden, daß sich in den zuerst durchlaufenden Ästen die richtige Hypothese befindet. Daher werden nur "erfolgversprechende" Äste verfolgt, indem die primären Merkmale umsortiert werden, so daß schon auf den oberen Ebenen bei der Traversierung des Interpretationsbaumes frühzeitig Sackgassen erkannt werden und die Kosten für das Zurückgehen[21] auf höhere Ebenen begrenzt werden.

In STABIL^{++} ist die Reihenfolge der zuzuordnenden primären Merkmale durch die innere Objektmodellstruktur starr vorgegeben. Bei der erschöpfenden Suche im Interpretationsbaum werden nur die algorithmisch weniger aufwendigen Bewertungen in Form der vorgestellten Restriktionen angewendet, so daß hiermit zunächst alle möglichen Hypothesen aufgestellt werden. Im Schritt der Hypothesenbewertung wird dann erst die weitere Auswahl der Hypothesen vorgenommen. Trotzdem gilt es, die Anzahl der möglichen Hypothesen und aufzustellenden Assoziationen gering zu halten.

Zu den Beispielinterpretationen in den vorhergehenden Abschnitten wurden schon Angaben zur maximalen Anzahl von Assoziationen pro Ebene des Interpretationsbaumes und zur maximalen Gesamtanzahl der Assoziationen gemacht. Ebenso war dort zu erkennen, daß der Einsatz der Restriktionen die Anzahl der real aufzubauenden Assoziationen stark beschränkt. In diesem Abschnitt wird aufgezeigt, wie die maximale Größe des Interpretationsbaumes zunächst generell durch die Anzahl von Modell- und Szenenmerkmalen beeinflußt wird. Daran schließt sich an, wie die Suchraumgröße durch die Reihenfolge der primären Merkmale in den Ebenen des Baumes und durch den gezielten Einsatz der Restriktionen beschränkt werden kann.

Maximale Größe

Geht man davon aus, daß zum Aufbau des Interpretationsbaumes einer Liste von n primären / Modellmerkmalen $\mathbf{F} = \{\mathbf{f}_1, \ldots, \mathbf{f}_n\}$ eine Liste von m Szenenmerkmalen $\mathbf{S} = \{\mathbf{s}_1, \ldots, \mathbf{s}_m\}$ gegenüber steht und diese alle das gleiche Basisattribut haben, wird die maximale Anzahl von Hypothesen H durch:

$$|H| = \frac{m!}{(m-n)!}$$

bestimmt.

Die Anzahl der maximal aufzustellenden Hypothesen entspricht auch der maximalen Anzahl von Assoziationen, die auf der untersten Ebene des Interpretationsbaumes aufzustellen sind. In den Ebenen darüber können, bei der Verwendung nur eines Basisattributes, entsprechend weniger Assoziationen aufgestellt werden. Die Gesamtanzahl aller möglichen Assoziationen ergibt sich somit durch die Summe der maximalen Anzahl von Assoziationen der Ebenen $0 \ldots n$:

$$\sum_{i=1}^{n} \frac{m!}{(m-i)!}$$

Werden zur Interpretation anstelle von primären Merkmalen mit nur einem Basisattribut primäre Merkmale mit k verschiedenen Basisattribute $(1) \ldots (k)$ verwendet, so reduziert sich

[21] engl. *backtracking*.

bei gleicher maximaler Anzahl aller Szenenmerkmale die Größe des Interpretationsbaumes. Generell ergibt sich für die maximale Anzahl von aufzustellenden Hypothesen:

$$|H| = \frac{m_1!}{(m_1 - n_1)!} \cdot \frac{m_2!}{(m_2 - n_2)!} \cdots \frac{m_k!}{(m_k - n_k)!}$$

wobei

$$\begin{aligned} \exists\, assoc \left\langle \mathbf{s}_i^{(l)}, \mathbf{f}_j^{(l)} \right\rangle, \quad & i = 1, \ldots, n_a,\ j = 1, \ldots, m_b, \\ & a = 1, \ldots, k,\ b = 1, \ldots, k, \\ & l = 1, \ldots, k \end{aligned} \tag{3.21}$$

mit

$$\mathbf{S} = \{\mathbf{s}_1^{(1)} \ldots \mathbf{s}_{m_1}^{(1)}\} \cup \{\mathbf{s}_1^{(2)}, \ldots, \mathbf{s}_{m_2}^{(2)}\} \cup \ldots \cup \{\mathbf{s}_1^{(k)}, \ldots, \mathbf{s}_{m_k}^{(k)}\}$$

$$\mathbf{F} = \{\mathbf{f}_1^{(1)}, \ldots, \mathbf{f}_{n_1}^{(1)}\} \cup \{\mathbf{f}_1^{(2)}, \ldots, \mathbf{f}_{n_2}^{(2)}\} \cup \ldots \cup \{\mathbf{f}_1^{(k)}, \ldots, \mathbf{f}_{n_k}^{(k)}\}$$

Für die Beispielinterpretation aus Abschn. 3.6.3 wurden acht primäre Merkmale verwendet. Falls für diese alle das gleiche Basisattribut verwendet wird, so ist die maximale Anzahl von Hypothesen $8! = 40320$, wenn man davon ausgeht, daß ebenfalls acht Szenenmerkmale vorhanden sind und keine Restriktionen angewendet werden. Durch die Verwendung von zwei primären Merkmalen mit dem Basisattribut der Farbe “grün”, drei Merkmalen mit der Farbe “cyan” und ebenfalls drei Merkmale mit der Farbe “gelb” reduziert sich die Anzahl entsprechend der Glg. 3.21 auf $2! \cdot 3! \cdot 3! = 72$.[22]

Reihenfolge der primären Merkmale

Bei der Verwendung von mehreren verschiedenen Basisattributen variiert die maximale Anzahl der Assoziationen in den einzelnen Ebenen und somit auch die maximale Gesamtanzahl von Assoziationen im Interpretationsbaum. In den Abb. 3.22 (a) – (d) ist dies beispielhaft für die Zuordnung von fünf primären Merkmalen zu fünf Szenenmerkmalen gezeigt. Hierbei sind jeweils zwei verschiedene Basisattribute verwendet worden. Zwei primäre Merkmale zeichnen sich durch das erste Basisattribut aus, die weiteren drei primären Merkmale durch das zweite. In der Abb. ist die Zuordnung eines Szenenmerkmals zu einem primären Merkmal mit dem ersten Basisattribut mit einer “gestrichelten” Kante dargestellt. Die Zuordnungen von Merkmalen mit dem zweiten Attribut sind mit durchgängigen Linien gekennzeichnet.

Für die vier beispielhaften Interpretationsbäume in den Abb. 3.22 (a) – (b) ergeben sich maximale Anzahlen von Assoziationen zwischen 34 und 45. An den Beispielen sind zwei Regeln zu erkennen, um eine geringe Anzahl von Assoziationen und damit eine geringe maximale Größe des Interpretationsbaumes zu erhalten: Zunächst sollten in der ersten Ebene des Interpretationsbaumes immer Zuordnungen zwischen primären Merkmalen und Szenenmerkmalen mit den Basisattributen vorgenommen werden, von denen die geringste Anzahl zur Verfügung stehen, vgl. hierzu in Abb. 3.22 die oberste mit der unteren Reihe. Desweiteren soll frühzeitig, d.h. auf oberen Ebenen, die Zuordnung aller Merkmale eines Basisattributes abgeschlossen sein. Hierdurch erreicht man, daß durch die nicht steigende Anzahl von Assoziationen in der Folgeebene bei der Zuordnung der letzten Merkmale eines Basisattributes der Baum in den oberen Ebenen schmal gehalten wird, vgl. hierzu in Abb. 3.22 (a) mit (b) und (c) mit (d).

[22] Auch hier wird vorausgesetzt, daß die Anzahl der detektierten Szenenmerkmale eines Basisattributes mit der Anzahl der primären Merkmale mit diesem Basisattribut übereinstimmt.

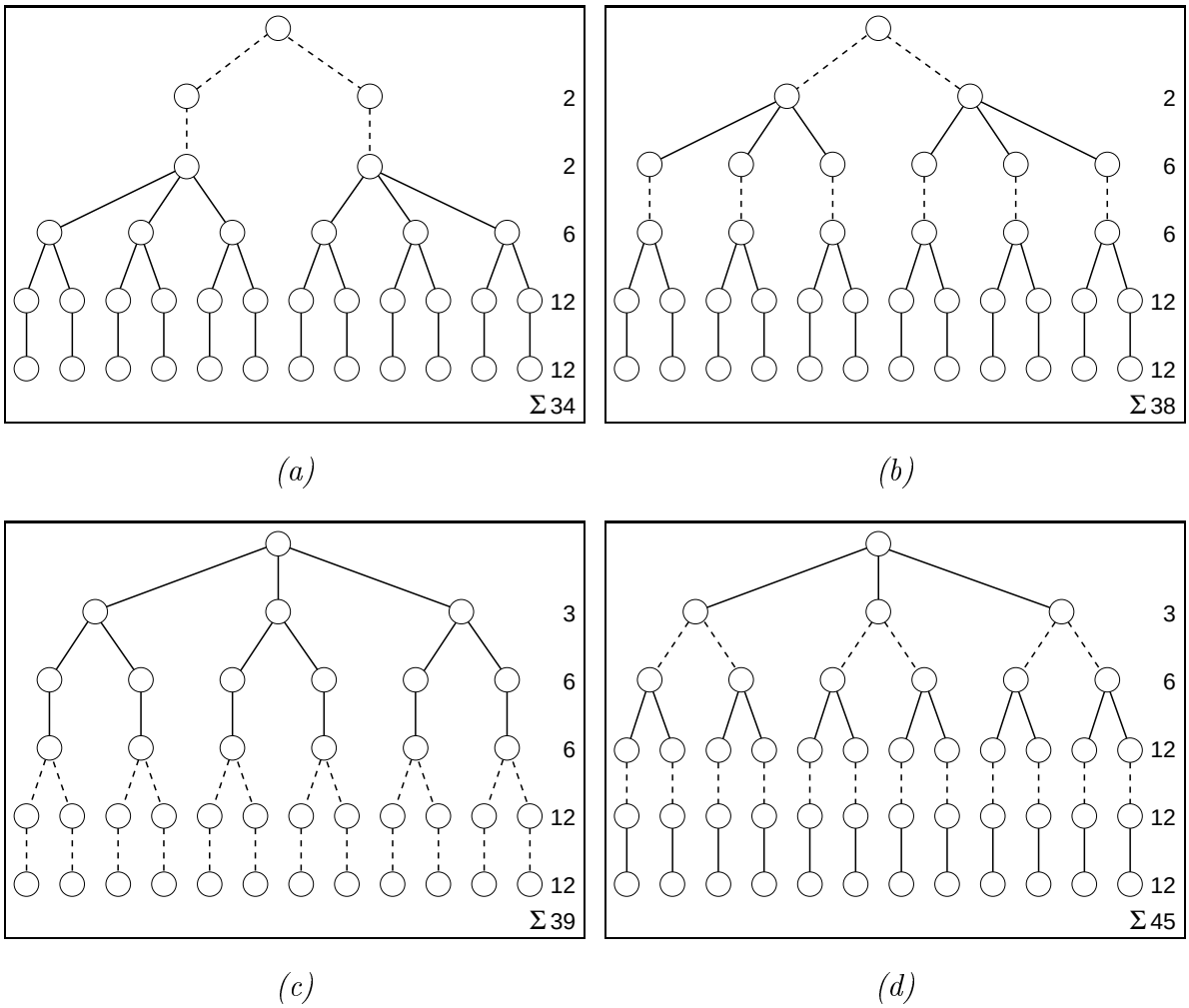


Abbildung 3.22: *Maximale Anzahl der Assoziationen in den Interpretationsbäumen bei verschiedener Reihenfolge der primären Merkmale.*

Durch die fehlende Möglichkeit der Umsortierung der primären Merkmale für die Zuordnungsreihenfolge im Interpretationsbaum kann die maximale Größe des Interpretationsbaumes nicht mehr bei der Suche beeinflußt werden. Es kann nur die innere Objektmodellstruktur entsprechend der o.a. Regeln angepaßt werden. Ist es anwendungsabhängig möglich, die Basisattribute der einzelnen primären Merkmale zu beeinflussen, so kann man auch hierüber die maximale Größe des Interpretationsbaumes steuern. Im Anh. D sind hierzu zwei weitere Beispielinterpretationen dargestellt, die sich zum einen durch eine Veränderung der hierarchischen, inneren Objektmodellstruktur und zum anderen durch eine Veränderung der Basisattribute auszeichnen. Zu diesen Beispielen sind dort auch die Maßzahlen der Interpretationsbäume aufgezeigt.

Anwendung der Restriktionen

Nachdem die maximale Größe des Interpretationsbaumes durch die Anzahl der Modell- / primären Merkmale und die Anzahl der detektierten Szenenmerkmale eines bestimmten Basisattributs abhängt, wird die reale Anzahl der aufzustellenden Assoziationen durch die anzuwendenden Restriktionen beschränkt. Die Signifikanz aller Restriktionen wird zunächst durch

die für die Restriktionen festgelegten Schwellenwerte bestimmt, vgl. Kap. 3.6.2. Kann man diese Grenzen anwendungsbedingt eng fassen, so kann die entsprechende Restriktion frühzeitig die Traversierung nicht erfolgversprechender Äste unterbinden.

Können keine signifikanten Unterschiede für die Merkmale bestimmt werden, so bringen die einstelligen Restriktionen auf der Merkmalsebene nicht den gewünschten Erfolg. In der Beispielinterpretation aus Abschn. 3.6.3 hat daher nur die Restriktion $restr^{(insideSsp)}(.)$ bei der Re-Detektion die Größe des Interpretationsbaumes stark reduziert. Dies liegt darin begründet, daß für jedes Objektmodellteil eine eigene 3D Position vorhergesagt wird. Liegen die einzelnen Positionen weit genug auseinander und werden durch eine entsprechend gute Qualität der Vorhersage entsprechend kleine 3D Suchräume bestimmt, so erhält man eine sehr mächtige Restriktion. Liegt jeweils, wie in der dargestellten Re-Detektion, nur noch ein 3D Szenenmerkmal in dem 3D Suchraum, so kann auf jeder Ebene des Interpretationsbaumes die Anzahl der gültigen Knoten auf eins reduziert werden.

Die mehrstelligen Restriktionen auf der Modellebene haben in der dargestellten Beispielinterpretation bei der initialen Detektion die Anzahl der gültigen Knoten beschränkt. Für diese gelten jedoch Einschränkungen entsprechend der hierarchischen, inneren Objektmodellstruktur. Jedes Objektmodellteil, außer dem ersten Objektmodellteil $omp_{0.1}$, hat in der hierarchischen, inneren Objektmodellstruktur ein Vorgängerobjektmodellteil. Dem entsprechend kann die Restriktion $restr^{(parentD)}(.)$ noch nicht auf der Ebene 1 des Interpretationsbaumes angewendet werden. Man sollte daher nach Möglichkeit die Anzahl der möglichen Knoten in der Ebene 1 des Baumes beschränken, wie im letzten Abschnitt beschrieben.

Für die Anwendung der Restriktion $restr^{(siblingD)}(.)$ muß jeweils auf den vorhergehenden Ebenen schon eine gültige Assoziation für ein Objektmodellteil aufgestellt worden sein, das das gleiche Vorgängerobjektmodellteil, wie das Objektmodellteil der aktuell betrachteten Ebene hat. Somit beeinflußt die hierarchische, innerer Objektmodellstruktur, in welcher Stufe diese tertiäre Restriktion angewendet werden kann. Nach Möglichkeit sollte die innere Modellstruktur dahingehend beeinflußt werden, daß schon in den oberen Ebenen des Interpretationsbaumes der Geschwister-Abstand überprüft werden kann.

In der vorgestellten Beispielinterpretation hat das erste Objektmodellteil $omp_{0.1}$ die Nachfolger $omp_{1.1}$, $omp_{1.2}$ und $omp_{1.3}$, vgl. Abb. 3.13. Nachdem das Objektmodellteil $omp_{1.1}$ keine Nachfolger hat, kann schon in der Ebene 3 des Interpretationsbaumes die Restriktion $restr^{(siblingD)}(.)$ für die Assoziationen der Objektmodellteile $omp_{1.2}$, $omp_{0.1}$ und $omp_{1.1}$ angewendet werden. Auf der Ebene 6 ist dies dann für die Objektmodellteile $omp_{1.3}$, $omp_{0.1}$ und $omp_{1.1}$ und $omp_{1.3}$, $omp_{0.1}$ und $omp_{1.2}$ möglich. In den weiteren Beispielen zum Interpretationsbaum im Anh. D ist in der zweiten Beispielinterpretation die hierarchische, innere Objektmodellstruktur dahingehend abgeändert, daß die Restriktion $restr^{(siblingD)}(.)$ erst in der Ebene 5 angewendet werden kann.

3.7 Hypothesenbewertung

3.7.1 Einführung

Im Teilschritt der Detektion des Interpretationsprozesses werden entsprechend der Traversierung des Interpretationsbaumes Hypothesen $H = h_1, \ldots, h_k$ für die Detektion einer Objektmodellinstanz obj aufgestellt. Durch die Verwendung von Restriktionen beim Aufbau des Interpretationsbaumes können die Hypothesen $h_i \in H$ als plausible Hypothesen bezeichnet werden. In einem weiteren Schritt muß nun aus den plausiblen Hypothesen eine Hypothese ausgewählt werden, die als Interpretation der Szene für die Detektion von obj akzeptiert wird, vgl. Alg. 3.6.

Für die Auswahl wird für jede einzelne Hypothese h_i ein Gütemaß / eine Qualität $q_i \in [0, 1]$ bestimmt. Es wird die Hypothese akzeptiert, die die beste Qualität aufweist. Erreicht keine der Hypothesen eine vorgegebene minimale Qualität q_{min}, so ist die Detektion von obj gescheitert, vgl. Alg. 3.1 und 3.3.

Bei der Bewertung der Hypothesen wird, ähnlich wie bei der Anwendung der Restriktionen $restr(.)$ bei der Traversierung des Interpretationsbaumes, eine Plausibilität anhand von Gütefunktionen $qual(.)$ überprüft. Jedoch werden neben dem einfachen Ausschlußkriterium jeweils noch ein Gütegrad $q \in [0, 1]$ bestimmt. Es wird hierzu zum einen bestimmt, mit welcher Güte die Restriktionen erfüllt waren, die beim Aufbau des Interpretationsbaumes angewendet wurden, so daß ein Teil der Gütefunktionen direkt auf den verwendeten Restriktionen aufbauen. Entsprechend der Einteilung der Restriktionen in Kap. 3.6.2 wird daher hier von *Gütefunktionen auf Merkmalsebene* und von *Gütefunktionen auf Modellebene* gesprochen. Zum anderen werden noch weitreichendere Gütefunktionen auf der Modellebene für die Beurteilung der Hypothesen herangezogen. Diese zusätzlichen Gütefunktionen haben ebenfalls auch Restriktionseigenschaften. Diese werden jedoch aufgrund ihrer Komplexität nicht schon während des Aufbaus des Interpretationsbaumes für jede aufgestellte Assoziation angewendet, sondern erst auf die plausiblen Hypothesen. Dies gilt auch für die Anwendung der sekundären Merkmale $\mathbf{f}$, die den Objektmodellteilen zugeordnet sind und die ebenfalls zur Bewertung der Hypothesen verwendet werden.

Wird mit den weiteren Gütefunktionen oder den sekundären Merkmalen eine Qualität $q = 0$ ermittelt, so ist dies mit einer nicht erfüllten Restriktion gleichzusetzen. In diesem Fall kann die Hypothese nicht akzeptiert werden und es wird dieser eine Gesamtqualität von $q = 0$ zugewiesen, so daß diese verworfen wird. Ansonsten bestimmt sich die Gesamtqualität einer Hypothese aus Einzelqualitäten q_i, $i = 1 \ldots n$. n ergibt sich aus der Anzahl der verschiedenen Gütefunktionen, die auf die Assoziationen einer Hypothese angewendet werden. Für die Güte einer Hypothese h_k gilt dann:

$$q_k = \frac{1}{n} \sum_{i=1}^{n} a_i \cdot q_i, \quad \sum_{i=1}^{n} a_i = 1 \tag{3.22}$$

Hierbei werden die Einzelqualitäten entsprechend der Faktoren a_i gewichtet. Die Gewichtungsfaktoren sind anwendungsbedingt entsprechend der Aussagekraft der Gütefunktionen zu wählen.

Eine Einzelqualität q_i ergibt sich wiederum aus dem Mittelwert aller Anwendungen einer Gütefunktion $qual_i(.)$ eines Typs auf die Assoziationen der Hypothese h_k. Läßt sich die Gütefunktion $qual_i(.)$ auf alle m Assoziationen $\{assoc_1, \ldots, assoc_m\}$ der Hypothese anwenden, so

$qual^{(originQ)}(.)$	**origin q**uality	Merkmalsebene
$qual^{(fitToPr)}(.)$	**fit to pr**ediction	Merkmalsebene
$qual^{(parentD)}(.)$	**parent d**istance	Modellebene
$qual^{(siblingD)}(.)$	**sibling d**istance	Modellebene
$qual^{(jointAng)}(.)$	**joint ang**les	Modellebene
$qual^{(inters)}(.)$	**inters**ection	Modellebene
$qual_{\mathbf{f}'}^{(distToXY)}(.)$	**dist**ance **to xy**-plane	sekundäres Merkmal
$qual_{\mathbf{f}'}^{(insFgReg)}(.)$	**ins**ide **f**ore**g**round **reg**ion	sekundäres Merkmal

Tabelle 3.9: *Gütefunktionen zur Bewertung der Hypothesen.*

gilt für die Einzelqualität q_i:

$$q_i = \frac{1}{m} \sum_{j=1}^{m} qual_i(assoc_j) \tag{3.23}$$

Wird eine Gütefunktion auf mehrere Assoziationen angewendet oder ist diese nicht auf alle Assoziationen einer Hypothese anwendbar, so ergeben sich entsprechend weniger Einzelqualitäten.

3.7.2 Bewertungskriterien

In diesem Abschnitt werden die einzelnen Gütefunktionen und somit die Bewertungskriterien für die Hypothesen erläutert. Hierbei werden zunächst die Gütefunktionen auf Merkmalsebene beschrieben. Daran schließt sich die Beschreibung der Gütefunktionen auf Modellebene. Schließlich wird noch die Anwendung der sekundären Merkmale zur Bewertung der Hypothesen erläutert. In der Tab. 3.9 ist eine Übersicht über die verschiedenen Gütefunktionen $qual^{type}(.)$ gegeben, wobei eine Erläuterung der sich aus den zugehörigen englischen Begriffen ergebenden Bezeichnungen $type$ aufgeführt ist.

Gütefunktionen auf Merkmalsebene

Die Gütefunktionen der Merkmalsebenen sind unäre Funktionen, denn diese werden jeweils auf einzelne Assoziationen einer Hypothese angewendet. Entsprechend der Einteilung der Restriktionen auf der Merkmalsebene werden diese Gütefunktionen in die generellen und die Funktionen, die explizit bei einem primären Merkmal mit $QUAL$ angegeben sind, eingeteilt.

Die generellen Gütefunktionen stützen sich auf die 3D Positionen $\vec{p}_{wcs}$ der Szenenmerkmale, die den Modellmerkmalen mit den Assoziationen der Hypothese zugeordnet sind. Hierzu wird die Güte der 3D Position berücksichtigt und wird beurteilt, wie nahe die 3D Position des Szenenmerkmals an der vorhergesagten Position liegt.

Ist für ein Szenenmerkmal keine 3D Position vorhanden, d.h. ist ein nil-Szenenmerkmal in einem leeren Knoten des Interpretationsbaumes verwendet worden, so kann keine direkte Aussage über die Güte vorgenommen werden. Jedoch muß mit einer Gütefunktion $qual(.)$ eine Qualität $q \neq 0$ betimmt werden, da ansonsten die Hypothese verworfen wird. Andererseits ist zu berücksichtigen, daß q kleiner sein muß als bei einer vergleichbaren Assoziation mit realem Szenenmerkmal oder geschätztem Szenenmerkmal. Hiermit wird sichergestellt, daß

die Hypothesen bevorzugt ausgewählt werden, in deren Assoziationen keine oder weniger nil-Szenenmerkmale zugeordnet sind.

Die explizit angegebenen Gütefunktionen sind, wie die vergleichbaren Restriktionen, von den Attributen $attr^{\mathbf{m}}$ der Modellmerkmale abhängig. Diese Gütefunktion sollte nur verwendet werden, wenn sich aus dem Vergleich der projizierten Modellmerkmale mit den Bildmerkmalen gesicherte Meßgrößen ableiten lassen. Zum Aufstellen explizit anzugebender Gütefunktionen sei auf die beschriebenen Restriktionen $restr^{(areaFnd)}(.)$ und $restr^{(exzentr)}(.)$, sowie auf die folgende Beschreibung der beiden generellen Gütefunktionen der Merkmalsebene verwiesen, so daß hier auf eine Beschreibung der Gütefunktionen $qual^{(areaFnd)}(.)$ und $qual^{(exzentr)}(.)$ verzichtet worden ist.

Güte des 3D Punktes: Wie bei der Restriktion $restr^{(originQ)}(.)$ wird bei der Gütefunktion $qual^{(originQ)}(.)$ die Güte q_μ der 3D Position berücksichtigt, die dem Szenenmerkmal $\mathbf{s}\mu$ zugeordnet ist und beim 2D / 3D Übergang gesetzt wurde. Es gilt somit für $qual^{(originQ)}(assoc_\nu)$:[1]

$$qual^{(originQ)}(assoc_\nu) = q_\mu$$

wobei mit $assoc_\nu$ dem primären Merkmal $\mathbf{f}_\nu$ des Objektmodellteils omp_ν das Szenenmerkmal $\mathbf{s}_\mu$ zugeordnet ist.

Für die geschätzten Szenenmerkmale nimmt die Güte bei wiederholter Vorhersage ab. Jedoch wird ein geschätztes Szenenmerkmal nur dann bei der Interpretation verwendet, wenn die Güte größer als eine minimale Qualität q_{min} ist, vgl. Kap. 3.2.3.

Wie oben angegeben, muß auch bei der Zuordnung von nil-Szenenmerkmalen zu Modellmerkmalen eine Güte bestimmt werden, wobei die Güte jedoch unterhalb der Güte von Assoziationen mit realen oder geschätzten Szenenmerkmalen liegen muß. Da die Güte für die 3D Positionen der geschätzten Szenenmerkmale minimal q_{min} betragen kann, wird für die Assoziation $assoc_\xi$ die Güte entsprechend

$$qual^{(originQ)}(assoc_\xi) = 0.1 \cdot q_{min}$$

gesetzt, wobei mit $assoc_\xi$ dem primären Merkmal $\mathbf{f}_\xi$ des Objektmodellteils omp_ξ ein nil-Szenenmerkmal $\mathbf{s}_{nil}$ zugeordnet ist.

Nähe zur Vorhersageposition: Mit der Gütefunktion $qual^{(fitToPr)}(.)$ soll die Nähe der 3D Position des Szenenmerkmals zur Vorhersageposition beurteilt werden. Dadurch wird beurteilt, wie wahrscheinlich die gewählte Zuordnung zur Bewegungsvorhersage des Objektmodells paßt. Liegen zwei Szenenmerkmale mit gleichem Basisattribut innerhalb des Suchbereiches eines Objektmodellteiles, so soll der Hypothese bei der Auswahl der Vorrang eingeräumt werden, bei der die 3D Position des zugeordneten Szenenmerkmals näher an der Vorhersageposition liegt. Daher kann die Gütefunktion $qual^{(fitToPr)}(.)$ mit der Restriktion $restr^{(insideSsp)}(.)$ verglichen werden.

Aufgrund der Anwendung der Restriktion $restr^{(insideSsp)}(.)$ beim Aufbau des Interpretationsbaumes liegen alle 3D Positionen der zu einem primären Merkmal $\mathbf{f}_\mu$ zugeordneten realen und geschätzten Szenenmerkmale innerhalb des Suchraumes des Objektmodellteiles omp_μ. Die schlechteste Güte $q = q_{min}$ soll bei der Anwendung von $qual^{(fitToPr)}(.)$ bestimmt werden, wenn die 3D Position des zugeordneten Szenenmerkmals soeben noch innerhalb des Suchraumes liegt. Dies entspricht der maximal möglichen Entfernung d_{max} von der Vorhersageposition

[1] Es ergeben sich unterschiedliche Güten für 3D Positionen, die mit dem Mono- und Stereoansatz bestimmt werden, vgl. Kap. 3.5.1. Für geschätzte Szenenmerkmale wird q aus Güte der verwendeten Vorhersageposition bestimmt.

$\tilde{\vec{p}}$. Eine Qualität von $q = 1$ soll erreicht werden, wenn die 3D Position $\vec{p}_{wsc}$ des zugeordneten Szenenmerkmals mit der Vorhersageposition zusammenfällt. Aufgrund dieser Überlegungen und der Forderung, daß die Qualität q mit größer werdendem Abstand $d = \|\vec{p}_{wsc} - \tilde{\vec{p}}\|$ exponentiell abnehmen soll, gilt für die Anwendung auf eine Assoziation $assoc_\nu$:

$$qual^{(fitToPredict)}(assoc_\nu) = \begin{cases} 1 & , \quad d = 0 \\ 1 - e^{\left(\frac{d_{max} \cdot \ln(1-q_{min})}{d}\right)} & , \quad \text{sonst} \end{cases}$$

Hierbei ist $\tilde{\vec{p}}$ der Vorhersagepunkt für das Objektmodellteil omp_ν.

Für Bestimmung der maximal möglichen Entfernung d_{max} muß zwischen den Vorhersagen $\tilde{\vec{p}}_0$ und $\tilde{\vec{p}}_e$ unterschieden werden.[2] Liegt für das Objektmodellteil eine Positionsvorhersage $\tilde{\vec{p}}_0$ vor, so ist der 3D Suchraum aus nur einer Kugel mit dem Durchmesser d_{p_0} bestimmt. Somit ergibt sich ein maximaler Abstand $d_{max} = \frac{1}{2} \cdot d_{p_0}$. Bei einer Vorhersageposition $\tilde{\vec{p}}_e$, die durch Extrapolation ermittelt wurde, kann die 3D Position des Szenenmerkmals innerhalb des keulenförmigen Suchbereiches liegen, vgl. Abb. 3.2. Der maximal mögliche Abstand von $\tilde{\vec{p}}_e$ wird für einen Punkt erreicht, der auf einer Gerade liegt, die durch $\tilde{\vec{p}}_e$ und $\tilde{\vec{p}}_0$ gebildet wird und der auf dem Rand der Suchraumkugel, die um $\tilde{\vec{p}}_0$ gebildet ist, liegt. Die Suchraumkugel um $\tilde{\vec{p}}_0$ hat einen Durchmesser von $0.75 \cdot d_p$, vgl. Glg. 3.6. Daher ergibt sich für $d_{max} = \|\tilde{\vec{p}}_e - \tilde{\vec{p}}_0\| + \frac{1}{2} \cdot 0,75 \cdot d_p$.

Da für nil-Szenenmerkmale keine 3D Position bestimmt ist, wird festgelegt, daß bei der Anwendung von $qual^{(fitToPr)}(.)$ auf eine Assoziation, bei der ein nil-Szenenmerkmal zugeordnet worden ist, die Qualität $q = 0.1 \cdot q_{min}$ sein soll. Desweiteren sei hier noch erwähnt, daß $qual^{(fitToPr)}(.)$ bei der Anwendung auf eine Assoziation, bei der ein geschätztes Szenenmerkmal zugewiesen wurde immer $q = 1$ liefert. Es ist daher zwingend sicherzustellen, daß die Gewichtungsfaktoren a_i in Glg. 3.22 so gewählt werden, daß mit der Gütefunktion $qual^{(fitToPr)}(.)$ nicht generell den Hypothesen mit geschätzten Szenenmerkmalen der Vorrang gegeben wird.

Gütefunktion auf Modellebene

Die Gütefunktionen auf der Modellebene basieren zunächst, wie die Restriktionen auf der Modellebene, auf der hierarchischen, inneren Objektmodellstruktur und den Translationsanteilen der geometrischen Struktur. In weiteren Funktionen werden noch die Rotationsanteile der geometrischen Struktur berücksichtigt, indem die Gelenkwinkel in den Knoten zwischen den einzelnen Objektmodellteilen auf Zulässigkeit überprüft werden. Schließlich wird noch mit der Überprüfung der gegenseitigen Durchdringung von Objektmodellteilen die äußere Struktur des Objektmodells in die Beurteilung der Hypothesen einbezogen.

Vater-Abstand: Die Gütefunktion $qual^{(parentD)}(.)$ basiert auf dem Vergleich der 3D Distanz $d_{\mathbf{m}}$ zwischen den Modellmerkmalen eines Objektmodellteiles und seinem Vorgängerobjektmodellteil und der 3D Distanz $d_{\mathbf{s}}$ der zugeordneten Szenenmerkmale, vgl. Glg. 3.16 für die vergleichbare Restriktion $restr^{(parentD)}(.)$. Durch die Anwendung der Restriktion ist beim Aufbau des Interpretationsbaumes die größten noch zugelassenen Abweichungen zwischen Modellpunkten und Szenenpunkten auf Δ_{max} begrenzt worden. Daher muß $qual^{(parentD)}(.)$ einen minimalen Wert von q_{min} bei der maximalen Abweichung Δ_{max} erreichen.

Stimmen die Abstände zwischen den Modellpunkten mit den Abständen der Szenenpunkte exakt überein, so soll die maximale Qualität $q = 1$ gesetzt werden. Ansonsten soll auch hier die

[2] Vgl. Kap. 3.2.4 zur Bestimmung der Suchräume der primären Merkmale.

Qualität exponentiell mit größer werdender Abweichung $d = |d_{\mathbf{s}} - d_{\mathbf{m}}|$ abnehmen, daher gilt für $qual^{(parentD)}(.)$ bei der Anwendung auf die Assoziationen $assoc_\mu$ und $assoc_\nu$:

$$qual^{(parentD)}(assoc_\mu, assoc_\nu) = \begin{cases} 1 & , \quad d = 0 \\ 1 - e^{\left(\frac{\Delta_{max} \cdot \ln(1-q_{min})}{d}\right)} & , \quad \text{sonst} \end{cases} \tag{3.24}$$

wobei das Objektmodellteil omp_ν in der hierarchischen, inneren Objektmodellstruktur das Vorgängerobjektmodellteil von omp_μ ist. Ist bei der Assoziation $assoc_\mu$ und / oder $assoc_\nu$ ein nil-Szenenmerkmal zugeordnet worden, so soll mit $qual^{(parentD)}(.)$ eine Qualität $q = 0.1 \cdot q_{min}$ bestimmt werden.

$qual^{(parentD)}(.)$ ist eine binäre Gütefunktion, da hier zwei Assoziationen miteinander verglichen werden. Die Gütefunktion läßt sich auf alle Assoziationen außer der Assoziation für das erste Objektmodellteil $omp_{0.1}$ anwenden. Dies ist bei der Mittelwertbildung entsprechend der Glg. 3.23 zu berücksichtigen.

Geschwister-Abstand: Die Gütefunktion $qual^{(siblingD)}(.)$, die den Geschwister-Abstand überprüft, ist von der Restriktion $restr^{(siblingD)}(.)$ abgeleitet und mit der Gütefunktion $qual^{(parentD)}(.)$ zu vergleichen. Auch hier wird ein 3D Abstand $d = |d_{\mathbf{s}} - d_{\mathbf{m}}|$ gebildet, vgl. Glg. 3.17. Für die Bewertung dieses Abstandes gilt das gleiche, wie in Glg. 3.24.

$qual^{(siblingD)}(assoc_\mu, assoc_\nu, assoc_\xi)$ ist eine tertiäre Gütefunktion. Diese kann nur angewendet werden, wenn für die, den drei Assoziationen zugehörenden Objektmodellteile omp_μ, omp_ν und omp_ξ gilt, daß omp_μ in der hierarchischen, inneren Objektmodellstruktur mit omp_ν das gleiche Vorgängerobjektmodellteil, wie omp_ξ hat. Dies ist bei der Mittelwertbildung entsprechend der Glg. 3.23 zu berücksichtigen.

Ist bei der Assoziation des jeweiligen Vorgängerobjektmodellteiles ein nil-Szenenmerkmal zugeordnet worden, so läßt sich der 3D Abstand d bilden und die Qualität q mit $qual^{(siblingD)}(.)$ wie o.a. bestimmen. Falls jedoch bei einer der anderen beteiligten Assoziationen ein nil-Szenenmerkmal zugeordnet wurde, so wird $q = 0.1 \cdot q_{min}$ gesetzt.

Gelenkwinkel: Für jedes Objektmodellteil omp_ν, außer dem ersten Objektmodellteil $omp_{0.1}$, wird mit der homogenen Transformationsmatrix ${}^{omp_\mu} = \mathbf{T}_{omp_\nu}$ die Transformation seines lokalen Koordinatensystems zum lokalen Koordinatensystem seines Vorgängerobjektmodellteiles omp_μ angegeben. Mit dem Rotationsanteil der Transformationsmatrix sind die Gelenkwinkel in den Knoten zwischen den beiden Objektmodellteilen bestimmt.

Dem Objektmodellteil omp_ν sind mit $restr_\nu^{\angle}$ Grenzwinkel als Winkelrestriktionen bekannt, die sich aus der Struktur des modellierten Objektes ergeben. Mit einer Gütefunktion $qual^{(jointAng)}(.)$ sind diese Winkelrestriktionen zunächst zu überprüfen und die Einhaltung der Restriktionen zu bewerten.

Mit der Generierung der Hypothesen durch die Traversierung des Interpretationsbaumes ist von der geometrischen Struktur jedoch nur der Translationsanteil bestimmt worden. Daher muß vor der Anwendung von $qual^{(jointAng)}(.)$ der Rotationsanteil noch bestimmt werden. Nachdem in jedem Knoten / Gelenk zunächst generell 3 Freiheitsgrade der Rotation vorhanden sind, jedoch jedes Gelenk nur durch einen 3D Punkt fixiert ist, müssen für die Bestimmung der Rotationen Heuristiken angenommen werden, die sich aus dem modellierten Objekt ergeben.

Es werden daher einzelne Objektmodellteile zu Kompositionen zusammengefaßt, wobei sich jeweils die Rotationen für die zusammengefaßten Objektmodellteile aus den 3D Positionen der Szenenmerkmale bestimmen lassen, die durch die Assoziationen den primären Merkmalen der Objektmodellteile zugeordnet sind. Jedem Objektmodell sind die Kompositionen mit

$COMPO$ bekannt. Im Anh. A ist die Bestimmung der Rotationen zwischen den Objektmodellteilen unter Anwendung der Kompositionen erläutert. Die Beschreibung der Kompositionen stützt sich dort auf die Modellierung des menschlichen Körpers ab.

Wird in einer Assoziation $assoc_\mu$ ein nil-Szenenmerkmal zugeordnet, so können für die Komposition, der das Objektmodellteil omp_μ angehört, keine Rotationen bestimmt werden. Für die Rotationen in den Objektmodellteilen der betreffenden Komposition werden dann die Rotationen des vorhergehenden Interpretationszyklus übernommen. Hierzu wird der Rotationsanteil der Transformationsmatrix verwendet, die für den Zeipunkt $t_{(-1)}$ in der Historie $HIST$ des entsprechenden Objektmodellteiles vermerkt ist. Das gleiche gilt, wenn ein Objektmodellteil in keiner Komposition des Objektmodells aufgenommen ist. Damit werden für den Rotationsanteil der Transformationsmatrix dieses Objektmodellteils immer die Werte des initialen Objektmodells beibehalten.

Innerhalb der Kompositionen eines Objektmodells gibt es eine Reihenfolge, die durch die hierarchische, innere Objektmodellstruktur bestimmt ist. Damit können die Rotationen nur in dieser Reihenfolge bestimmt werden. Dies begründet sich darin, daß die 3D Positionen $\vec{p}_{wcs}$ der Szenenmerkmale im Bezug zum Weltkoordinatensystem wcs des Szenenmodells definiert sind, jedoch die Bestimmung der Rotationen in den lokalen Koordinatensystemen der Objektmodellteile durchgeführt werden muß. Hierzu müssen die 3D Positionen $\vec{p}_{wcs}$ entsprechend der hierarchischen, inneren Objektmodellstruktur jeweils von dem lokalen Koordinatensystem eines Objektmodellteils zum nächsten transformiert werden. Ist die Kette der kaskadierten Transformation unterbrochen, da für ein Objektmodellteil, aufgrund der Zuordnung eines nil-Szenenmerkmals in der zugehörigen Assoziation, der Rotationsanteil nicht bestimmt werden konnte, so können für alle nachfolgenden Objektmodellteile keine Rotationen bestimmt werden, vgl. Glg. 2.4. Auch hier wird auf die Historie $HIST$ zurückgegriffen.

Sind die Rotationsanteile der Transformationsmatrizen der Objektmodellteile bestimmt worden, so ist damit für eine Hypothese die komplette geometrische Struktur ermittelt. Die Gütefunktion $qual^{(jointAng)}(.)$ kann nun für alle die Assoziationen angewendet werden, bei denen für das zugehörige Objektmodellteil im aktuellen Interpretationszyklus die Rotationen bestimmt worden sind.

Für die Anwendung von $qual^{(jointAng)}(.)$ auf die Assoziation $assoc_\nu$ werden aus der Transformationsmatrix ${}^{omp_\mu}\mathbf{T}_{omp_\nu}$ des Objektmodellteils omp_ν die drei Rotationswinkel α_ν, β_ν und γ_ν jeweils entsprechend des Rotations-Systems bestimmt, in dem die Grenzwinkel in den Winkelrestriktionen $restr_\nu^{\angle}$ angegeben sind.[3] Für jeden der drei Winkel muß bestimmt werden, ob dieser zwischen dem minimalen und dem maximalen Grenzwinkel liegt und eine entsprechende Qualität q bestimmt werden. Der Funktionswert von $qual^{(jointAng)}(.)$ ergibt sich somit aus:

$$qual^{(jointAng)}(assoc_\nu) = \begin{cases} 0 & , \quad \bigvee_{i=1}^{3} f_a(\varphi_i, \varphi_{i_{min}}, \varphi_{i_{max}}) = 0 \\ \frac{1}{3}\sum_{i=1}^{3} f_a(\varphi_i, \varphi_{i_{min}}, \varphi_{i_{max}}) & , \quad \text{sonst} \end{cases} \tag{3.25}$$

wobei $\varphi_i \in \{\alpha_\nu, \beta_\nu, \gamma_\nu\}$ und sich die minimalen und maximalen Grenzwinkel $\varphi_{i_{min}}$, $\varphi_{i_{max}}$ aus $restr_\nu^{\angle}$ ergeben.

Es ist jedoch nicht sinnvoll, bei Winkeln genau innerhalb der Grenzen eine Qualität $q = 1$ zu setzen und direkt ober- und unterhalb der Grenzen die Qualität $q = 0$ zu setzen. Vielmehr muß ein Unschärfebereich an den Grenzwerten betrachtet werden. Hierdurch wird erreicht, daß zum einen die Qualität von $q = 1$ nur innerhalb eines sehr wahrscheinlichen Kernbereiches

[3] Im Anh. A.7 sind die Grenzwinkel für die Modellierung des menschlichen Körpers angegeben. Vgl. auch Anh. B.5 zur Bestimmung der Rotations-Systeme in Transformationsmatrizen.

der Winkelgrenzen erreicht wird. Zum anderen wird eine Hypothese, bei der ein Gelenkwinkel in einem Objektmodellteil omp_ν etwas ober- oder unterhalb der Grenzwinkel liegt nicht verworfen. Jedoch muß bei der Anwendung von $qual^{(jointAng)}(.)$ auf die zugehörige Assoziation $assoc_\nu$ eine geringere Qualität bestimmt werden. Man erhält somit um die Grenzwinkel einen Unschärfebereich. Liegt der Gelenkwinkel jedoch außerhalb des Unschärfebereiches, so muß weiterhin eine Qualität $q = 0$ bestimmt werden.

Man erhält somit verschiedene Definitionsbereiche für die Überprüfung der Gelenkwinkel mit der Funktion $f_a(.)$, die in Glg. 3.25 für die Bestimmung von $qual^{(jointAng)}(.)$ verwendet wird. Es soll die Qualität q in den Unschärfebereichen exponentiell ansteigen und abfallen, wobei für den minimalen und maximalen Grenzwert $q = 0,5$ gelten soll. Es ergibt sich somit für $f_a(.)$:

$$f_a(\varphi, \varphi_{min}, \varphi_{max}) = \begin{cases} 0 & , & & & \varphi & < & (\varphi_{min} - \epsilon/2) \\ 1 - e^{-\left(\frac{a}{\varphi - \varphi_{min} - b}\right)} & , & (\varphi_{min} - \epsilon/2) & \geq & \varphi & < & \varphi_{min} \\ e^{\left(\frac{a}{\varphi - \varphi_{min} + b}\right)} & , & \varphi_{min} & \geq & \varphi & < & (\varphi_{min} + \epsilon/2) \\ 1 & , & (\varphi_{min} + \epsilon/2) & \geq & \varphi & \leq & (\varphi_{max} - \epsilon/2) \\ e^{-\left(\frac{a}{\varphi - \varphi_{max} - b}\right)} & , & (\varphi_{max} - \epsilon/2) & > & \varphi & \leq & \varphi_{max} \\ 1 - e^{\left(\frac{a}{\varphi - \varphi_{max} + b}\right)} & , & \varphi_{max} & > & \varphi & \leq & (\varphi_{max} + \epsilon/2) \\ 0 & , & & & \varphi & > & (\varphi_{max} + \epsilon/2) \end{cases}$$

wobei $0° < \varphi \leq 360°$, $0° < \varphi_{min} < \varphi_{max} \leq 360°$ und

$$\begin{aligned} a &= \epsilon/2 \cdot \ln(1 - c) \\ b &= \frac{a}{\ln(0,5)} \end{aligned}$$

Mit ϵ wird hierbei die Breite des Unschärfebereiches eingestellt, der von der Größe des zulässigen Winkelbereiches abhängen muß. Soll der Unschärfebereich z.B. 10% der Breite des Winkelbereiches haben, so gilt $\epsilon = 0,1 \cdot |\varphi_{max} - \varphi_{min}|$ bei $0° < \varphi_{min} < \varphi_{max} \leq 360°$. Mit c wird wiederum eingestellt, wie schnell im Unschärfebereich die Gütefunktion fällt oder steigt. Zudem wird der minimale Wert im Unschärfebereich mit c und der maximale Wert mit $1 - c$ festgelegt, daher muß $c \ll 0,5$ sein.[4]

Durchdringung von Objektmodellteilen: Mit der Überprüfung der Durchdringung von Objektmodellteilen wird die äußere Objektmodellstruktur berücksichtigt. Es wird hierzu mit einer Gütefunktion $qual^{(inters)}(.)$ für jede Assoziation der Hypothese bestimmt, ob das zugehörige Objektmodellteil sich mit anderen Objektmodellteilen des Objektmodells überschneiden. Es sollen hiermit die Hypothesen verworfen werden, bei denen sich die Volumenkörper, die den Objektmodellteilen zugeordnet sind, in nicht erlaubter Weise durchdringen und überschneiden. Dies basiert auf dem Grundsatz, daß jeweils zwei Teile des modellierten Objektes nicht in einander liegen und sich nicht kreuzen können.

Hierbei ist zu beachten, daß bei zwei in der hierarchischen, inneren Objektmodellstruktur aufeinander folgenden Objektmodellteilen sehr wohl eine (Teil-)Durchdringung beabsichtigt sein kann. Dies liegt darin begründet, daß eine Verformung der Volumenkörper in dem Objektmodell nicht berücksichtigt ist. So überlappen z.B. bei der Modellierung des menschlichen Körpers die Objektmodellteile des Oberarms und des Unterarms für einen Winkel $\alpha \neq 0°$ in der

[4] Aufgrund der Eigenschaften der Winkeldarstellung gelten die Angaben für positive Grenzwinkel. Bei negativen Grenzwinkeln und somit Grenzwinkelbereiche, die $0°$ oder $360°$ einschließen, sind die Funktionen den Grenzbereichen entsprechend anzupassen.

Nähe des Gelenkpunktes, vgl. z.B. Abb. 3.2. Liegt, aufgrund der gewählten Modellierung, der Ursprung des lokalen Koordinatensystems eines Objektmodellteiles sogar innerhalb des Volumenkörpers seines Vorgängerobjektmodellteiles, so ergeben sich immer Überlappungen beider Volumenkörper.

Daher werden die, jeweils in der inneren Struktur aufeinander folgenden, Objektmodellteile von der gegenseitigen Überprüfung der Durchdringung ausgenommen. Damit wird die Gesamtbeurteilung der Hypothese nicht beeinflußt, denn mit der Gütefunktion $qual^{(jointAng)}(.)$ ist anhand der Gelenkwinkel die Lage zweier, in der inneren Modellstruktur aufeinander folgender, Objektmodellteile hinreichend geprüft worden.

Bei der Bestimmung der Qualität q mit $qual^{(inters)}(assoz_\nu)$ soll $q = 1$ erreicht werden, wenn der Volumenkörper des Objektmodellteils omp_ν keinen Volumenkörper eines weiteren Objektmodellteils durchdringt. Liegt nur eine Teildurchdringung der äußeren Bereiche der Volumenkörper vor, so soll dies noch zulässig sein. Durchdringt jedoch der Kern des Volumenkörpers von omp_μ Volumenkörper eines anderen Objektmodellteils, so ist $q = 0$ zu setzen und die Hypothese zu verwerfen. Geht man davon aus, daß bei der Modellierung der Objekte, sowie bei der Bestimmung der geometrischen Struktur Ungenauigkeiten auftreten, so ist dies zu berücksichtigen. Es wird daher für $qual^{(inters)}(.)$ auf eine Differenzierung der Qualität bei Teildurchdringung verzichtet. Damit hat $qual^{(inters)}(.)$ rein restriktiven Charakter, so daß $q \in [0:1]$ gilt.

Aufgrund dieser Überlegungen kann man die Eigenschaften der Volumenkörper ausnutzen und bei der Überprüfung der Durchdringung nur betrachten, ob die Rotationsachse des entsprechenden Volumenkörpers einen anderen Volumenkörper schneidet. Es wird, wie gefordert, die Teildurchdringung von Objektmodellteilen zugelassen, falls die Durchdringung nicht den Kern des Volumenkörpers und somit die Rotationsachse umfaßt. Weiterhin wird hierdurch erreicht, daß die Durchdringung der Volumenkörper algorithmisch einfach überprüft werden kann. Für $qual^{(inters)}(.)$ gilt bei der Anwendung auf die Assoziation $assoc_\nu$ somit:

$$qual^{(inters)}(assoc_\nu) = \begin{cases} 0 & , \quad \bigvee_{i=1}^{n} f_d(vol_\nu, vol_i) = \text{wahr}, \quad i \neq \nu, \mu, \xi \\ 1 & , \quad \text{sonst} \end{cases}$$

Hierbei sind omp_μ und omp_ξ jeweils das Vorgänger- und das Nachfolgeobjektmodellteil von omp_ν in der hierarchischen, inneren Objektmodellstruktur. Mit der Funktion $f_d(vol_\nu, vol_\mu)$ ist zu bestimmen, ob die Rotationsachse des Volumenkörpers vol_ν den Volumenkörper vol_μ schneidet. Falls die Rotationsachse den Volumenkörper schneidet, so gilt $f_d(vol_\nu, vol_i) = \text{wahr}$.

Sekundäre Merkmale

Die sekundären Merkmale $\mathbf{f}'$, die einem Objektmodellteil explizit zugeordnet werden können, sind weitere Heuristiken zur Überprüfung und zur Verifikation von Hypothesen. Hierzu werden für eine Hypothese alle sekundären Merkmale $\mathbf{f}'_j \in \mathbf{F}'$ der zugehörigen Objektmodellinstanz obj nacheinander überprüft. In die Bestimmung der Gesamtgüte einer Hypothese gehen, entsprechend der Glg. 3.22, die Qualitäten der Gütefunktionen $qual_{\mathbf{f}'}(.)$ der sekundären Merkmale einzeln ein.

Nachdem einem Objektmodellteil, für das ein sekundäres Merkmal definiert ist, nicht zwangsläufig ein primäres Merkmal definiert sein muß, werden die Gütefunktionen $qual_{\mathbf{f}_\nu}(.)$ nicht auf die Assoziationen einer Hypothese, sondern auf die Objektmodellteile der Objektmodellinstanz angewendet. Hierzu wird jedoch vorausgesetzt, daß für alle Objektmodellteile von obj die Transformationsmatrizen ${}^{omp_\mu}\mathbf{T}_{omp_\nu}$ entsprechend der zu beurteilenden Hypothese gelten.

Im Kap. 2.4.6 sind beispielhaft zwei sekundäre Merkmale für das Modell des menschlichen Körpers vorgestellt worden, für die hier nun die Gütefunktionen erläutert werden. Dies ist zum einen die Überprüfung des Abstandes des Objektmodellteils der Füße zur xy-Ebene des Weltkoordinatensystems. Zum anderen ist dies die Überprüfung des Vorhandenseins einer entsprechenden Repräsentation des Oberkörpers unterhalb des Objektmodellteils des Kopfes. Die Beispiele lassen sich auf die Modellierung anderer Objekte übertragen. Ebenso ist es möglich, weitere Heuristiken, die sich aus den Eigenschaften eines Objektes ergeben, mit den sekundären Merkmalen zu überprüfen und zur Beurteilung der Hypothesen zu verwenden.

Position der Füße: Für die Überprüfung der Position des lokalen Koordinatensystems des Objektmodellteils omp_μ, z.B. der Füße, mit dem sekundären Merkmal $\mathbf{f}'_\mu$ sind kein weiteres Modellmerkmal $\mathbf{m}_\mu$ und weitere Bildverarbeitungsoperationen IP_μ notwendig. Es kann, entsprechend einer Transformationsmatrix ${}^{wcs}\mathbf{T}_{omp_\mu}$ die 3D Position $\vec{p}_{wcs}$ des lokalen Koordinatensystems von omp_μ im Weltkoordinatensystem bestimmt werden.[5] Für $\vec{p}_{wcs}$ kann der Abstand d zur xy-Ebene des Weltkoordinatensystems ermittelt werden.

Läßt man, aufgrund von zu berücksichtigenden Ungenauigkeiten zu, daß der Abstand maximal d_{max} betragen darf, so soll eine entsprechende Gütefunktion $qual_{\mathbf{f}'}^{(distToXY)}(.)$ bei dem maximal zulässigen Abstand eine Qualität $q = q_{min}$ bestimmen. Die maximale Qualität von $q = 1$ soll erreicht werden, wenn das Objektmodellteil der Füße die Ebene soeben berührt. Aufgrund dieser Überlegungen und der Forderung, daß der Wert der Funktion zwischen dem maximalen und minimalen Wert exponentiell fallen soll, ergibt sich für $qual_{\mathbf{f}'}^{(distToXY)}(.)$:

$$qual_{\mathbf{f}'}^{(distToXY)}(omp_\mu) = \begin{cases} 1 & , \quad d = 0 \\ 1 - e^{\left(\frac{d_{max} \cdot \ln(1-q_{min})}{d}\right)} & , \quad 0 < d \leq d_{max} \\ 0 & , \quad \text{sonst} \end{cases}$$

Gegenüber den Gütefunktionen, die sich aus den Restriktionen ableiten, die beim Aufbau des Interpretationsbaumes verwendet werden, muß hier noch die restriktive Eigenschaft berücksichtigt werden. Daher wird eine Qualität $q = 0$ bestimmt, wenn der Abstand $d > d_{max}$ ist. Wird für eine Anwendung von $qual_{\mathbf{f}'}^{(distToXY)(.)}$ eine Qualität $q = 0$ bestimmt, so wird die Hypothese verworfen.

Repräsentation des Oberkörpers: Ist bei der Anwendung von STABIL^{++} zur Personendetektion nur das Objektmodellteil für den Kopf mit dem primären Merkmal einer "hautfarbenen" Ellipse verwendet worden, so kann hiermit zunächst nur die hypothetische Lage eines Kopfes bestimmt werden. Über ein entsprechendes sekundäres Merkmal soll sichergestellt werden, ob unterhalb des vermeintlichen Kopfes noch eine Repräsentation des Oberkörpers vorhanden ist. Hierzu wird angenommen, daß die Projektion des Körpers eines Objektes generell in den segmentierten Vordergrundregionen $REG^{(fg)}$ liegt. Die Vordergrundregionen sind hierzu schon für jedes Bild, das zur Interpretation herangezogen wird, bestimmt worden, vgl. Kap. 3.4.2 zur Segmentierung.

Dem sekundären Merkmal $\mathbf{f}'_\mu$, das dem Objektmodellteil omp_μ des Rumpfes zugeordnet ist, ist ein Modellmerkmal $\mathbf{m}_\mu$ und mit $IP_\mu = ip(.)$ eine Bildverarbeitungsoperation bekannt, um den Wert der Gütefunktion $qual_{\mathbf{f}'}^{(insFgReg)}(.)$ bestimmen zu können. Das Modellmerkmal gibt

[5] ${}^{wcs}\mathbf{T}_{omp_\mu}$ ergibt sich aus der Kaskadierung der Transformationsmatrizen der einzelnen Objektmodellteile, die in der hierarchischen, inneren Objektmodellstruktur auf dem Pfad vom ersten Objektmodellteil $omp_{0.1}$ zum Objektmodellteil omp_μ liegen.

die Erscheinungsform an, die zu überprüfen ist. Hier ist es die Repräsentation entsprechend des Volumenkörpers vol des Objektmodellteiles. Daher muß als erster Schritt von dem Modellmerkmal $\mathbf{m}$ des Volumenkörpers zu einem Szenenmerkmal $\mathbf{s}$ übergegangen werden. Hierzu wird die Lage des Volumenkörpers aus dem lokalen Koordinatensystem des Objektmodellteils entsprechend den bei der Interpretation bestimmten Transformationsmatrizen in das Weltkoordinatensystem des Szenenmodells transformiert.

Der transformierte Volumenkörper wird anschließend in die zur Interpretation verwendeten Bilder $img_j \in IMG$, $j = 1 \ldots n$ projiziert. Als Projektionen der Volumenkörper vol_{ell} und vol_{trCone} werden als Näherung die Projektion von Kugeln und Zylindern verwendet. Vgl. hierzu die Projektion von kugelförmigen und zylindrischen Suchräumen in Kap. 3.4.2. Man erhält somit in jedem Bild img_j eine Region $reg_j^{(proj)}$ mit der Fläche $a_j^{(proj)}$.

Mit dem Bildverarbeitungsoperator $ip(.)$ werden für jedes Bild die Regionen bestimmt, in denen sich die Regionen $reg_j^{(proj)}$ und die Vordergrundregionen überschneiden. Für diese Schnittregionen wird ein Maß $a_j^{(inters)}$ für die Fläche bestimmt. Mit der Gütefunktion werden nun die Größen der Flächen verglichen. Hierzu wird ein mittlerer Quotient a über alle Regionen entsprechend

$$a = \frac{1}{n} \sum_{j=1}^{n} \frac{a_j^{(inters)}}{a_j^{(proj)}} \tag{3.26}$$

gebildet. Unterschreitet a einen Schwellenwert a_{min}, so soll die Gütefunktion eine Qualität $q = 0$ bestimmen. Für $a = a_{min}$ soll die Qualität $q = q_{min}$ betragen und für größere a exponentiell ansteigen, bis bei $a = 1$ eine Qualität $q = 1$ erreicht wird. Damit gilt für $qual_{\mathbf{f}'}^{(insFgReg)}(.)$ bei der Anwendung auf das Objektmodellteil omp_μ:

$$qual_{\mathbf{f}'}^{(insFgReg)}(omp_\mu) = \begin{cases} 0 & , \quad a < a_{min} \\ 1 - e^{\left(\frac{(a_{min}-1)\cdot \ln(1-q_{min})}{a-1}\right)} & , \quad a_{min} \leq a < 1 \\ 1 & , \quad \text{sonst} \end{cases}$$

Es ist bei der Anwendung von $qual_{\mathbf{f}'}^{(insFgReg)}(.)$ darauf zu achten, daß die projizierte Region des Volumenkörpers des Objektmodellteiles eine ausreichende Größe besitzt. Ist dies nicht der Fall, dann ist die Aussagekraft von $qual_{\mathbf{f}'}^{(insFgReg)}(.)$ unzuverlässig. Dies bedeutet, daß die Gütefunktion nicht angewendet werden soll, wenn das Objektmodellteil aufgrund seiner Lage und der Lage des Sichtbereiches der Kamera im Bild nicht hinreichend genug sichtbar ist. Um dies zu bestimmen, wird zunächst die Region bestimmt, die der projizierte Volumenkörper in der Bildebene der Kamera einnimmt. Anschließend wird diese Region auf die Region des effektiv nutzbaren Bildausschnittes reduziert. Zur Beurteilung wird das Verhältnis beider Regionen ermittelt.

Für das Beispiel bedeutet dies, daß bei einer zu detektierenden Person, der Oberkörper zunächst im Bild sichtbar ist, wenn die Person von der Kamera entfernt steht und der Kopf im oberen Bildteil abgebildet ist. Bewegt sich die Person auf die feststehende Kamera zu, dann wird ein zunehmend geringerer Anteil des Oberkörpers im Bild abgebildet, falls die Kamera z.B. an der Decke des Raumes montiert ist. Ab dem Unterschreiten eines Grenzwertes von z.B. 30% sichtbarer Fläche des Oberkörpers sollte bei der Anwendung von $qual_{\mathbf{f}'}^{(insFgReg)}(.)$ für das entsprechende Bild keine Güte über den Flächenquotient mehr ermittelt werden. Dies ist bei der Bestimmung des mittleren Quotienten in Glg. 3.26 zu berücksichtigen.

3.8 Hypothesenauswahl

3.8.1 Akzeptieren von Hypothesen

Nach der Bewertung der Hypothesen, die im Teilschritt der Detektion des Interpretationsprozeßes für eine Objektmodellinstanz obj aufgestellt wurden, wird die Hypothese mit dem besten Qualitätsmaß ausgewählt und ist als Interpretation der Szene für obj zu akzeptieren, vgl. Alg. 3.3. Für die Verwaltung der Objektmodellinstanzen ist hierbei zwischen der initialen Detektion und der Re-Detektion zu unterscheiden.

Initiale Detektion

Ist bei der Detektion des initialen Objektmodells eine Hypothese mit einer Qualität $q > q_{min}$ gefunden worden, so ist durch das System in der Szene ein weiteres Objekt detektiert worden. Aus der Hypothese wird daher eine neue Objektmodellinstanz erzeugt, die an die Liste der aktuell gefundenen Objektmodellinstanzen angehängt wird.[1]

Die neue Instanz ist damit eine Kopie des initialen Objektmodells, die zunächst die gleiche hierarchische, innere Objektmodellstruktur aufweist. Auch die Merkmale, die zur äußeren Objektmodellstruktur gehören, werden übernommen. Die einzelnen Instanzen unterscheiden sich jedoch durch die Lage im Szenenmodell und durch die Orientierung der einzelnen Objektmodellteile zueinander. Hierzu muß zum einen die Transformationsmatrix ${}^{wcs}\mathbf{T}_{obj}$ bestimmt werden und die Rotationsanteile der Transformationsmatrizen der einzelnen Objektmodellteile gesetzt werden. Hiermit wird die "Haltung" der Modellinstanz entsprechend des detektierten Objektes gesetzt.

Ist für das erste Objektmodellteil $omp_{0.1}$ ein primäres Merkmal bestimmt worden, so ist diesem in der ausgewählten Hypothese ein reales Szenenmerkmal zugeordnet worden.[2] Somit läßt sich aus der 3D Position $\vec{p}_{wcs}$ des Szenenmerkmals die Position der Objektmodellinstanz bestimmen. Hierzu wird ausgenutzt, daß in ${}^{wcs}\mathbf{T}_{obj}$ nur der Translationsanteil bestimmt werden muß und daß die Position des lokalen Koordinatensystems von $obj_{0.1}$ zum Koordinatensystem des Objektmodells fix ist. Vgl. zur Bestimmung von ${}^{wcs}\mathbf{T}_{obj}$ auch Glg. A.5.

Falls für $omp_{0.1}$ kein primäres Merkmal definiert wurde, so wird in der hierarchischen, inneren Modellstruktur das erste Objektmodellteil omp_{μ} gesucht, für das ein primäres Merkmal definiert wurde. Nimmt man alle Transformationen zwischen den Objektmodellteilen, die in der inneren Struktur auf dem Pfad von $omp_{0.1}$ zu omp_{μ} als fix an, so läßt sich ${}^{wcs}\mathbf{T}_{obj}$ äquivalent der Überlegungen zu Glg. A.5 bestimmen, nur daß hier mehrere Transformationsmatrizen zu kaskadieren sind.

Für die neue Objektmodellinstanz sind weiterhin noch die Translationen der geometrischen Struktur und die Ausdehnung der Volumenkörper der äußeren Struktur zu setzen. Hiermit ist die "Größe" der Modellinstanz entsprechend des detektierten Objektes zu setzen. Ist das zu detektierende Objekt in seinen Ausmaßen exakt bekannt und vermessen, so können schon für das initiale Objektmodell die Größe der einzelnen Objektmodellteile entsprechend gesetzt werden.[3] Für die neue Objektmodellinstanz werden dann die Werte aus dem initialen Objektmodell übernommen.

[1] Jedoch wurde die Interpretation schon mit einer "initialen" Objektmodellinstanz durchgeführt, so daß dort nicht zwischen initialer und Re-Detektion unterschieden werden muß. Die "initiale" Instanz muß jedoch für weitere initiale Detektionen wieder freigegeben werden.

[2] Bei der initialen Detektion werden keine geschätzten oder nil-Szenenmerkmale verwendet, vgl. Kap. 3.6.4.

[3] Die maximale 3D Zuordnungstoleranz kann für die Restriktionen dann entsprechend klein gewählt werden, vgl. Kap. 3.6.2..

Sind die Objekte, die in der Szene auftauchen können, in ihrer spezifischen Größe[4] nicht bekannt, so repräsentieren die Ausmaße der Objektmodellteile des initialen Objektmodells nur Standardgrößen. In dem Beispiel zur Personendetektion wird daher als Größe des Modells eine Durchschnittsgröße von z.B. 1,75 m angenommen. Nachdem das System eine Zuordnung von 3D Szenenmerkmalen zu 3D Modellmerkmalen vornimmt, werden sich, entsprechend der spezifischen / individuellen Objekte, Differenzen zwischen den Abständen der Modellpunkte und der Szenenpunkte ergeben. Daher ist die Objektmodellinstanz anzupassen, dieser Vorgang der Adaption, der auch für die Re-Detektion betrachtet werden muß, ist weiter unten beschrieben.

Weiterhin ist noch der Zeitstempel, der dem Interpretationszyklus durch die verwendeten Bilder zugewiesen ist, für die Modellinstanz als erster Eintrag in T zu übernehmen.

Re-Detektion

Das Akzeptieren einer Hypothese bei der Re-Detektion einer Objektmodellinstanz unterscheidet sich zunächst grundlegend zur initialen Detektion dadurch, daß keine neue Objektmodellinstanz angelegt werden muß. Vielmehr müssen die Modellstrukturen der Instanz an die ausgewählte Hypothese angepaßt werden. Dies gilt jedoch nur, falls die Qualität der Hypothese $q > q_{min}$ ist, ansonsten wird die Objektmodellinstanz verworfen. Aufgrund der Möglichkeit der Verwendung von geschätzten und nil-Szenenmerkmalen bei der Re-Detektion wird für eine Objektmodellinstanz auch dann noch eine Hypothese aufgestellt, wenn die Merkmale nur noch teilweise zu detektieren sind. Die Qualität q sinkt dabei jedoch ab. Wird in mehreren aufeinanderfolgenden Interpretationsschritten die Merkmale geschätzt, so nimmt q weiter ab. Die Bestimmung der Qualität wird über die Qualitäten der verwendeten Vorhersagen bestimmt, vgl. Kap. 3.6.4. Eine Objektmodellinstanz wird daher, auch wenn die Merkmale nicht mehr in den Bildern zu detektieren ist, nicht gleich verworfen. Es wird dann die komplette Objektmodellinstanz geschätzt, in dem ein “geschätzter” Interpretationsbaum verwendet wird.

Man erreicht hiermit implizit durch den Schritt der Hypothesengenerierung, daß eine Instanz bei kurzzeitiger Fremdverdeckung noch erhalten bleibt. Aufgrund der, als Schätzung verwendeten, Positionsvorhersagen der Objektmodellteile “bewegt” sich die Instanz weiter, so daß das Objekt auch feststehende Hindernisse passieren kann. In der Anwendung zur Personendetektion ist dies z.B. beim Passieren einer Säule der Fall.

Aufgrund der Verwendung der nil-Szenenmerkmale bei der Re-Detektion kann es sein, daß zur Bestimmung der Position der Objektmodellinstanz im Szenenmodell nicht das erste Objektmodellteil in der inneren Struktur, für das ein primäres Merkmal definiert wurde, verwendet werden kann. Es muß dann zur Bestimmung von ${}^{wcs}\mathbf{T}_{obj}$ entsprechend der Vorgehensweise bei der initialen Detektion verfahren werden und das nächste Objektmodellteil omp_μ verwenden werden, bei dem in der Hypothese ein reales oder geschätztes Szenenmerkmal zugeordnet ist. Für alle Objektmodellteile, die in der hierarchischen, inneren Struktur auf dem Pfad von $omp_{0.1}$ zum omp_μ liegen, wird für die geometrische Struktur die Transformationsmatrix aus der Historie für den Zeitpunkt $t_{(-1)}$ übernommen. Weiterhin werden für die Rotationsanteile der Transformationsmatrizen der einzelnen Objektmodellteile die für die Hypothese bestimmten Rotationen gesetzt.

Die Translationsanteile der Transformationsmatrizen und die Größe der Volumenkörper der einzelnen Objektmodellteile können auch hier wieder übernommen werden, wenn ein zuvor explizit bekanntes und vermessenes Objekt detektiert wird. Ansonsten sind auch hier die Strukturen, wie im folgenden Abschnitt beschrieben, anzupassen.

[4] Bei der Anwendung auf ein Personenmodell auch individuelle Größe.

Auch bei der Re-Detektion ist weiterhin noch der Zeitstempel, der dem Interpretationszyklus durch die verwendeten Bilder zugewiesen ist, für die Objektmodellinstanz als oberster Eintrag in T für t_0 zu übernehmen.

3.8.2 Adaption der Modellstrukturen

Wie im letzten Abschnitt erläutert, kann nur dann von einem Größen-invarianten Objektmodell ausgegangen werden, wenn das zu detektierende Objekt mit seinen spezifischen Ausmaßen bekannt und vermessen ist. Bei der Anwendung zur Bewegungserfassung von Personen kann hierzu ein individuelles Objektmodell erstellt werden.

Sind die zu detektierenden Objekte, abgesehen von ihrer grundlegenden Struktur und Erscheinung nicht bekannt, so wird ein Objektmodell mit mittlerer zu erwartender Größe und Ausdehnung verwendet. Nachdem in STABIL^{++} bei der 3D/3D-Zuordnung von Modellpunkten zu Szenenpunkten Toleranzen zugelassen werden, entstehen hierdurch Modell-Szenen-Differenzen, die es zu bewerten gilt.

Es ist zu beachten, daß die Differenzen jedoch zum einen durch Größenunterschiede zwischen Modell und realem Objekt, jedoch auch durch Ungenauigkeiten bei der Bestimmung der 3D Positionen der Szenenmerkmale begründet sein können. Die Art einer Adaption des Modells an die Struktur, die sich durch die Punkte im Szenenraum ergibt, ist daher anwendungsabhängig zu wählen.

Ist das Ziel der Anwendung, in dem zur Interpretation verwendeten Videobild das Objektmodell überlagert darzustellen, so können die Translationsanteile der Transformationen in den Objektmodellteilen einfach durch eine gewichtete Addition der Modell-Szenen-Differenz adaptiert werden. Bezeichnet man mit $\vec{t}_{omp_\mu} = [x, y, z]^T$ den für die Anpassung neu zu setzenden Translationsanteil der Transformationsmatrix ${}^{omp_\mu}\mathbf{T}_{omp_\nu}$, so gilt:[5]

$$\begin{array}{lcl} \vec{t}_{omp_\mu} & = & \vec{d}_{\mathbf{m}} + a \cdot \left(\vec{d}_{\mathbf{m}} - \vec{d}_{\mathbf{s}} \right) \\ \text{wobei} & & \\ \vec{d}_{\mathbf{s}} & = & {}^{\mathbf{s}}\vec{p}^{\,\nu}_{omp_\mu} - {}^{\mathbf{s}}\vec{p}^{\,\mu}_{omp_\mu} = {}^{\mathbf{s}}\vec{p}^{\,\nu}_{omp_\mu} \\ \vec{d}_{\mathbf{m}} & = & {}^{\circ}\vec{p}^{\,\nu}_{omp_\mu} - {}^{\circ}\vec{p}^{\,\mu}_{omp_\mu} = {}^{\circ}\vec{p}^{\,\nu}_{omp_\mu} \end{array} \tag{3.27}$$

In Anwendungen, deren Ziel es ist, die Größen der einzelnen Objektmodellteile zu vermessen,[6] sind zur Adaption die möglichen Fehler bei der Bestimmung der 3D Positionen der Szenenmerkmale zu berücksichtigen. Hierzu sollte das sich bewegende Objekt durch das System über einen längeren Zeitraum beobachtet werden. Man erhält somit in der Historie der einzelnen Objektmodellteile für jeden Interpretationszyklus die den primären Merkmalen zugeordneten Szenenmerkmale mit den zugehörigen 3D Positionen.

Das Modell ist in der Form anzupassen, daß die Größe der einzelnen Objektmodellteile so bestimmt wird, daß die Summe der Modell-Szenen-Differenzen über alle Objektmodellteile und alle Einträge in der Historie minimal werden.[7] Es ist dabei sicherzustellen, daß eine genügend große Anzahl von repräsentativen Mengen von Szenenmerkmalen aus der Historie verwendet werden. Diese Form der Adaption kann auch off-line, d.h. nach der eigentlichen Interpretation einer Bildsequenz durchgeführt werden. Jedoch kann dann eine Anpassung des Modells an das

[5] $\vec{t}_{omp_\mu}$ kann in ${}^{omp_\mu}\mathbf{T}_{omp_\nu}$ direkt gesetzt werden, wenn die $(\mathcal{TR})$-Notation verwendet wird,vgl. Glg. B.16.
[6] Z.B. zur Vermessung von Personen bei der Erfassung von anthropometrischen Daten.
[7] Methode der kleinsten Quadrate; (engl.) *least-squares*.

reale Objekt nicht schon während der Interpretation bei den Restriktionen und den Gütefunktionen auf der Modellebene zu einer restriktiveren Auswahl von Assoziationen und Hypothesen verwendet werden.

Zusätzlich zur Anpassung der Translationsanteile der Transformationsmatrizen und damit der Größe der Objektmodellteile, müssen auch die Volumenkörper der äußeren Objektmodellstruktur angepaßt werden. Liegen die beiden Koordinatensysteme zweier in der inneren Struktur aufeinanderfolgender Objektmodellteile nach der Adaption weiter auseinander und ist die z-Achse des Vorgängerobjektmodellteiles in diese Richtung orientiert, so muß der zugehörige Volumenkörper in Richtung der Rotationsachse vergrößert werden.

Aufgrund der Verwendung von Punkt- und Flächenmerkmalen wird die Oberfläche der Objektmodellteile nicht explizit bestimmt, so daß generell auch keine Anpassungen der Durchmesser der Volumenkörper vorgenommen werden kann. Hat ein Objektmodellteil jedoch mehrere Nachfolgeobjektmodellteile, wobei die Knoten / Gelenke zu diesen Teilen jeweils auf der Oberfläche des Objektmodellteils liegt, so kann die Ausdehnung des Volumenkörpers in mehrere Richtungen angepaßt werden. Bei der eingeführten Modellierung des menschlichen Körpers ist dies z.B. bei dem Objektmodellteil des Rumpfes der Fall. Mit der Position des Objektmodellteils des Rumpfes kann die Ausdehnung in Richtung der z-Achse bestimmt werden. Mit den Positionen der Objektmodellteile des linken und rechten Oberarms kann die Ausdehnung der abschließenden Ellipse des Volumenkörpers vol_{trCone} in Richtung der x-Achse bestimmt werden, vgl. Abb. 2.2.

3.9 Zusammenfassung

Basierend auf den in Kap. 2 definierten Modellen der Szene, der zu detektierenden Objekte und der Kamera als Sensoren wird mit dem in diesem Kapitel vorgestellten Interpretationsprozeß die Szene interpretiert. Aufgrund der Wiederholung der Interpretationszyklen erhält man für einzelne Objektmodellinstanzen in den Historien $HIST$ Daten, die zusammen mit den Modellstrukturen die “Haltung” und 3D Position des detektierten Objektes für aufeinanderfolgende Zeitpunkte repräsentieren.

Diese Daten sind das Resultat der Interpretation, welches angewendet und ausgewertet werden kann. Im Interpretationszyklus ist hierzu als letzter Teilschritt die durch die Akteure act auszuführenden Aktionen vorgesehen. Die Aktionen und somit die Akteure sind anwendungsabhängig zu definieren. Es kann z.B. für die Anwendung in einem Überwachungssystem allein die Position der Objektmodellinstanzen verwendet werden, um durch die 3D Lokalisation eines Objektes in einem definierten 3D Gefahrenbereich einen Alarm auszulösen. Für eine andere Anwendung, z.B. zur Beobachtung von Bewegungsabläufen, kann es wiederum interessant sein, die Winkel in den Gelenken / Knoten zwischen den Objektmodellteilen zu speichern, um diese später zu analysieren.

Im folgenden Kapitel wird zunächst mit einem Modell/Modell-Vergleich die Anwendung des Systems auf künstlich erzeugte Aufnahmen einer gehenden Person angewendet. Anhand dieser Experimente wird eine Untersuchung der Interpretationsgüte aufgezeigt. Daran schließt sich die Beschreibung von Anwendungen an. Hierzu werden einerseits Anwendungen zur Objektdetektion und -verfolgung und andererseits Anwendungen zur Bewegungserfassung erläutert.

4 Experimente und Anwendungen

4.1 Experimente mit Modell/Modell-Vergleich

Die Konfiguration des artikularen Objektmodells[1] ist durch die Rotationen in den Gelenken zwischen den einzelnen Objektmodellteilen bestimmt. Durch eine kontinuierliche Veränderung der Rotationswinkel und einer Veränderung der Position des Modells erhält man eine animierte Modellbewegung. Sind zudem Bewegungsmuster für die einzelnen Gelenkwinkel bekannt, so kann mit dem Objektmodell ein bestimmter Bewegungsablauf nachgebildet werden.

Projiziert man die äußere Objektmodellstruktur des animierten Objektmodells unter Berücksichtigung der Eigenverdeckung in die Bildebene einer fiktiven Kamera, so erhält man für jeden Zeitpunkt eine künstliche Aufnahme. Mit so erstellten Sequenzen liegen Aufnahmen eines künstlichen Objektes vor, für das die exakten Bewegungsdaten des Objektes bekannt sind. Auf die aus dem Modell heraus erstellten Sequenzen wird der Interpretationsprozeß angewendet. Vergleicht man die Konfiguration der detektierten Objektmodellinstanz mit denen des animierten Modells, so spricht man von einem Modell/Modell-Vergleich. Der Modell/Modell-Vergleich wird zur Verifikation des vorgestellten Konzepts und zur Bestimmung der Detektionsgüte angewendet.

Für den in diesem Abschnitt vorzustellenden Modell/Modell-Vergleich wurde die Modellierung des menschlichen Körpers zugrunde gelegt. Hierzu ist die in Kap. 2.3.3 vorgestellte Standardmodellierung verwendet worden. Die Gelenkwinkel für die Objektmodellteile der Arme und Beine wurden kontinuierlich so verändert, daß eine Gehbewegung animiert wurde.

Im folgenden Abschnitt wird zunächst die Animation des Modells und die Erstellung der künstlichen Aufnahmen dargestellt, daran schließt sich die Erläuterung der Detektion und der Bewegungserfassung an. Im weiteren werden ausgewählte Daten der erfaßten Bewegung den Daten der Animation gegenübergestellt.

4.1.1 Modellanimation

Basierend auf den von Murray in [Mur67] veröffentlichten Gelenkwinkelverläufen des Schulter-, Ellbogen-, Hüft- und Kniegelenkes bei Gehbewegungen, ist das Personenmodell animiert worden. Hierzu sind 37 zeitlich folgende Rotationswinkel für die entsprechenden Objektmodellteile gewählt worden, um eine Bewegung mit einer Länge von 1,5 Gehzyklen zu animieren. Bei einer Zeitdifferenz von 40 ms zwischen den Einträgen legt das animierte Modell in 1,52 s eine Stecke von zwei Metern zurück, vgl. Abb. 4.1.

Diese animierte Bewegung des Objektmodells wird von drei fiktiven Kameras aufgenommen, wobei ein Versuchsaufbau entsprechend Abb. 4.2 gewählt wurde.[2] Durch die Wahl von

[1] Bei der Anwendung auf Personen kann von der Haltung des Modells gesprochen werden.

[2] Mit den fiktiven Kameras wird eine perspektivische Projektion entsprechend des Lochkameramodells abgebildet, jedoch ohne radiale Verzerrung, vgl. Kap. 2.5.4.

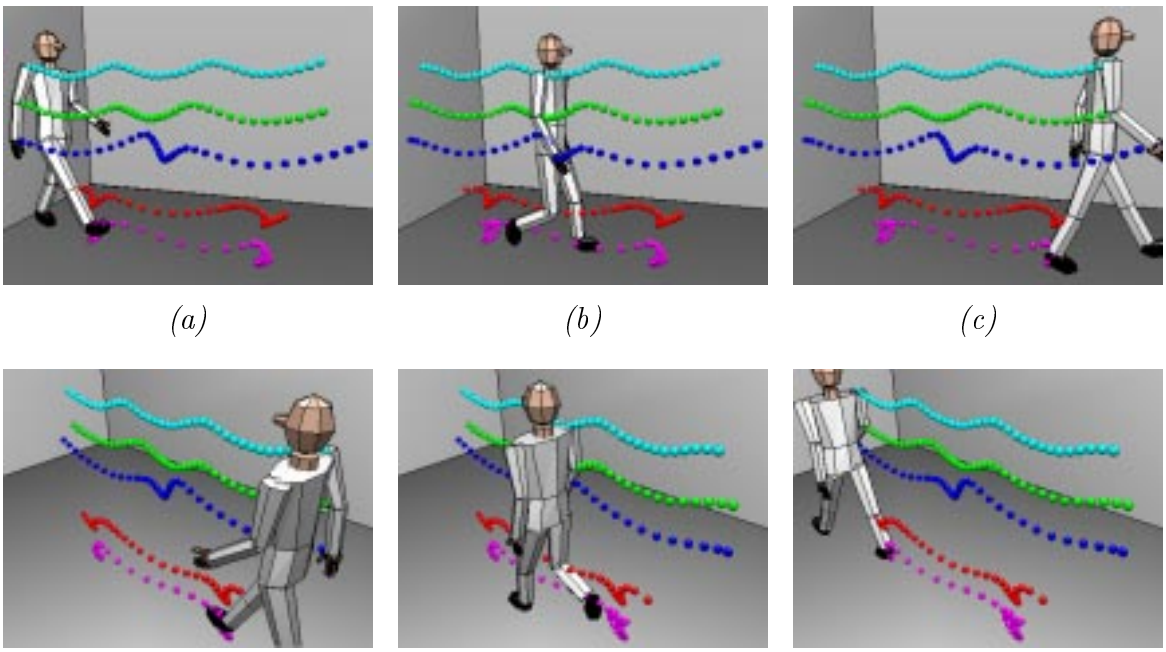


Abbildung 4.1: *Beobachtung des animierten Objektmodells aus zwei beliebigen Ansichten; Trajektorien für die Ursprünge der lokalen Koordinatensysteme der Objektmodellteile des rechten Oberarms, Unterarms, Unterschenkels, Fußes und der rechten Hand; Bildnummern der Sequenz: (a)/(d) 0, (b)/(e) 18, (c)/(f) 38.*

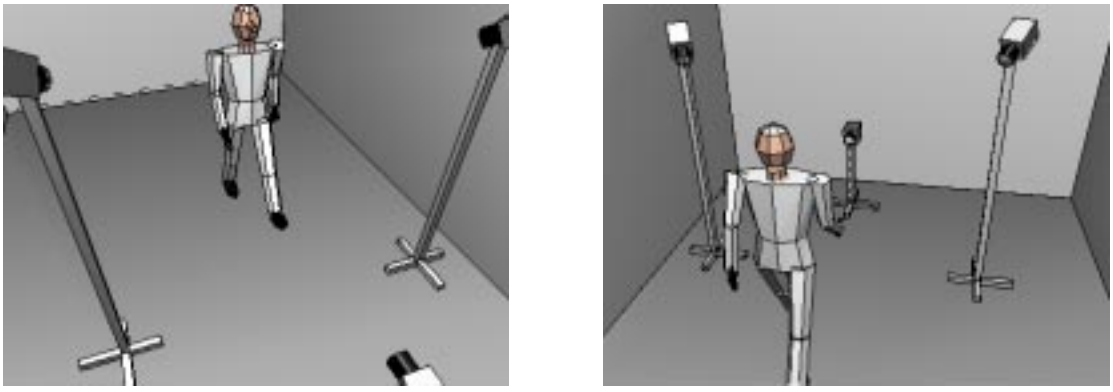


Abbildung 4.2: *Versuchsaufbau zur Beobachtung der animierten Gehbewegung mit 3 Kameras; das Objektmodell ist für die Bildnummer 38 der Bewegung abgebildet.*

drei Kameras wird sichergestellt, daß die Objektmodellteile zu jedem Zeitpunkt der Bewegung in mindestens zwei Kamerabildern sichtbar sind. Hierdurch wird gewährleistet, daß die 3D Positionen der Szenenmerkmale über den Stereo Ansatz vermessen und nicht über den Monoansatz geschätzt werden.

In den Projektionen der äußeren Objektmodellstruktur des animierten Modells werden neben den Volumenkörpern auch die primären Merkmale dargestellt. In Anlehnung an Anwendungen zur Bewegungserfassung unter ergonomischen Gesichtspunkten,[3] werden als Merkma-

[3] Vgl. die Darstellung einer entsprechenden Anwendung in Kap. 4.3.1.

Objektmodellteil	Nr.	Farbe	$\vec{t}$ [m]		
			x	y	z
Hüfte	0.1	cyan	0,00	-0,08	0,00
rechter Oberschenkel	1.1	grün	-0,06	0,00	0,00
Rumpf	1.2	grün	0,00	-0,08	0,05
linker Oberschenkel	1.3	grün	0,06	0,00	0,00
rechter Unterschenkel	2.1	gelb	0,00	0,00	0,00
rechter Oberarm	2.2	cyan	0,00	0,00	0,00
Hals	2.3	grün	0,00	-0,08	0,00
linker Oberarm	2.4	gelb	0,00	0,00	0,00
linker Unterschenkel	2.5	cyan	0,00	0,00	0,00
rechter Fuß	3.1	cyan	0,00	0,00	0,00
rechter Unterarm	3.2	gelb	0,00	0,00	0,00
Kopf	3.3				
linker Unterarm	3.4	cyan	0,00	0,00	0,00
linker Fuß	3.5	gelb	0,00	0,00	0,00
rechte Hand	4.1	cyan	0,00	0,00	0,00
linke Hand	4.2	gelb	0,00	0,00	0,00

Tabelle 4.1: *Farben der primären Merkmale des Modells und deren Verschiebungsvektoren $\vec{t}$ bei der Vollkörperdarstellung für den Modell/Modell-Vergleich.*

le farbige Markierungen gewählt, die als Bänder um die Gelenke gelegt werden oder als runde Punkte auf der Oberfläche des Körpers angebracht werden.

Bei der Animation werden als Merkmale farbige Kugeln dargestellt. Diese Kugeln liegen für Objektmodellteile, die einen kreisförmigen Querschnitt haben, mit ihrem Mittelpunkt im Ursprung des lokalen Koordinatensystems. Bei einem entsprechend großen Durchmesser der Kugeln ergibt sich in der Darstellung ein farbiges Band, welches sich am Knotenpunkt jeweils um die zwei angrenzenden Objektmodellteile legt. Für alle weiteren Objektmodellteile sind die Kugeln um einen, für das Modellmerkmal anzugebenden, Verschiebungsvektor $\vec{t}$ auf die Oberfläche des entsprechenden Volumenkörpers verschoben, vgl. Glg. 2.8. Die Farben der zugeordneten Merkmale und die Verschiebungsvektoren sind in Tab. 4.1 für die einzelnen Objektmodellteile aufgelistet.

In der Abb. 4.3 sind für die drei fiktiven Kameras jeweils fünf Bilder der erzeugten Sequenzen abgebildet, in denen die Volumenkörper des Objektmodells und die farbigen Merkmalskugeln dargestellt sind. Will man mit dem Modell/Modell-Vergleich eine Aussage über die Auswirkung der notwendigen Korrekturen der Verschiebungsvektoren $\vec{t}$ bei der Bestimmung der 3D Positionen der Szenenmerkmale treffen,[4] so kann hierzu ein “durchsichtiges” Objektmodell animiert werden, bei dem die Mittelpunkte der farbigen Kugeln, die die primären Merkmale darstellen, immer im Ursprung der lokalen Koordinatensysteme der Objektmodellteile liegen. In Abb. 4.4 ist aus einer Sequenz jeweils ein Bild der drei fiktiven Kameras abgebildet, wobei das projizierte Modell nur durch Gitterlinien dargestellt ist. Hierdurch sind die Merkmalskugeln innerhalb des Modells sichtbar. Eine geringe Störung bei der Extraktion der Merkmale in den Bilder ist jedoch auch noch hier zu erwarten. Dies ist zum einen durch die Gitterlinien und zum anderen durch die unterschiedliche Beleuchtung und die Reflexionen gegeben. Daher ist

[4] Vgl. Glg. 3.5 in Kap. 3.2.4 und Glg. 3.10 in Kap. 3.5.2.

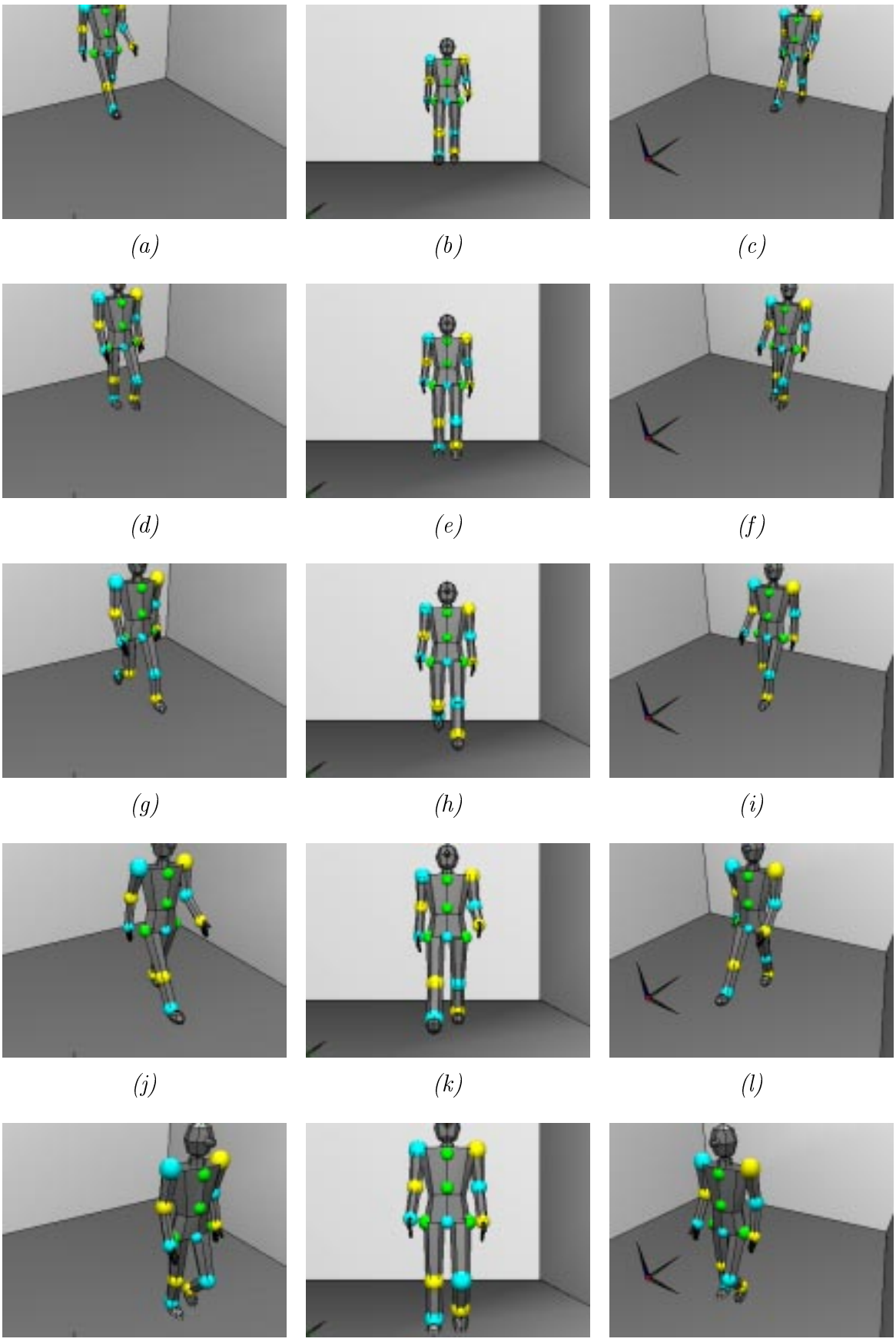
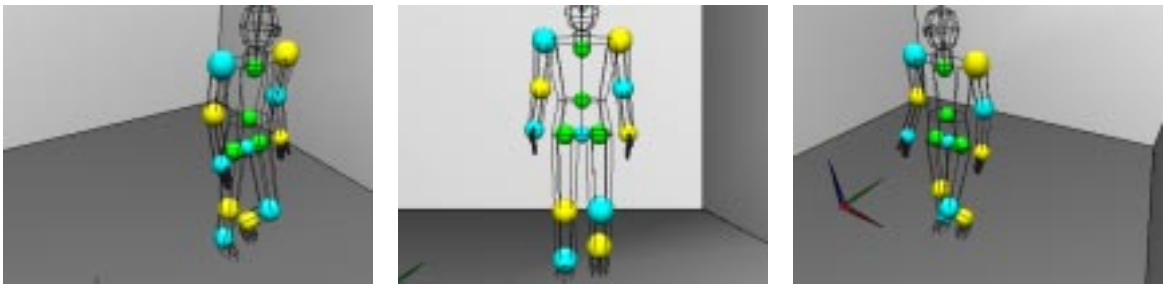
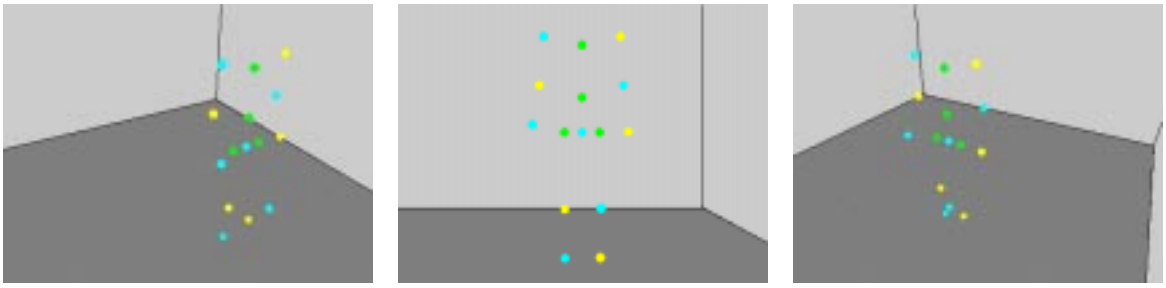


Abbildung 4.3: *Künstliche Aufnahmen von animiertem Bewegungsvorgang (1); Objektmodell als Vollkörper projiziert; Kameranummer: (a)/(m) 1, (b)/(n) 2, (c)/(o) 3, Bildnummer: (a)–(c) 0, (d)–(f) 8, (g)–(i) 16, (j)–(l) 24, (m)–(o) 32.*



(a) (b) (c)

Abbildung 4.4: *Künstliche Aufnahmen von animiertem Bewegungsvorgang (2); Objektmodell als Gitterkörper projiziert; Kameranummer: (a) 1, (b) 2, (e) 3, Bildnummer: jeweils 32.*



(a) (b) (c)

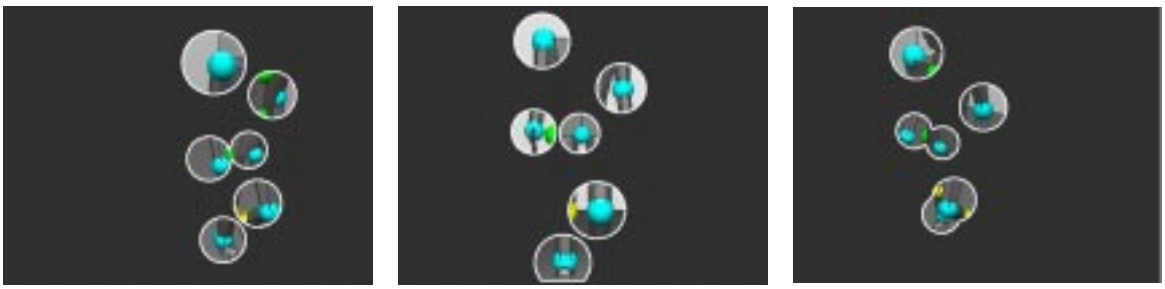
Abbildung 4.5: *Künstliche Aufnahmen von animiertem Bewegungsvorgang (3); Objektmodell ohne Körper projiziert; Kameranummer: (a) 1, (b) 2, (e) 3, Bildnummer: jeweils 32.*

Sequenz Nr.	Darstellung der Volumenkörper	Beleuchtung	Durchmesser der Merkmalskugeln [m]
1	Vollkörper	Schatten / Reflexionen	0,10 – 0,15
2	Gitterkörper	Schatten / Reflexionen	0,10 – 0,15
3	ohne Körper	homogen	0,05

Tabelle 4.2: *Aufnahmebedingungen der Testsequenzen für den Modell/Modell-Vergleich.*

Kamera Nr.	Brennweite	CCD-Chip Größe	Bildgröße
1	0.059 mm	$^1/_2$”	384 × 288
2	0.072 mm	$^1/_2$”	384 × 288
3	0.050 mm	$^1/_2$”	384 × 288

Tabelle 4.3: *Kameraparameter der drei fiktiven Kameras für den Modell/Modell-Vergleich.*



(a) (b) (c)

Abbildung 4.6: *Projizierte 3D Suchräume für die Modellmerkmale mit dem Basismerkmal "cyan"; Kameranummer: (a) 1, (b) 2, (c) 3, Bildnummer: jeweils 32.*

bei einer dritten Testsequenz auf die Darstellung der Volumenkörper verzichtet worden, so daß nur noch die Merkmalskugeln dargestellt sind, vgl. Abb. 4.5. Weiterhin ist bei dieser Sequenz eine homogene Ausleuchtung gegeben, so daß keine Störungen durch Schatten und Reflexionen auftreten. In der Tab. 4.2 sind für die drei unterschiedlichen Testsequenzen die Aufnahmebedingungen zusammengefaßt. Bei allen Testaufnahmen ist pro Kanal das Kamerarauschen durch ein weißes Rauschen mit einer Amplitude von fünf Grauwertstufen simuliert worden. Die Kameraparameter der drei fiktiven Kameras sind in Tab. 4.3 dargestellt.

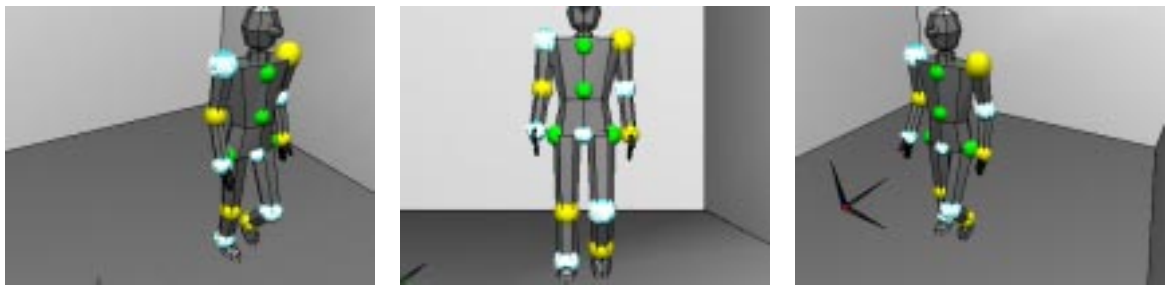
4.1.2 Detektion und Bewegungserfassung

In diesem Abschnitt werden die Schritte des Interpretationsprozeßes zur Erfassung der Bewegung des animierten Bewegungsablaufes dargestellt. Es wird hier jeweils die Testsequenz 1 mit der Vollkörperdarstellung verwendet. Stellvertretend für alle Merkmale werden die Schritte für die Merkmale mit dem Basisattribut der Farbe "cyan" in den Abbildungen gezeigt. Für die Merkmale mit den anderen Basisattributen der Farben "gelb" und "grün" ergeben sich ähnliche Darstellungen.

In der Abb. 4.6 sind die projizierten 3D Suchräume für die Objektmodellteile dargestellt, deren primäres Merkmal als Basisattribut die Farbe "cyan" haben. Aufgrund der hohen Bildwiederholrate von 25 Bildern / sek bei den künstlich erzeugten Bildsequenzen ist der Vorhersagezeitraum klein, so daß die Bewegungsrichtung sich nicht in der Form der Suchräume wiederspiegelt, vgl. auch Kap. 3.2 zur Bestimmung der Suchräume. Die Abbildungen zeigen daher kreisförmige Projektionen der Suchbereiche. Innerhalb dieser Bildregionen wird mit dem Farbklassifikator für die Extraktion der Bildmerkmale segmentiert, vgl. Kap. 3.4.2. Die als "cyanfarben" klassifizierten Bereiche sind in Abb. 4.7 hell dargestellt.

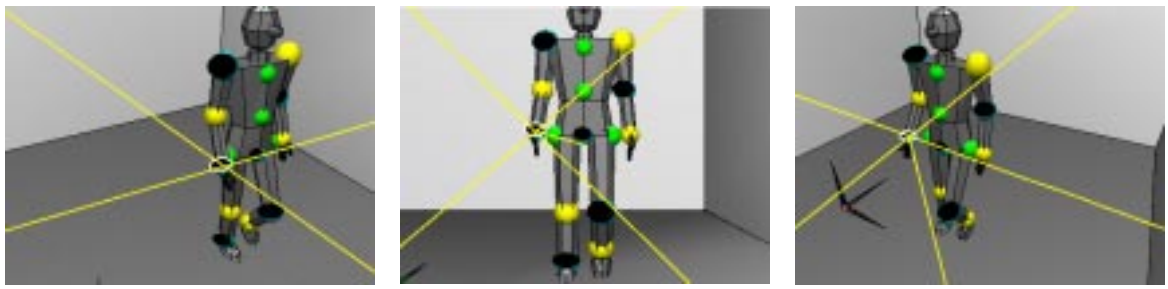
In einem weiteren Schritt werden die Bildbereiche ausgewählt, die als Bildmerkmale einer Farbellipse mit dem Basisattribut "cyan" angesehen werden. Diese ausgewählten Bildbereiche sind in der Abb. 4.8 als dunkle Ellipsen markiert. Die Bildmerkmale liegen in den Bildern aller drei Bildsequenzen vor, so daß für den 2D/3D Übergang ein Mehrfachstereoansatz verwendet werden kann. Für das dem Objektmodellteil $obj_{4.1}$ der rechten Hand zugehörige Merkmal sind in der Abb. 4.8 die projizierten 3D Sichtstrahlen der Bildmerkmale aus den jeweils anderen Kameras in den Aufnahmen der drei fiktiven Kameras dargestellt. Die 3D Position des zugehörigen Szenenmerkmals ergibt sich aus der Mittelung der quasi Schnittpunkte der drei Strahlen, vgl. Kap. 3.5.3.

Bei der Generierung der Hypothesen wird aus den Positionen der 3D Szenenmerkmale und



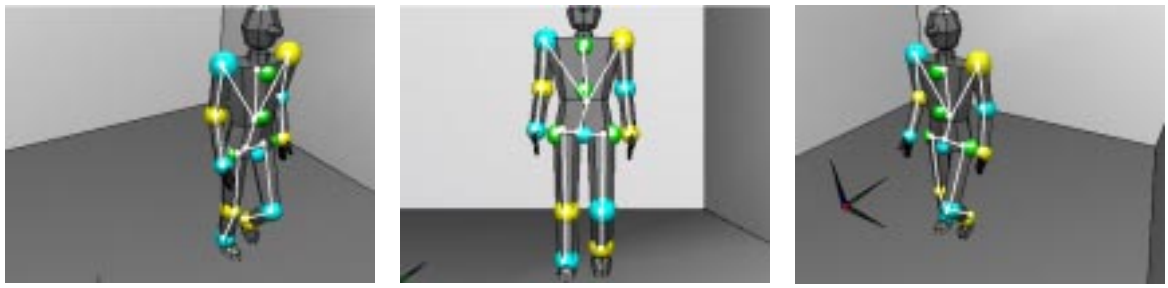
(a) (b) (c)

Abbildung 4.7: *Extraktion der Bildmerkmale: Durch den Farbklassifikator als "cyan-farben" segmentierte Bildbereiche; Kameranummer: (a) 1, (b) 2, (c) 3, Bildnummer: jeweils 32.*



(a) (b) (c)

Abbildung 4.8: *2D Bildmerkmale mit Basisattribut "cyan", dunkel dargestellt; Position des 3D Szenenmerkmals für das zu $omp_{4.1}$ passende Merkmal am Schnittpunkt der Sichtstrahlen; Kameranummer: (a) 1, (b) 2, (c) 3, Bildnummer: jeweils 32.*



(a) (b) (c)

Abbildung 4.9: *Detektierte innere Objektmodellstruktur für Testsequenz 1; Kameranummer: (a) 1, (b) 2, (c) 3, Bildnummer: jeweils 32.*

unter Berücksichtigung der Modellstruktur die Lage der inneren Struktur bestimmt, vgl. Kap. 3.6. Für das Bildtripel mit der Bildnummer 32 ist in Abb. 4.9 die so detektierte innere Objektmodellstruktur den Sequenzbildern überlagert dargestellt. In dem weiteren Prozeß der Interpretation werden noch die geometrische und die äußere Modellstruktur berücksichtigt und die Gelenkwinkel zwischen den Objektmodellteilen bestimmt, vgl. Kap. 3.7 und Anh. A. Basierend auf der so ermittelten Haltung und Konfiguration des Modells ist in der Abb. 4.10 das

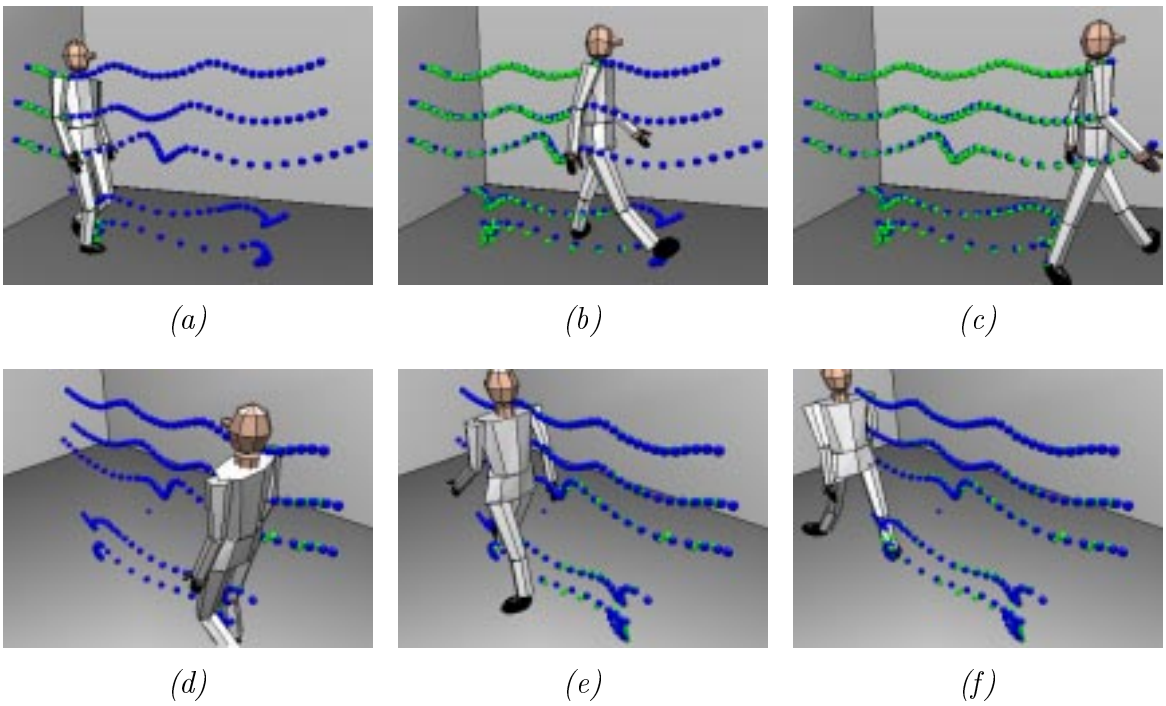


Abbildung 4.10: *Projektion des detektierten Objektmodells; dargestellt aus zwei beliebigen Ansichten; Trajektorien für die Ursprünge der lokalen Koordinatensysteme der Objektmodellteile des rechten Oberarms, Unterarms, Unterschenkels, Fußes und der rechten Hand; hell – Detektion; dunkel – Animation; Bildnummern der Sequenz: (a)/(d) 6, (b)/(e) 23, (c)/(f) 36.*

detektierte Objektmodell jeweils als Projektion zu drei verschiedenen Zeitpunkten der Sequenz abgebildet. Für die Objektmodellteile des rechten Oberarms, Unterarms, Unterschenkels, Fußes und der rechten Hand sind die Trajektorien bis zu dem entsprechenden Zeitpunkt durch helle Kugeln dargestellt. Zusätzlich sind mit dunklen Kugeln die zugehörigen Trajektorien der animierten Bewegung abgebildet, vgl. auch Abb. 4.1. Die Abweichungen zwischen der vorgegebenen animierten Bewegung und der erfaßten Bewegung wird anhand der Trajektorien nur schlecht sichtbar, daher sei hier auf den folgenden Abschnitt zur Gegenüberstellung der Bewegungsdaten verwiesen.

4.1.3 Gegenüberstellung der Bewegungsdaten

Für den Modell/Modell-Vergleich werden die Bewegungsdaten, die für das in den künstlich erzeugten Bildsequenzen detektierte Objekt erfaßt worden sind, mit den für die Animation zugrunde gelegten Daten verglichen. Es sollen hier zum einen die erfaßten 3D Positionen der Ursprünge der lokalen Koordinatensysteme und zum anderen die für die detektierte Objektmodellinstanz bestimmten Gelenkwinkel betrachtet werden. Weiterhin werden noch Daten zur Adaption der Modellstrukturen betrachtet. Für die Darstellungen sind exemplarisch einzelne Objektmodellteile ausgewählt worden.

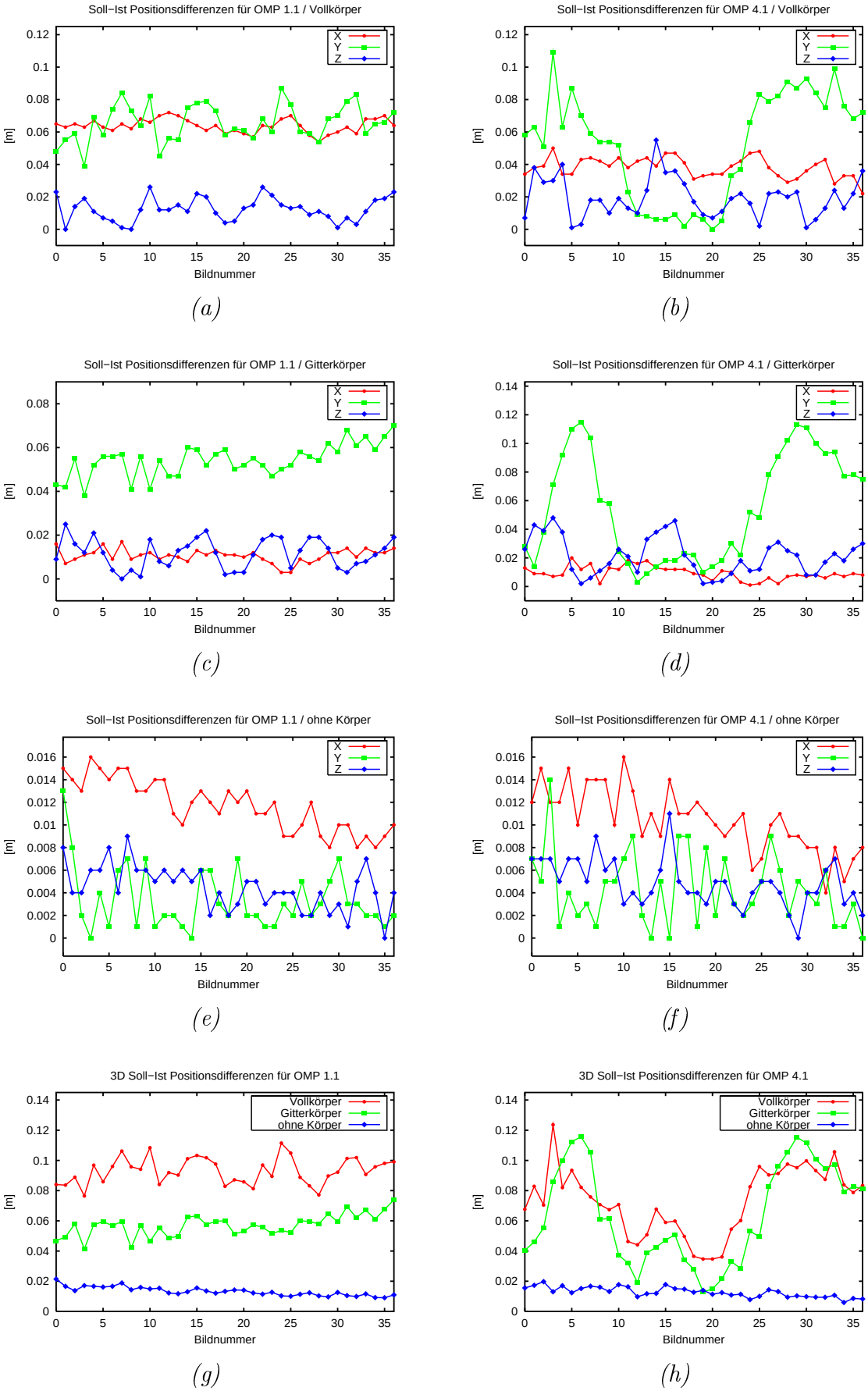


Abbildung 4.11: *Soll-Ist Positionsdifferenzen der Ursprünge der lokalen Koordinatensysteme; (a)/(g): $omp_{1.1}$ (rechter Oberschenkel); (b)/(h): $omp_{4.1}$ (rechte Hand); (a)–(f): Δx, Δy und Δz; (g)–(h): $\sqrt{\Delta x^2 + \Delta y^2 + \Delta z^2}$.*

3D Positionen

Zum Vergleich der Animation und der Detektion können die absoluten Differenzen zwischen den 3D Positionen der Objektmodellteile im Weltkoordinatensystem gebildet werden. Bezeichnet man hierzu den 3D Punkt des Ursprungs des lokalen Koordinatensystems eines Objektmodellteils aus der Animation als ${}^{\circ}\vec{p}^{\,a}_{wcs} = [x^a_{wcs}, y^a_{wcs}, z^a_{wcs}]^T$ und den entsprechenden Punkt der detektierten Objektmodellinstanz als ${}^{\circ}\vec{p}^{\,d}_{wcs} = [x^d_{wcs}, y^d_{wcs}, z^d_{wcs}]^T$, so können Differenzen für die einzelnen Komponenten entsprechend

$$\begin{aligned} \Delta x &= \left| x^a_{wcs} - x^d_{wcs} \right| \\ \Delta y &= \left| y^a_{wcs} - y^d_{wcs} \right| \\ \Delta z &= \left| z^a_{wcs} - z^d_{wcs} \right| \end{aligned}$$

und für eine 3D Positionsdifferenz entsprechend

$$\Delta\vec{p} = \left\| {}^{\circ}\vec{p}^{\,a}_{wcs} - {}^{\circ}\,\vec{p}^{\,d}_{wcs} \right\|$$

gebildet werden.

In der Abb. 4.11 sind für die Objektmodellteile $omp_{1.1}$ des rechten Oberschenkels und $omp_{4.1}$ der rechten Hand die Diagramme mit Positionsdifferenzen dargestellt. In den oberen drei Reihen (a)–(f) sind jeweils die Positionsdifferenzen für die drei Komponenten abgebildet; (a)–(b): Vollkörperdarstellung, (c)–(d): Gitterkörperdarstellung, (e)–(f): ohne Darstellung des Körpers. Diese Reihen unterscheiden sich durch die Testsequenz, in der das Objekt detektiert wurde: Testsequenz 1 mit der Vollkörperdarstellung, Testsequenz 2 mit der Gitterkörperdarstellung und Testsequenz 3 ohne Darstellung des Körpers. In der untersten Reihe sind die 3D Positionsdifferenzen für die Daten aus allen drei Testseqeunzen abgebildet.

Es ist in den Diagrammen zu erkennen, daß die Genauigkeit, mit der die 3D Positionen bestimmt werden, um eine Größenordnung bei der dritten Testsequenz gegenüber den anderen beiden Sequenzen steigt. Zwischen den Diagrammen der ersten beiden Testsequenzen ist ein signifikanter Unterschied nur bei dem Objektmodellteil $omp_{1.1}$ des rechten Oberschenkels zu erkennen. Dies liegt darin begründet, daß bei dem Objektmodellteil $omp_{4.1}$ der rechten Hand der Mittelpunkt der Merkmalskugel im Ursprung des lokalen Koordinatensystems liegt, vgl. Tab. 4.1. Für das Objektmodellteil $omp_{1.1}$ ist hingegen ein Verschiebungsvektor $\vec{t}$ in Richtung der x-Achse des lokalen Koordinatensystems angegeben. Nach dem die animierte Gehbewegung in Richtung der negativen y-Achse des Weltkoordinatensystems verläuft, kann die Auswirkung des Versatzes des Merkmals in der Positionsdiffferenz für die x-Komponente beobachtet werden: In dem Diagramm in Abb. 4.11 (a) ist die Abweichung der x-Komponente mit den Abweichungen in der y-Komponente nahezu identisch. Für die Testsequenz mit dem Gittermodell liegt hingegen in Abb. 4.11 (c) die Abweichung der x-Komponente mit der Abweichung der z-Komponente auf einer Höhe.

Dieser Sachverhalt spiegelt sich auch in den 3D Positionsdifferenzen in den Diagrammen der Abb. 4.11 (g) und (h) wieder: Für beide Objektmodellteile sind die Abweichungen bei der dritten Testsequenz am geringsten. Für das Objektmodellteil $omp_{4.1}$ kann für die erste und zweite Sequenz praktisch kein Unterschied festgestellt werden, wohingegen für das Objektmodell $omp_{1.1}$ die 3D Abweichung zwischen der ersten und der zweiten Sequenz auf die Hälfte sinkt.

Gelenkwinkel

Zur Beurteilung der gesamten Detektionsgüte und der Bewegungserfassung sollen in diesem Abschnitt Gelenkwinkelverläufe verglichen werden. Hierzu werden die Gelenkwinkel des rech-

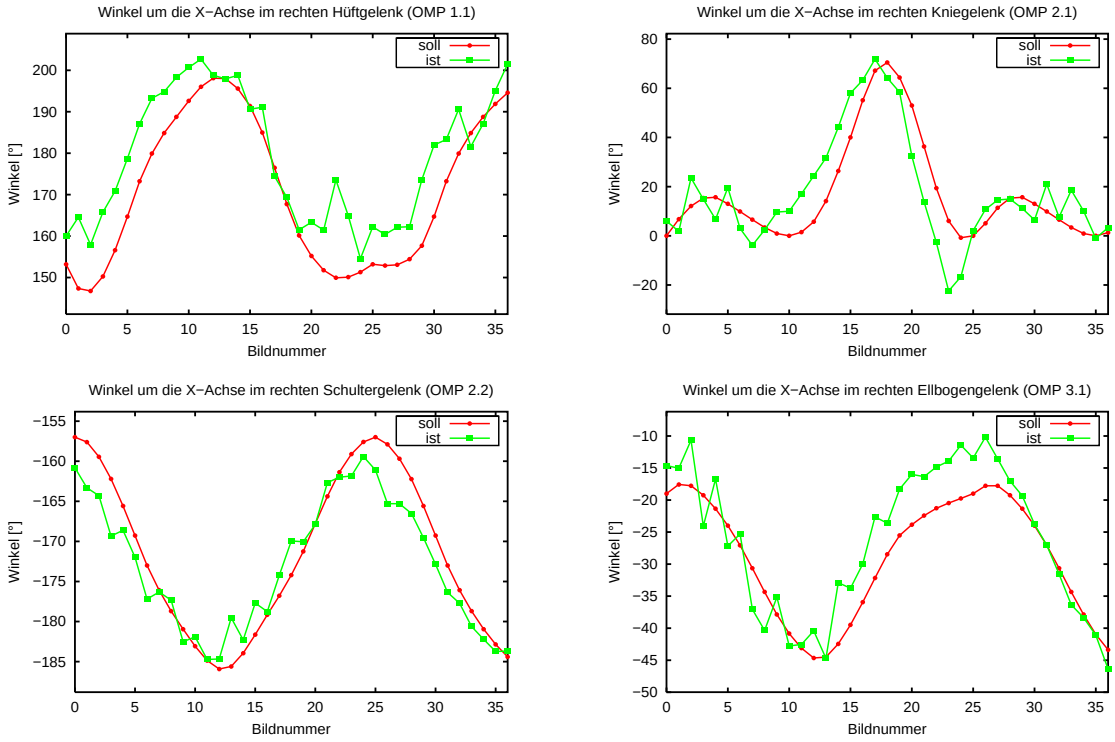


Abbildung 4.12: *Testsequenz 1 mit Vollkörperdarstellung: Winkel um die X-Achsen der Objektmodellteile $omp_{1.1}$, $omp_{2.1}$, $omp_{2.2}$ und $omp_{3.1}$; soll: Winkelvorgaben; ist: ermittelte Winkel.*

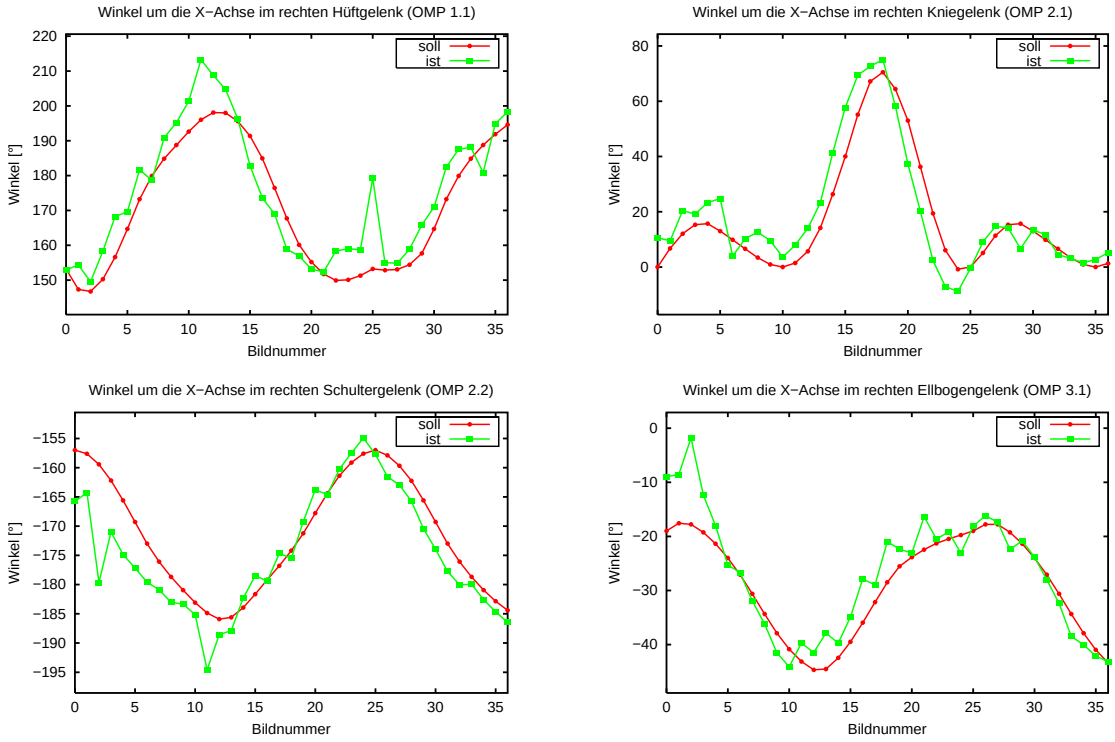


Abbildung 4.13: *Testsequenz 2 mit Gitterkörperdarstellung: Winkel um die X-Achsen der Objektmodellteile $omp_{1.1}$, $omp_{2.1}$, $omp_{2.2}$ und $omp_{3.1}$; soll: Winkelvorgaben; ist: ermittelte Winkel.*

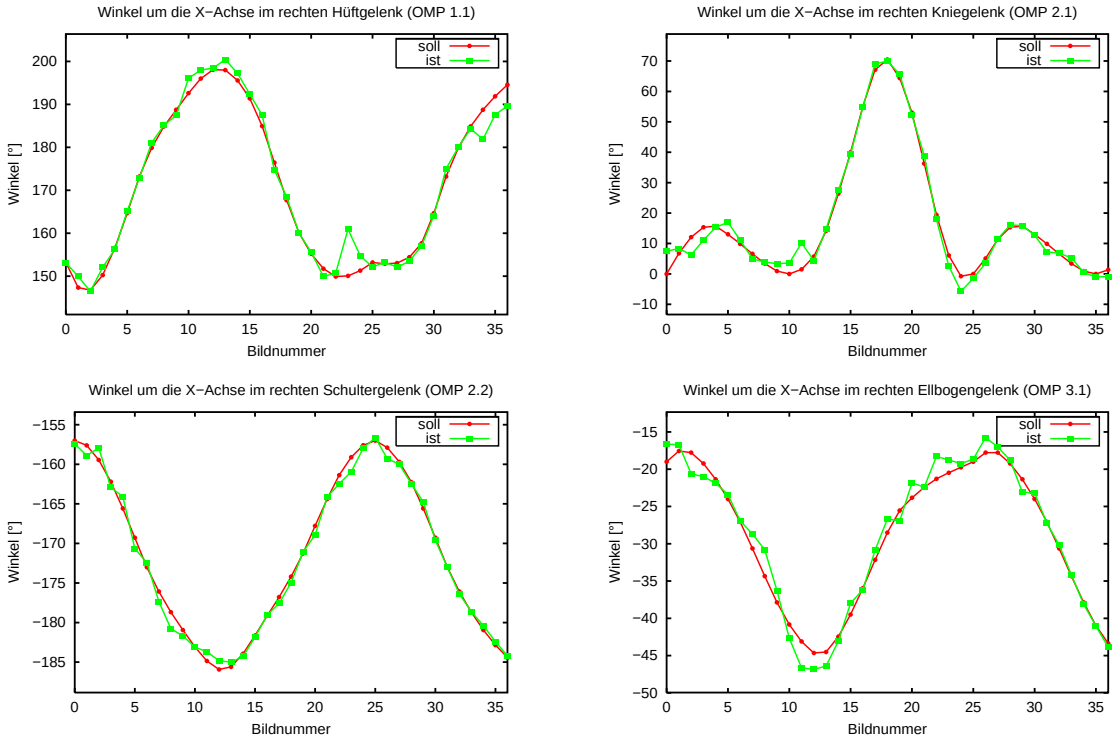
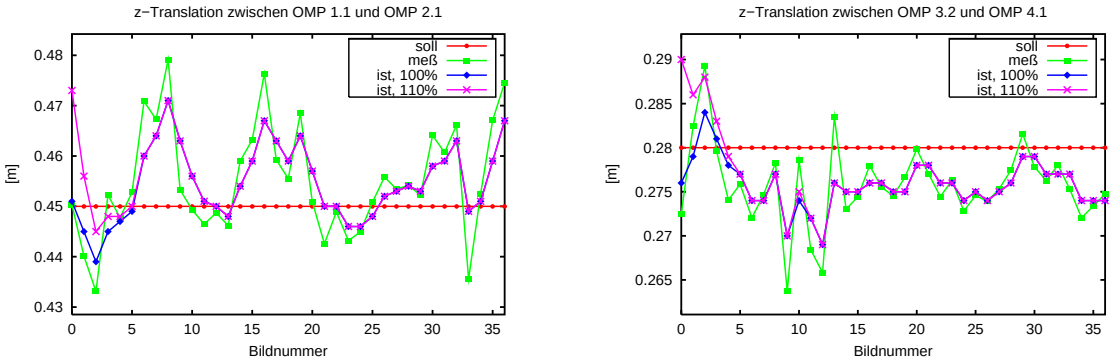


Abbildung 4.14: *Testsequenz 3 ohne Darstellung des Körpers: Winkel um die X-Achsen der Objektmodellteile $omp_{1.1}$, $omp_{2.1}$, $omp_{2.2}$ und $omp_{3.1}$; soll: Winkelvorgaben; ist: ermittelte Winkel.*

ten Schulter-, Ellbogen-, Hüft- und Kniegelenkes, mit denen das Objektmodell animiert wurde, und die für das detektierte Objektmodell ermittelten Gelenkwinkel betrachtet. Auch hier wird zwischen den drei Testsequenzen unterschieden: Abb. 4.12 zeigt die vier Gelenkwinkelverläufe für die Sequenz mit der Vollkörperdarstellung, Abb. 4.13 für die Gitterdarstellung und Abb. 4.14 für die Sequenz ohne Darstellung des Körpers. Die Kurven, die mit “soll” gekennzeichnet sind, entsprechen den Verläufen bei der Animation; die mit “ist” gekennzeichneten Kurven sind die durch die Detektion bestimmten Gelenkwinkel.

Die beste Übereinstimmung wird erwartungsgemäß bei der dritten Testsequenz erreicht. Jedoch folgen die Kurven der Gelenkwinkelverläufe der anderen beiden Testsequenzen ebenfalls den Kurven der Animation. Die Abweichungen fallen bei der Testsequenz mit der Gittermodelldarstellung geringer aus als bei der Vollkörperdarstellung. Dies kann insbesondere bei den Gelenkwinkeln des Hüftgelenkes festgestellt werden. Dies liegt darin begründet, daß das Merkmal des zugehörigen Objektmodellteils $omp_{1.1}$ des rechten Oberschenkels in der Testsequenz 1 aus dem Ursprung des lokalen Koordinatensystems heraus verschoben ist, vgl. Tab. 4.1.

Die Ausreißer in den Gelenkwinkelverläufen, wie z.B. für das Kniegelenk in Abb. 4.12 bei Bildnummer 23, begründen sich auf Abweichungen in der 3D Positionsbestimmung. Der negative Winkel ergibt sich durch die durch die Zulassung von Überstreckungen bei der entsprechenden I-Komposition, vgl. Abb. 4.10 (b) und (e) und Anh. A.6. Sind die Winkel der zweiten Objektmodellteile in einer I-Komposition, z.B. für das Knie- oder Ellbogengelenk, nahe null, so kann es durch leichte Abweichungen einer 3D Position zu einem anderen Winkel im ersten Objektmodellteil der Komposition, wie z.B. des Hüft- oder Schultergelenkes, kommen. Vgl. hierzu die Drehung des linken Beines um den Oberschenkelknochen in den Abb. 4.10 (b)/(e) und (c)/(f).



(a) *(b)*

Abbildung 4.15: *Adaption der Modellstruktur: Länge der Objektmodellteile des rechten Oberschenkels (a) und des rechten Unterarms (b); jeweils für die Testsequenz 1 mit Vollkörperdarstellung*

Adaption der Modellstrukturen

Schließlich soll der Modell/Modell-Vergleich noch genutzt werden, um die Adaption der Modellstruktur aufzuzeigen. Hierzu ist für die Detektion ein Objektmodell verwendet worden, das in der Ausdehnung um 10% größer ist, als das Objektmodell, welches bei der Animation verwendet wurde. Diese vergrößerten Ausmaße beziehen sich zum einen auf die Translationsanteile der geometrischen Struktur, als auch auf die Ausdehnung der Volumenkörper.

Aufgrund der 3D Zuordnungstoleranzen wird auch mit dem zugrundegelegten größeren Objektmodell eine gültige Instanz in den künstlichen Aufnahmen gefunden. Die Instanz wird jedoch in ihrer Größe an die Messungen angepaßt. Hierzu kann z.B. das in Kap. 3.8.2 mit Glg. 3.27 beschriebene Vorgehen angewendet werden.

In der Abb. 4.15 sind die Translationen in Richtung der z-Achse zwischen den Objektmodellteilen $omp_{1.1}$ und $omp_{2.1}$ und den Objektmodellteilen $omp_{3.2}$ und $omp_{4.1}$ angegeben. Dies entspricht den Längen der Objektmodellteile $omp_{1.1}$ des rechten Oberschenkels und $omp_{3.2}$ des rechten Unterarms.

In den Diagrammen ist zum einen die konstante Soll-Länge angegeben, die der Länge der entsprechenden Objektmodellteile in dem zur Animation zugrunde gelegten Objektmodell entspricht. Zum anderen gibt die mit “meß” bezeichnete Kurve den 3D Abstand zwischen den 3D Positionen der Szenenmerkmale wieder, die dem entsprechenden Objektmodellteil und dem entsprechenden Vorgängerobjektmodellteil zugeordnet worden sind. Mit den beiden weiteren Kurven sind die nach der Adaption gesetzten Längen der Objektmodellteile einmal bei der Verwendung des Objektmodells in der Originalgröße und einmal bei der Verwendung des um 10% größeren Objektmodells angegeben. Die Kurven sind entsprechend mit “ist, 100%” und “ist, 110%” gekennzeichnet.

In beiden Diagrammen ist zu erkennen, daß bei der Detektion über die Bildsequenz hinweg bis zu dem Bild mit der Nummer sechs / sieben bei beiden Objektmodellteilen die Größenunterschiede des Modells durch die Adaption ausgeglichen worden sind. Weiterhin zeigen die Diagramme, daß durch die Adaption die Schwankungen, die sich aus den 3D Meßpositionen ergeben, durch die verwendete Adaption geglättet werden.

4.2 Anwendung zur Objektdetektion und -verfolgung

Bei der Objektdetektion werden zunächst Objekte in einer Szene erfaßt und deren 3D Position bestimmt. Werden die Objekte in einer Bildfolge beobachtet, so können mit dem vorgestellten Konzept weiterhin die einzelnen Objekte derart verfolgt werden, daß 3D Bewegungsbahnen erfaßt werden. Eine konkrete Anwendung ist die Beobachtung von Personen. Anwendungsgebiete ergeben sich überall dort, wo 3D Positionen von Personen bestimmt und verwendet werden müssen.

Ein klassisches Anwendungsgebiet zur Personendetektion und -verfolgung ist die videobasierte Überwachung in der Sicherheitstechnik. In den meisten Videoüberwachungsanlagen werden die Videosignale visuell durch das Wachpersonal ausgewertet. Lediglich durch mechanische, Infrarot- oder Videobewegungssensoren wird die Aufmerksamkeit auf einen bestimmten Monitor gelenkt oder wird eine Aufzeichnung gestartet. Soll eine Person in einem größeren zu überwachenden Bereich verfolgt werden, so muß durch manuelle Betätigung das Signal einer im Sichtbereich angrenzenden Kamera aufgeschaltet werden. Das gleiche gilt für die Verfolgung von Personen mit einer Schwenk- / Neigekamera, denn auch dort muß das Wachpersonal den Kameraschwenk manuell über ein Bedienpult ausführen.

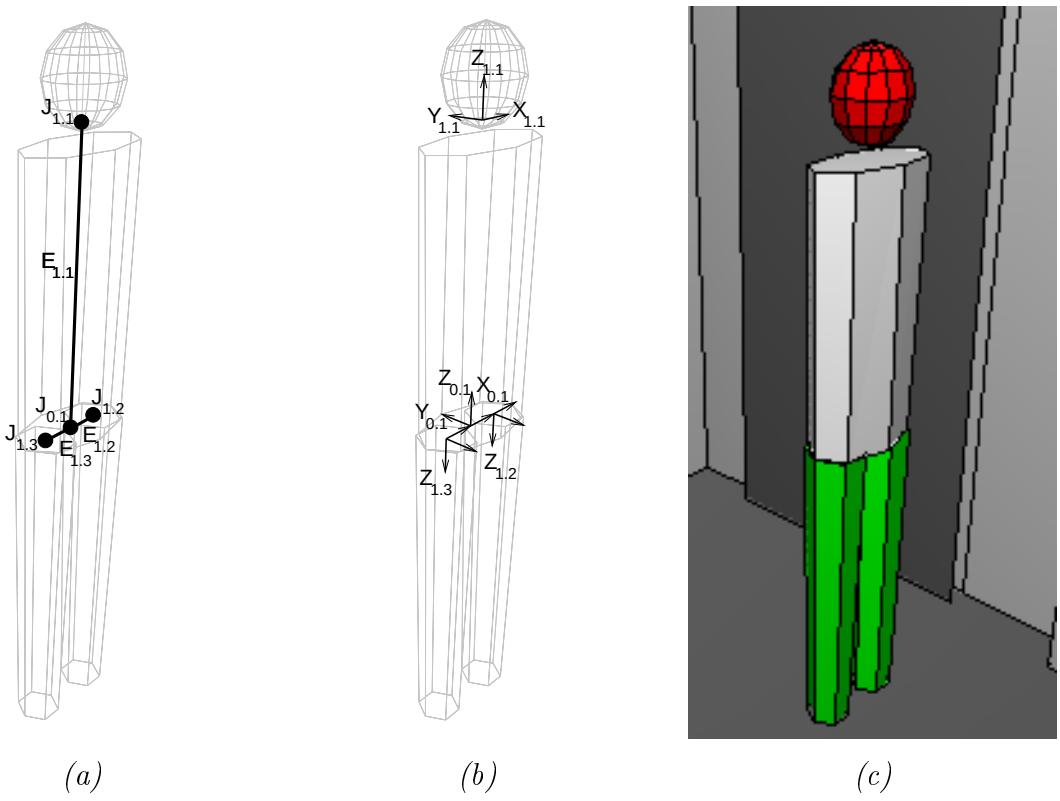
In diesem Abschnitt wird daher die Anwendung von STABIL^{++} zur Personendetektion und -verfolgung gezeigt, wobei für die detektierten Personen die 3D Positionen ermittelt werden, die einen Bezug zum Weltkoordinatensystem des Szenenmodells haben. Projiziert man die Einzelpositionen einer Bewegungsbahn in die xy-Ebene des Weltkoordinatensystems, so erhält man den von einer Person zurückgelegten Weg, der sich in einen Grundriß einzeichnen läßt. Werden in dem Grundriß des Gebäudes Alarmzonen markiert, so kann bei der Detektion einer Personenposition innerhalb dieser Zonen das Videosignal der entsprechenden Kamera gezielt auf einen Monitor oder auf eine Aufzeichnungskomponente aufgeschaltet werden. Verläßt eine Person den Sichtbereich einer Kamera, so kann durch die in STABIL^{++} implizit realisierte Objektübergabe[5] die Aufzeichnung auf eine weitere Kamera umgeschaltet werden.

Ein weiterer Aspekt der Anwendung ist die Nutzung der ermittelten 3D Position. Es können neben der Positionierung von Schwenk- / Neigekameras zur Erweiterung des Sichtbereiches des Systems auch zusätzliche aktive Kameras angesteuert werden. Diese Kameras können z.B. so positioniert werden, daß der mittlere Sichtstrahl auf die 3D Position zeigt, die für das Objektmodellteil des Kopfes einer Person ermittelt wurde. Weiterhin kann über die Größe des Objektmodellteils die Brennweite der Kamera so verändert werden, daß der Kopf formatfüllend abgebildet wird und Portraitaufnahmen der detektierten Personen aufgezeichnet werden können. Die aktive Ansteuerung einer weiteren Kamera oder das gezielte Aufschalten eines Videosignals zur Aufzeichnung sind Aktionen, die nach der Detektion auszuführen sind. Daher sind diese als Auswertung des Detektionsergebnisses zu bezeichnenden Aktionen in den Interpretationsprozeß als Akteure zu integrieren, vgl. Kap. 3.1.5.

4.2.1 Modellierung

Das Ziel von Anwendungen zur Personendetektion und -verfolgung ist es, die 3D Position von Personen in der zu observierenden Szene zu bestimmen. Die Haltung der Personen ist dabei zunächst unerheblich und soll auch nicht erfaßt werden. Somit wird für das artikulare Objektmodell in STABIL^{++} eine starre Konfiguration angenommen. Sollen aufrecht gehende und stehende Personen detektiert werden, so wird eine entsprechende Konfiguration in den Gelenken / Kno-

[5] Vgl. Kap. 3.3.1.



(a) *(b)* *(c)*

Abbildung 4.16: *(a) hierarchische, innere, (b) geometrische und (c) äußere Struktur des Objektmodells für die Personendetektion und -verfolgung.*

ten zwischen den Objektmodellteilen gesetzt. Diese Annahme entspricht im wesentlichen der Heuristik, die für die Tiefenschätzung über die Höhe eines Objektmodellteils mit dem monokularen Ansatz verwendet wird, vgl. Kap. 3.5.2. Weiterhin kann ein Objektmodell mit wesentlich weniger Objektmodellteilen verwendet werden, als in dem eingeführten Standardmodell für den menschlichen Körper, vgl. Kap. 2.3.

Es wird das Objektmodell auf die Objektmodellteile reduziert, für die primäre oder sekundäre Merkmale sinnvoll nutzbar und daher zu definieren sind. Dies sind je ein Modellteil für den Rumpf, den Kopf und die beiden Beine.[6] Das Objektmodellteil des Rumpfes umfaßt hier die bei der Standardmodellierung als Hüfte und als Rumpf bezeichneten Körperteile und ist das ausgezeichnete Wurzelelement $omp_{0.1}$ der hierarchischen inneren Modellstruktur. Alle weiteren Objektmodellteile liegen auf der zweiten Hierarchieebene. In der Abb. 4.16 sind die drei Modellstrukturen für das zu verwendende Objektmodell dargestellt, vgl. auch die Auflistung der Objektmodellteile in Tab. 4.4.

Für die Detektion von Personen mit videobasierten Systemen eignet sich im besonderen Maße der Kopf. Dies ist darin begründet, daß aufgrund der üblichen Beleuchtung von oben der Bereich des Kopfes am besten ausgeleuchtet ist. Weiterhin ist der Kopf im Vergleich zur Breite des Körpers relativ schmal, so daß auch bei Personengruppen im Bereich der Köpfe Verdeckungen erst später auftreten, als in den Bereichen des restlichen Körpers. Nimmt man weiterhin an, daß bekleidete Personen betrachtet werden, so zeichnet sich der Kopf auch durch die Hautfarbe des Gesichts aus.

Zur Unterscheidung zu anderen “hautfarbenen” Körperbereichen, wie z.B. der Hände und

[6] Die Beine können für die Anwendung jedoch auch zu einem Objektmodellteil zusammengefaßt werden.

Objektmodellteil	Nr.	primäres Merkmal	sekundäres Merkmal
Rumpf	0.1	–	Projektion ist im Vordergrund enthalten
Kopf	1.2	hautfarbene Ellipse	–
rechtes Bein	1.3	–	Ende des OMP ist auf dem Boden
linkes Bein	1.4	–	Ende des OMP ist auf dem Boden

Tabelle 4.4: *Objektmodellteile und deren Merkmale für die Personendetektion und -verfolgung.*

der Arme kann als weiteres Merkmal des Kopfes seine elliptische Form verwendet werden. Für das Objektmodell zur Personendetektion wird daher dem Objektmodellteil des Kopfes das primäre Merkmal einer "hautfarbenen" Ellipse zugeordnet. Zur Unterscheidung des Kopfes von den ebenfalls in Form von "hautfarbenen" Ellipsen im Videobild erscheinenden Händen bei langärmeliger Kleidung, wird ein sekundäres Merkmal für das Objektmodellteil des Rumpfes verwendet. Mit diesem Merkmal wird überprüft, ob unter dem vermeintlichen Objektmodellteil des Kopfes noch ein weiteres Körperteil im Videobild abgebildet worden ist. Hierzu wird an der durch das primäre Merkmal des Kopfes bestimmten 3D Position das Objektmodell in die Videobilder projiziert. In den Bildbereichen, in denen das Objektmodellteil des Rumpfes abgebildet wird, wird überprüft, ob dort Vordergrund detektiert worden ist, vgl. auch Kap. 2.4.6.

Weiterhin kann über die Größe der Ellipse, mit denen "hautfarbene" Objektteile abgebildet werden, und über die Höhe der Ellipsenmittelpunkte im Bild die Abbildung von Köpfen und Händen unterschieden werden. Die Hände werden von einer Person meist unterhalb des Kopfes gehalten. Wird für den zugehörigen Sichtstrahl eine Tiefenschätzung vorgenommen, so wird aufgrund der üblichen Anordnung der Kameras an der Decke eine 3D Position ermittelt, die näher an der Kamera liegt als die tatsächliche 3D Position der Hände. Wird das Modellteil des Kopfes an dieser vermeintlichen Position ins Bild projiziert, so ist, aufgrund des Größenunterschiedes zwischen Kopf und Händen, die projizierte Fläche größer als die extrahierte "hautfarbene" Fläche. Diese Überprüfung wird während des Interpretationsprozeßes bei der Zuordnung von Szenenmerkmalen zu Modellmerkmalen durch die Restriktion $restr^{(areaFnd)}(.)$ vorgenommen. Werden "hautfarbene" Regionen oberhalb des Kopfes extrahiert, so wird durch die Tiefenschätzung eine Position ermittelt, die in größerer Entfernung zur Kamera liegt. Liegen diese Positionen außerhalb des für das Szenenmodell definierten Observierungsbereich SSP_s, so können diese verworfen werden.

Ein weiteres sekundäres Merkmal, das bei der Personendetektion Verwendung findet, ist die Überprüfung der 3D Position der "Füße" zur xy-Ebene des Weltkoordinatensystems. Mit dieser Ebene ist der Boden des Observierungsraumes definiert, so daß überprüft wird, ob die Person mit den "Füßen" auf dem Boden steht. Nachdem für das zu verwendende Objektmodell keine Objektmodellteile der Füße definiert worden sind, ist dies gleichzusetzen mit der Überprüfung der 3D Position des Endes der Objektmodellteile der Beine.[7] Die Notwendig-

[7] Größte Ausdehnung in Richtung der z-Achse des lokalen Koordinatensystems.

keit dieses sekundären Merkmals ist nur gegeben, wenn Teile des zu überwachenden Bereiches durch Sichtbereiche von mehr als einer Kamera abgedeckt werden und somit für den 2D/3D Übergang der Stereoansatz zur Anwendung kommt. Dies ist darin begründet, daß hierdurch die Höhe des Kopfes über dem Boden nicht in einer heuristisch festgelegten Höhe[8] angenommen wird, sondern die 3D Position vermessen wird, vgl. Kap. 2.4.6.

4.2.2 Beispiele

In diesem Abschnitt werden beispielhafte Anwendungen von STABIL^{++} zur Personendetektion und -verfolgung gezeigt. Ein Anwendungsgebiet sind Überwachungsaufgaben in der Sicherheitstechnik, daher wird an drei Beispielen die Beobachtung eines Innenraumes gezeigt:

1. Verfolgung einer Person mit einer stationären Kamera, s. Abb. 4.17 und 4.18.
2. Verfolgung von mehreren Personen, wobei der Observierungsraum durch die Sichtbereiche von zwei stationären Kameras abgedeckt wird, s. Abb. 4.19 und 4.20.
3. Aktive Verfolgung einer Person mit einer Schwenk- / Neigekamera, s. Abb. 4.21 und 4.22.

Für die Darstellung werden jeweils die Eingabebilder von den Kameras gezeigt, wobei an den Positionen der detektierten Personen das Objektmodell überlagert dargestellt ist. Zusätzlich sind jeweils in weiteren Abbildungen Projektionen des Szenenmodells als virtueller 3D Raum abgebildet. In diesem Raum sind ebenfalls die äußeren Modellstrukturen der detektierten Objektmodellinstanzen projiziert worden. Weiterhin sind zur Darstellung der bei der Verfolgung erfaßten Bewegung die Trajektorien des Objektmodellteiles des Kopfes als Kugeln dargestellt, wobei sich die Darstellung der Trajektorien für die einzelnen Personen in der Helligkeit unterscheiden. Die Verweise auf die entsprechenden Abbildungen sind in der o.a. Auflistung angegeben.

Alle gezeigten Aufnahmen sind mit einer Installation des Systems im Labor des Bayerischen Forschungszentrums für Wissensbasierte Systeme, München entstanden. Es wurden im Hinblick auf die Aufbereitung der Beispiele in Form der graphischen Darstellungen eine *offline* Verarbeitung gewählt und daher File-Cameras verwendet, vgl. Kap. 2.5.3.[9] Hiermit war es möglich, eine feste Bildwiederholrate für die Verarbeitung festzulegen. Für die drei Beispiele sind in Tab. 4.5 neben den Bildwiederholraten Informationen zu den verwendeten Kameras gegeben. In dem ersten Beispiel und in dem dritten Beispiel sind mit der *SpeedDome*- und der *SpeedDome Ultra*-Kamera jeweils eine handelsübliche Kuppelkamera zum Einsatz gekommen. Im ersten Beispiel ist diese, jedoch als feststehende Kamera betrieben worden. Der Kameraschwenk im dritten Beispiel muß hier aufgrund der o.a. Problematik ebenfalls als Simulation mit File-Cameras gezeigt werden. Hierzu ist der Kameraschwenk bei der Aufnahme der Bildsequenz manuell ausgeführt worden.

Zu dem zweiten Beispiel – zur Verfolgung von mehreren Personen – ist noch anzumerken, daß für den 2D/3D Übergang von Bildmerkmalen zu Szenenmerkmalen der monokulare und der binokulare Ansatz verwendet wird. Immer dann, wenn das primäre Merkmal des Kopfes in beiden Kameras sichtbar ist, wird die 3D Position des Szenenmerkmals über den Stereoansatz gemessen. Ist die Person nur in einer Kamera sichtbar, so wird die Tiefeninformation über den monokularen Ansatz geschätzt. Zur Verdeutlichung, ob eine Person nur in einer oder in beiden Kameras sichtbar ist, sind in den Abb. 4.19 und 4.20 die Personen gekennzeichnet.

[8] In Richtung der z-Achse des Weltkoordinatensystems.

[9] Mit den graphischen Ausgaben wird die Verarbeitungsgeschwindigkeit erheblich herabgesetzt, so daß bei der Verwendung von Live-Cameras keine gleichmäßigen Trajektorien zu bestimmen sind.

Beispiel	Kamera-hersteller	Kameratyp	CCD-Chip-größe	Brennweite	Bildwieder-holrate
1	Sensormatic / AD	SpeedDome	$1/2$”	8,0 mm	6
2	JAI	730	$1/2$”	8,5 mm	12,5
3	Sensormatic / AD	SpeedDome Ultra	$1/4$”	4,0 mm	8,33

Tabelle 4.5: *Kameraspezifikationen für die Beispiele der Personendetektion und -verfolgung.*

Ansicht	ψ	θ
1	318,8°	23,4°
2	302,8°	29,7°
3	292,4°	34,9°

Tabelle 4.6: *Schwenkwinkel ψ und Neigewinkel θ der Ansichten bei der Personendetektion und -verfolgung mit aktiver Kamera.*

Für das Bildpaar mit der Nummer 25 ist die Person B in beiden Kameras sichtbar, die beiden anderen Personen jedoch nur in einer Kamera. Aufgrund der Personenbewegung ist im Bildpaar mit der Nummer 35 die Person B nur noch in der zweiten Kamera sichtbar, wohingegen die erste Person in beiden Kameras sichtbar ist. In Abb. 4.20 sind für jede der drei verfolgten Personen getrennt der zurückgelegte Weg eingezeichnet. Diese eindeutige Trennung basiert auf der Verwendung von 3D Suchräumen für die primären Merkmale der Objektmodellteile der Köpfe. Mit den Suchräumen wird die Bewegungsrichtung der Personen über Positionsvorhersagen berücksichtigt, so daß hierüber das Korrespondenzproblem der Zuordnung von ermittelten Szenenmerkmalen zur entsprechenden Objektmodellinstanz gelöst wird.

Bei dem dritten Beispiel, dem Beispiel zur aktiven Verfolgung mit einer Schwenk- / Neigekamera ist kein kontinuierlicher Kameraschwenk verwendet worden. Vielmehr sind für die Kamera ausgewählte Ansichten definiert worden, mit denen die Verkehrsfläche im Raum ausreichend abgedeckt wird. Für die in dem Beispiel zu verfolgende Personenbewegung sind drei Ansichten verwendet worden: Mit der ersten Ansicht wird der hintere Teil des Raumes nahe der Tür abgebildet, Bildnummern 1 - 43. In der zweiten Ansicht ist die Kamera weiter nach rechts geschwenkt und nach unten geneigt worden, Bildnummern 43 - 59. Der Schwenk nach rechts und das Neigen nach unten ist für die dritte Ansicht weiter geführt worden, ab Bildnummer 60. Die für die Ansichten verwendeten Schwenkwinkel ψ und Neigewinkel θ sind in Tab. 4.6 angegeben, vgl. hierzu auch Anh. C.5. In der Abb. 4.21 (d) ist eine Aufnahme in der Phase des Kameraschwenks dargestellt. Für die Zustände zwischen den verschiedenen Ansichten liegen keine exakten Kalibrierdaten vor. Zudem sind in dem Bild aufgrund der Bewegungsunschärfe keine gültigen Bildmerkmale zu extrahieren. Die Position der zu verfolgenden Person wird daher geschätzt. Dem eingeblendeten Objektmodell ist diese Position zugewiesen, jedoch sind noch die äußeren Kalibrierdaten verwendet worden, die für die erste Ansicht zutreffen. Der Wechsel der Ansichten kann auch in der in Abb. 4.22 dargestellten Trajektorie festgestellt werden. Dort sind zwei kleinere Unterbrechungen zu erkennen, denn für den Zeitraum der Kamerapositionierung kann keine gesicherte 3D Position lokalisiert werden. In den Abbildungen ist zusätzlich noch die Kamera dargestellt. Diese Darstellung hat nur symbolischen Charakter, da die verwendete Schwenk- / Neigekamera eine Kuppelkamera ist. Jedoch kann anhand der kastenförmigen Repräsentation des Kameragehäuses die unterschiedliche Orientierung für die drei Ansichten deutlich gemacht werden.

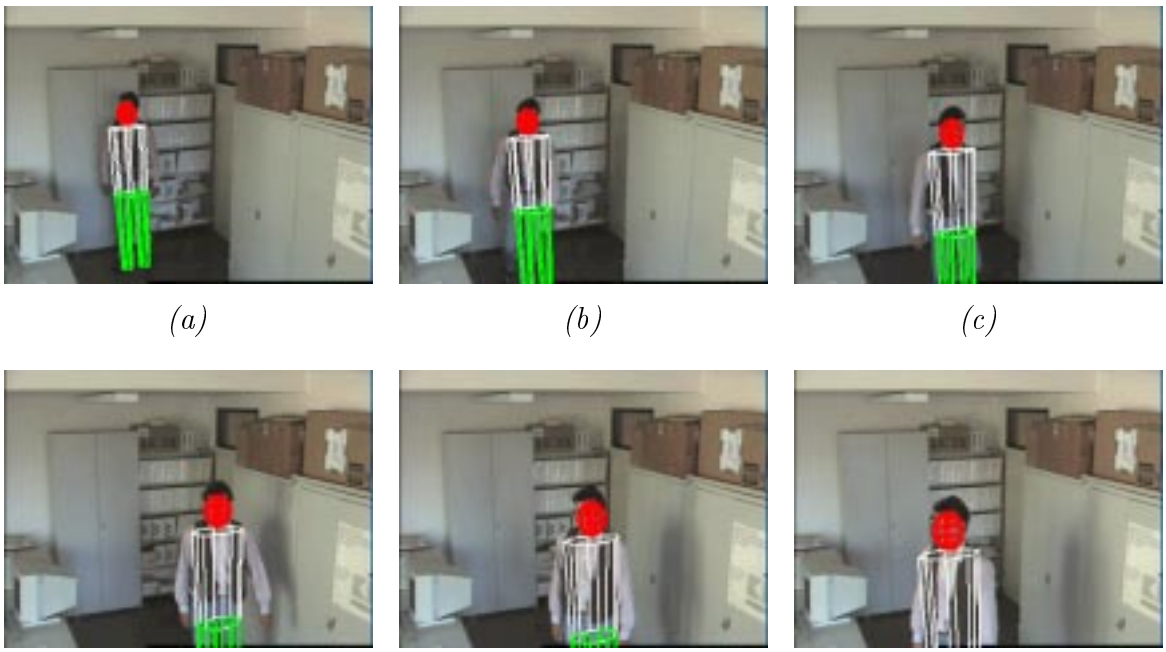


Abbildung 4.17: *Personendetektion und -verfolgung: Eingabebilder mit überlagertem Objektmodell an der ermittelten Position; Bildnummern: (a) 3, (b) 10, (c) 17, (d) 25, (e) 30, (f) 34.*

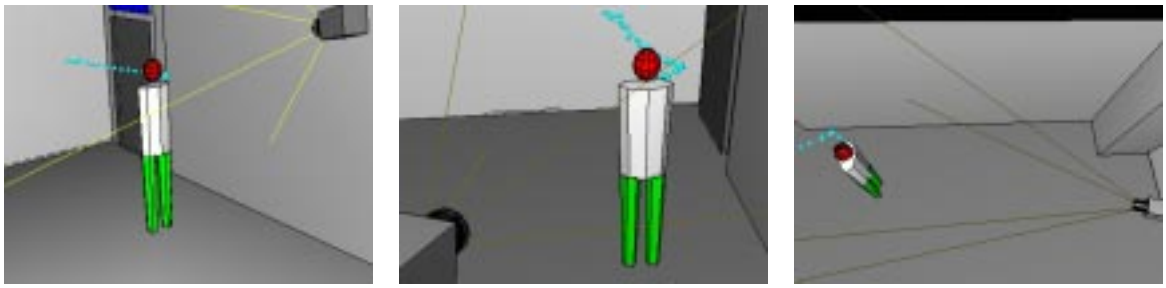


Abbildung 4.18: *Personendetektion und -verfolgung: virtueller 3D Raum aus drei beliebigen Ansichten mit detektierter Person und zurückgelegtem Weg bis Bildnr. 30.*

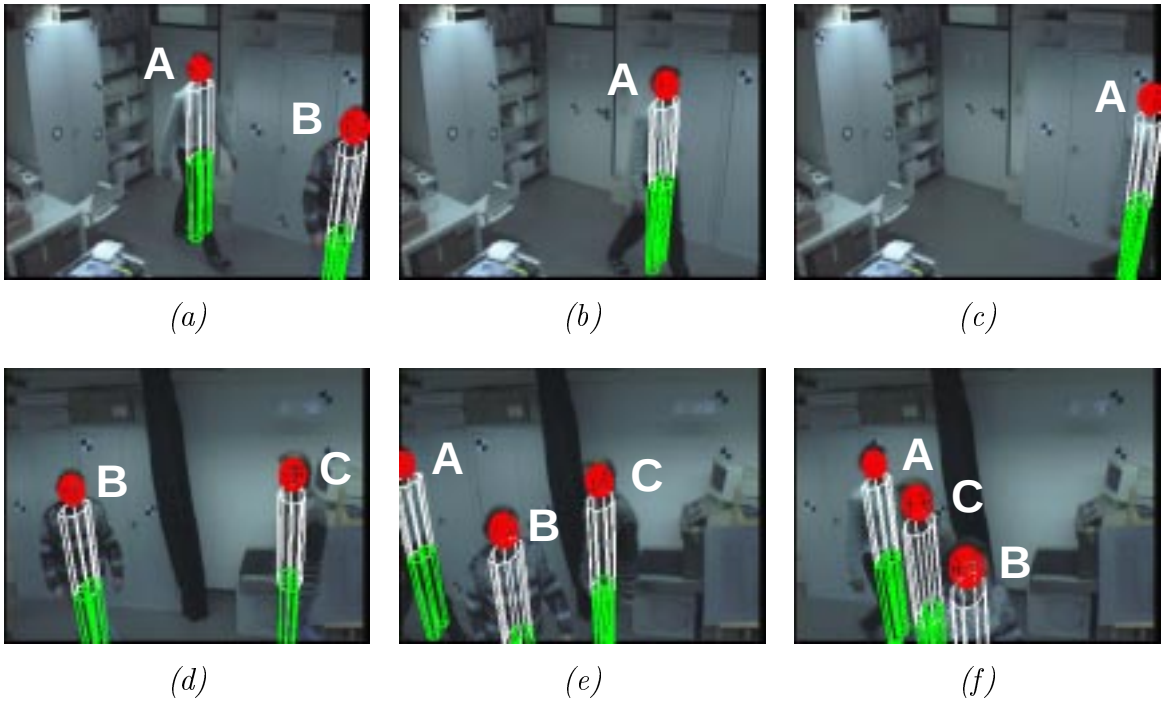


Abbildung 4.19: *Detektion und Verfolgung von drei Personen mit zwei Kameras: Eingabebilder mit überlagertem Objektmodell; Kameranummer: (a)–(c) 1, (d)–(f) 2; Bildnummer: (a)/(d) 25, (b)/(e) 35, (c)/(f) 45.*

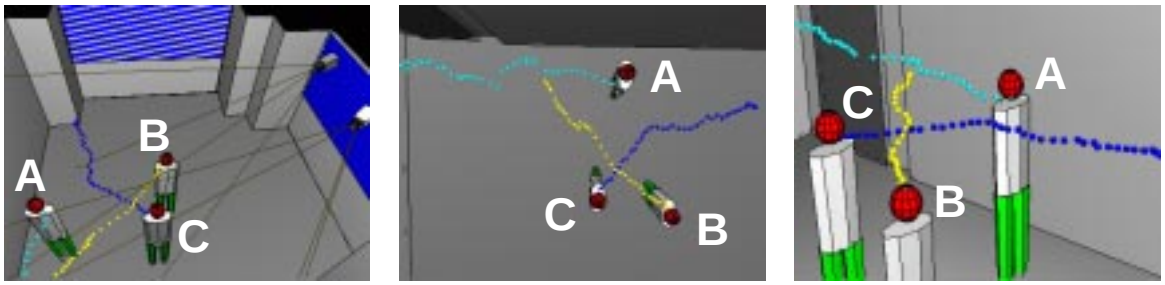
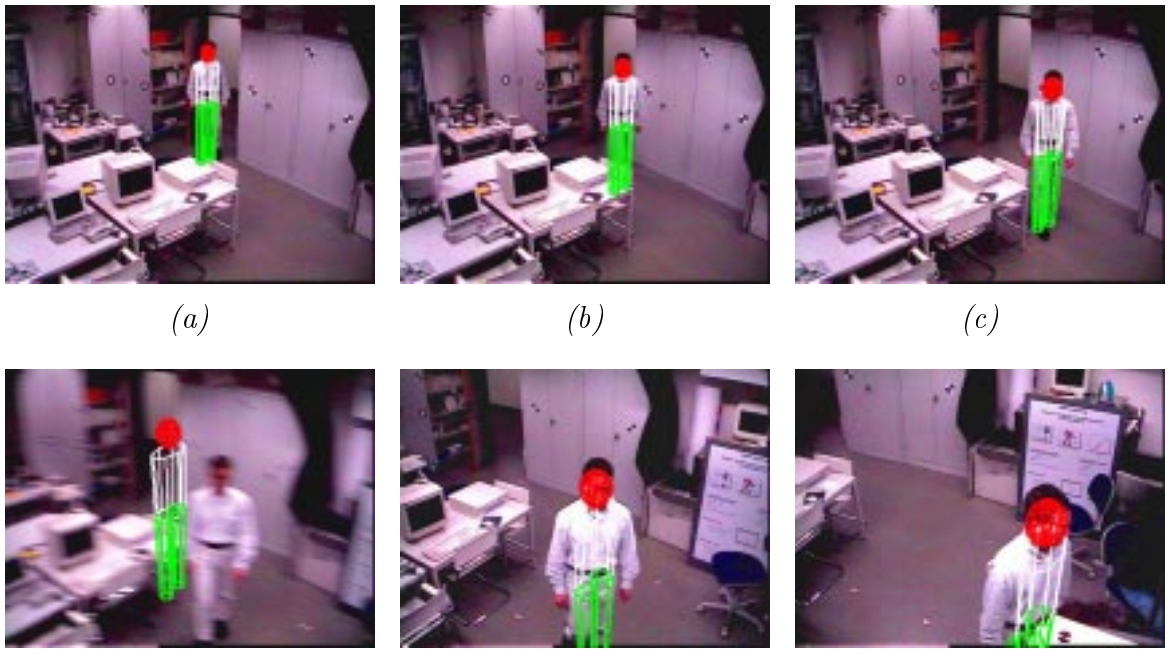


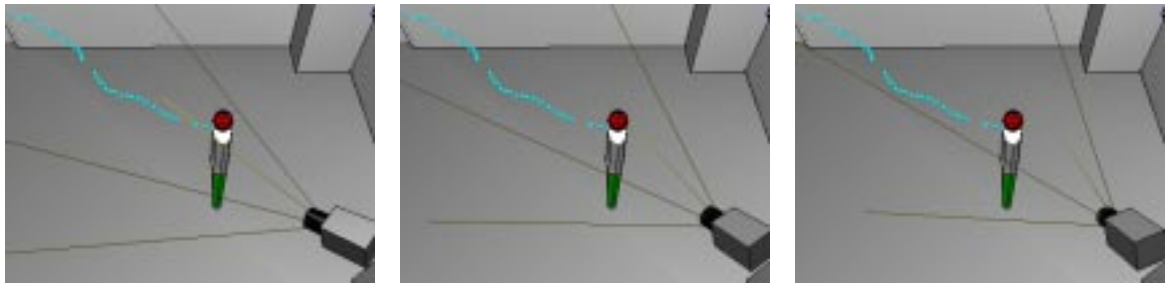
Abbildung 4.20: *Detektion und Verfolgung von drei Personen mit zwei Kameras: virtueller 3D Raum aus drei beliebigen Ansichten mit detektierten Personen und zurückgelegten Wegen bis Bildnummer 45.*



(a) (b) (c)

(d) (e) (f)

Abbildung 4.21: *Detektion und Verfolgung von einer Person mit aktiver Kamera: Eingabebilder mit überlagertem Objektmodell; Bildnummer / Ansicht: (a) 10 / 1, (b) 20 / 1, (c) 30 / 1, (d) 43 / 1-2, (e) 55 / 2, (f) 65 / 3.*



(a) (b) (c)

Abbildung 4.22: *Detektion und Verfolgung von einer Person mit aktiver Kamera: virtueller 3D Raum aus einer beliebigen Ansicht mit detektierter Person und zurückgelegtem Weg bis Bildnummer 65; unterschiedliche Kameraorientierung: (a) Ansicht 1, (b) Ansicht 2, (c) Ansicht 3.*

4.3 Anwendung zur Bewegungserfassung

Die Bestimmung der Konfiguration und Haltung von artikularen Objekten ist das Ziel von Anwendungen zur Bewegungserfassung. Dies steht im Gegensatz zu den Anwendungen zur Objektdetektion und -verfolgung. Daher muß hier ein weitaus detaillierteres Objektmodell verwendet werden. In der Regel werden alle die Körperteile als Objektmodellteil modelliert, die in den Gelenkpunkten verbunden sind und für die es gilt, die Gelenkwinkel zu ermitteln.

In diesem Kapitel wird zunächst eine Anwendung zur Erfassung von Bewegungen von Personen vorgestellt, wobei das Ziel eine Beobachtung von Bewegungsabläufen unter ergonomischen Gesichtspunkten ist. Daran schließen sich Überlegungen zu einer Anwendung in der Tiermedizin an.

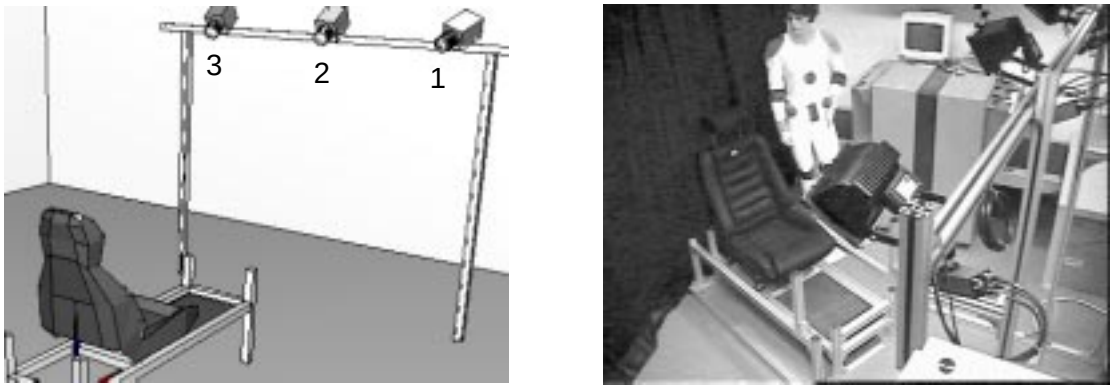
4.3.1 Anwendung für die Ergonomie

Zur Beurteilung von Körperhaltungen unter ergonomischen Gesichtspunkten werden Menschmodelle / Personenmodelle verwendet, vgl. z.B. [Sei94]. Sollen auch Bewegungsabläufe untersucht werden, so werden diese zunächst mit Videokameras aufgezeichnet und anschließend analysiert. Um exakte Daten über die Haltung zu den einzelnen Zeitpunkten einer Bewegung zu erhalten, müssen Gelenkwinkel bestimmt werden, mit denen die Haltung des Modells beschrieben wird. Ein Weg, der im Bereich der Ergonomie verwendet wird, ist die Einblendung des Menschmodells in die Videoaufnahmen mit anschließender manueller Anpassung der Haltung, vgl. [Arl99], S. 22.

Diese Vorgehensweise ist insbesondere bei der Beobachtung und Analyse längerer Bewegungsabläufe aufgrund des Aufwandes nicht praktikabel. Die automatische Erfassung der Gelenkwinkelkonfigurationen in Videobildfolgen kann diesen Prozeß erheblich beschleunigen. Um dies zu zeigen, ist mit dem System STABIL^{++} ein Versuchsaufbau zur Beobachtung eines Einstiegsvorgangs in einen PKW realisiert worden. Für diesen Aufbau ist ein Autositz auf einem Gestell montiert worden, wobei die Höhe der Sitzfläche der Höhe in einem realen PKW entspricht. Weiterhin sind Seitenschweller und Türholm, die den Einstiegsvorgang erschweren, durch weitere Profilstangen an dem Gestell simuliert worden.

Damit es, trotz der Bewegung der zu beobachtenden Person, zu jedem Zeitpunkt möglich ist, die 3D Positionen der Objektmodellteile über einen Stereoansatz zu lokalisieren, werden für den Versuchsaufbau drei Kameras verwendet. Hierdurch wird erreicht, daß trotz Verdeckungen immer in jeweils zwei Kameras die Merkmale eines Objektmodellteils sichtbar sind. Ein Überblick über den Versuchsaufbau ist in Abb. 4.23 gegeben.

Für die Erfassung der Bewegung und somit zur Bestimmung der Gelenkwinkel ist es notwendig, die 3D Positionen der Gelenke des zu beobachtenden menschlichen Körpers zu bestimmen. Daher wurden die Gelenkpunkte bei einer Testperson mit farbigen Bändern und runden Marken gekennzeichnet. Hierzu kann ein Vergleich zu dem Beispiel der animierten Gehbewegung im Kap. 4.1 gezogen werden. Das zu verwendende Objektmodell entspricht daher auch der in Kap. 2.3.3 vorgestellten Standardmodellierung des menschlichen Körpers. Die farbigen Markierungen werden als primäre Merkmale der Objektmodellteile verwendet, so daß in den Videobildern Bildmerkmale in der Form von farbigen Ellipsen extrahiert werden müssen. Die geometrische Form der Ellipsen wird jedoch nicht restriktiv überprüft, da sich die Form und Größe der in den Bildern sichtbaren Abbildungen der Bänder und Marken aufgrund der Körperbewegung verändern. Ein Überblick über die Objektmodellteile des verwendeten Modells, dem Basisattribut der primären Merkmale und der Position des Modellmerkmals im Bezug auf das jeweilige lokale Koordinatensystem ist in Tab. 4.7 gegeben.



(a) (b)

Abbildung 4.23: *Versuchsaufbau zur Beobachtung eines Einstiegsvorgangs mit 3 Kameras; (a) virtuelle Darstellung des Aufbaus, (b) reale Gesamtübersicht.*

Objektmodellteil	Nr.	Farbe	$\vec{t}$ [m]		
			x	y	z
Hüfte	0.1	cyan	0,00	-0,05	0,00
rechter Oberschenkel	1.1	grün	-0,05	0,00	0,00
Rumpf	1.2	grün	0,00	-0,07	0,05
linker Oberschenkel	1.3	grün	0,05	0,00	0,00
rechter Unterschenkel	2.1	gelb	0,00	0,00	0,00
rechter Oberarm	2.2	cyan	0,00	0,00	0,00
Hals	2.3				
linker Oberarm	2.4	gelb	0,00	0,00	0,00
linker Unterschenkel	2.5	cyan	0,00	0,00	0,00
rechter Fuß	3.1	cyan	0,00	0,00	0,00
rechter Unterarm	3.2	gelb	0,00	0,00	0,00
Kopf	3.3				
linker Unterarm	3.4	cyan	0,00	0,00	0,00
linker Fuß	3.5	gelb	0,00	0,00	0,00
rechte Hand	4.1	cyan	0,00	0,00	0,00
linke Hand	4.2	gelb	0,00	0,00	0,00

Tabelle 4.7: *Farben der primären Merkmale des Modells und deren Verschiebungsvektoren $\vec{t}$ zur Beobachtung eines Einstiegvorgangs.*

In der Abb. 4.24 sind die Eingabebilder von den drei Kameras dargestellt.[10] Hierbei handelt es sich um Aufnahmen, die mit einer Memory-Camera in den Hauptspeicher des Systems aufgenommen worden sind, vgl. Kap. 2.5.3. Hierbei wurde eine Bildwiederholrate von 8 Bildern / sek bei einer Bildgröße von 384 $\times$ 288 Bildpunkten erreicht. Außer der externen Synchronisation der Kameras sind keine weiteren Maßnahmen zur Sicherstellung einer gleichzeitigen Aufnahme getroffen worden.[11] Daher ist der zeitliche Versatz zwischen jeweils zwei Aufnah-

[10] Es wurden drei Kameras vom Typ JAI 730, $^1/_2$"-CCD-Chip mit 6 mm Objektiven verwendet.
[11] Es könnten jedoch auch synchronisierbare Digitalisierungskarten (*framegrabber*) verwendet werden.

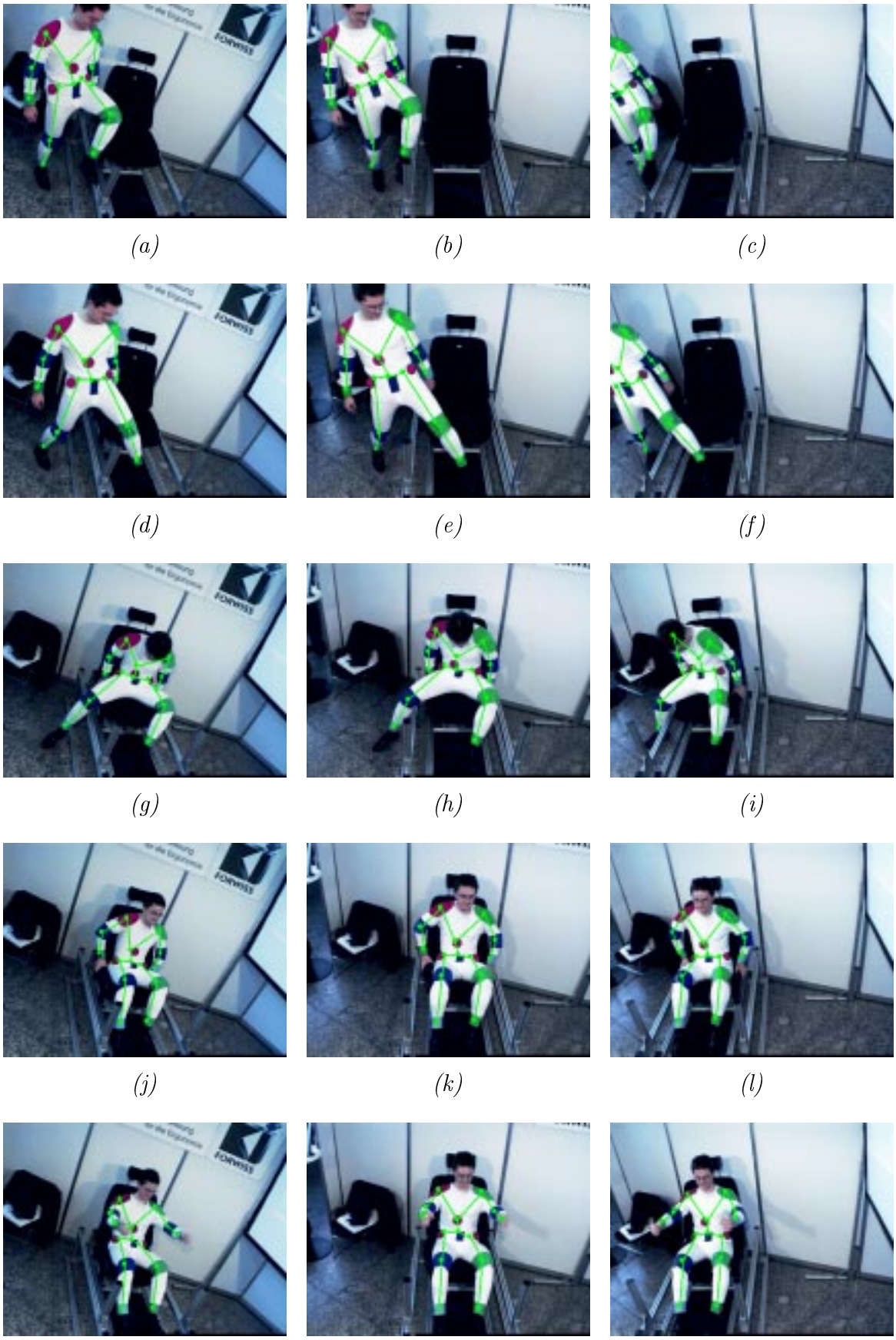


Abbildung 4.24: *Eingabebilder von drei Kameras mit überlagerter innerer Modellstruktur bei der Beobachtung eines Einstiegsvorgangs; Kameranr.: (a)/(m) 1, (b)/(n) 2, (c)/(o) 3, Bildnr.: (a)–(c) 10, (d)–(f) 20, (g)–(i) 30, (j)–(l) 45, (m)–(o) 60.*

men innerhalb eines Aufnahmetripels im Durchschnitt 40 oder 80 ms. In der Abbildung sind die Bilder der Sequenz bis zu der Bildnummer 60 dargestellt. Dies entspricht einer Länge von ca. 7,5 sek für den Bewegungsablauf des sich in den Sitz setzen und anschließend die Arme auf die Höhe des Lenkrades heben.

Die detektierte innere Objektmodellstruktur ist in der Abb. 4.24 zusätzlich dargestellt. Nachdem diese den Eingabebildern überlagert dargestellt ist, kann an dieser im besonderen erkannt werden, in welchen Kameras zu den verschiedenen Aufnahmezeitpunkten die Markierungen sichtbar sind und wo diese aufgrund von Verdeckungen unsichtbar bleiben. Für die ausgewählten Aufnahmezeitpunkte sind die Marken jeweils immer in mindestens zwei Ansichten sichtbar und daher können die 3D Positionen der Szenenmerkmale über den Stereoansatz bestimmt werden. Die Verwendung des monokularen Ansatzes für den 2D/3D Übergang von Bildmerkmalen zu Szenenmerkmalen ist bei dieser Anwendung aufgrund der zu erwartenden Ungenauigkeit nicht zu empfehlen. Daher wird hier, im Gegensatz zu den Anwendungen der Personendetektion und -verfolgung nur der Stereoansatz verwendet. Dies bedeutet, daß die 3D Position eines Objektmodellteiles geschätzt werden, wenn die zugehörigen Markierungen nicht in mindestens zwei Kameras sichtbar sind.

Zur Visualisierung des Ergebnisses der Bestimmung der Haltung und Konfiguration der Objektmodellinstanz kann die äußere Objektmodellstruktur projiziert werden. In der Abb. 4.25 ist hierzu das Modell und der Sitz auf dem Gestell jeweils aus der Sicht der drei Kameras dargestellt. Die Zeitpunkte entsprechen denen der in Abb. 4.24 dargestellten Eingabebilder. Es ist zu beachten, daß bei der Erfassung der Bewegung die Haltung des Kopfes nicht mit ermittelt wurde. Daher entspricht die Orientierung des Kopfes in der Projektion des Modells nicht der Haltung der Person, die in den Eingabebildern zu sehen ist. Vielmehr ist für die Gelenkwinkel zwischen den Objektmodllteilen des Rumpfes, des Halses und des Kopfes eine fest eingestellte Rotation angenommen worden.

Die Modellierung des Objektes durch 3D Volumenkörper und die Bestimmung der geometrischen Struktur erlaubt es, das Objektmodell, basierend auf den 3D Daten in eine virtuelle CAD Umgebung einzublenden. Dies ist in der Abb. 4.26 jeweils aus drei beliebigen Ansichten gezeigt. Hierdurch ist es schon zum Zeitpunkt der Konstruktion möglich, durch die Erfassung von Bewegungen in einfachen Konstruktionen, Bewegungsabläufe basierend auf realen Bewegungsdaten in einer Simulation der zu konstruierenden Umgebung zu betrachten. Dieser Gesichtspunkt gilt ebenfalls für die Darstellung in Abb. 4.27, denn dort sind zur Verdeutlichung des Bewegungsablaufes beim Einsteigen in einen PKW die Trajektorien von den Objektmodellteilen der Hände und Füße eingeblendet.

Zur Analyse der Bewegung können ebenso die Gelenkwinkelverläufe der detektierten Modellinstanz verwendet werden. In der Abb. 4.28 sind als Beispiel die Gelenkwinkelverläufe für die Winkel um die X-Achsen in den Hüftgelenken (a) ($omp_{0.1}$ / $omp_{1.1}$ und $omp_{0.1}$ / $omp_{1.3}$), sowie in den Kniegelenken (b) ($omp_{1.1}$ / $omp_{1.2}$ und $omp_{1.3}$ / $omp_{2.5}$) abgebildet. Anhand der Kurven ist nachzuvollziehen, daß die beobachtete Person zunächst mit dem rechten Bein einen kleinen Schritt nach vorn gemacht hat. Anschließend wurde das linke Bein angehoben und dabei das Knie gebeugt. Als der linke Fuß über den Seitenschweller hinweg gehoben war, wurde ab Bildnummer 10 das Bein wieder gestreckt und etwas abgesetzt. Ab der Bildnummer 20 bewegte die Person das Gesäß langsam nach unten. Dies bewirkt ein Abknicken in den Hüftgelenken. Der rechte Fuß steht bis zur Bildnummer 30 auf dem Boden, anschließend wird das rechte Bein angehoben, so daß sich ca. bei Bildnummer 36 die größte Beugung im rechten Hüftgelenk ergibt. Nachdem beide Füße auf dem Bodenblech abgestellt sind, werden ab Bildnummer 40 die Beine leicht gestreckt, was in den Diagrammen an den sich leicht verringernden Winkeln in den Kniegelenken erkannt werden kann.

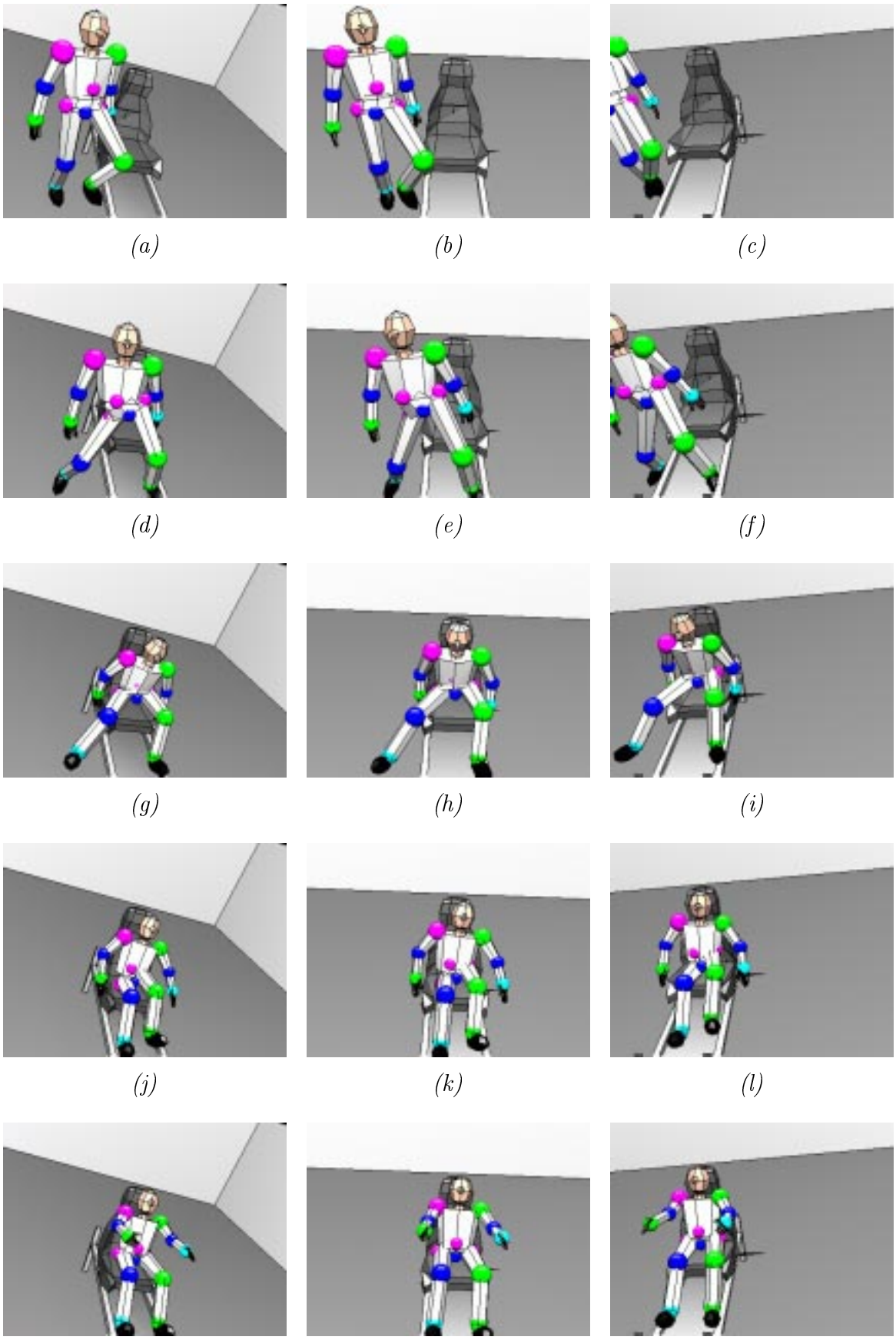


Abbildung 4.25: *Projektionen des detektierten Objektmodells aus Sicht der Kameras; Kameranummer: (a)/(m) 1, (b)/(n) 2, (c)/(o) 3, Bildnummer: (a)–(c) 10, (d)–(f) 20, (g)–(i) 30, (j)–(l) 45, (m)–(o) 60.*

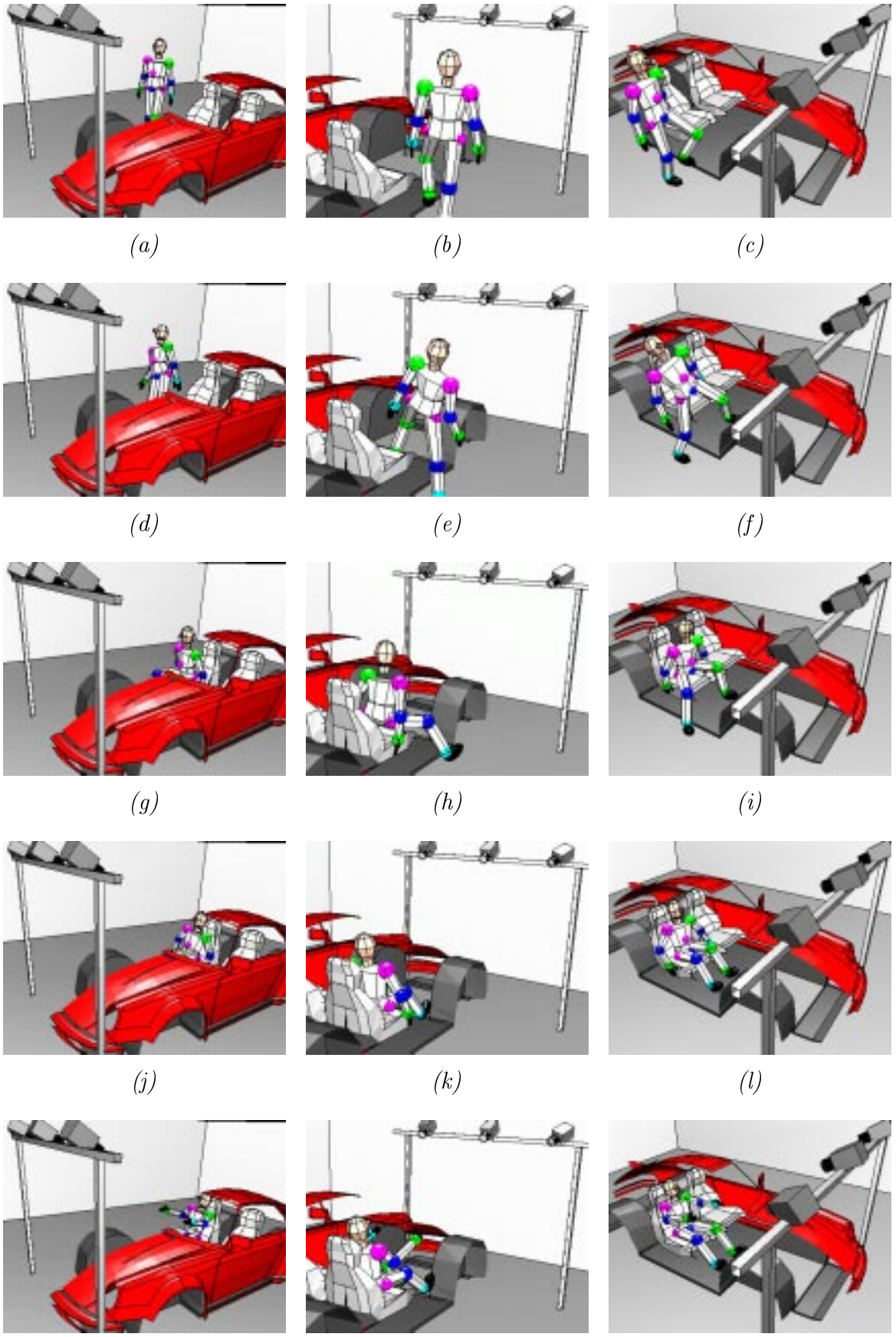
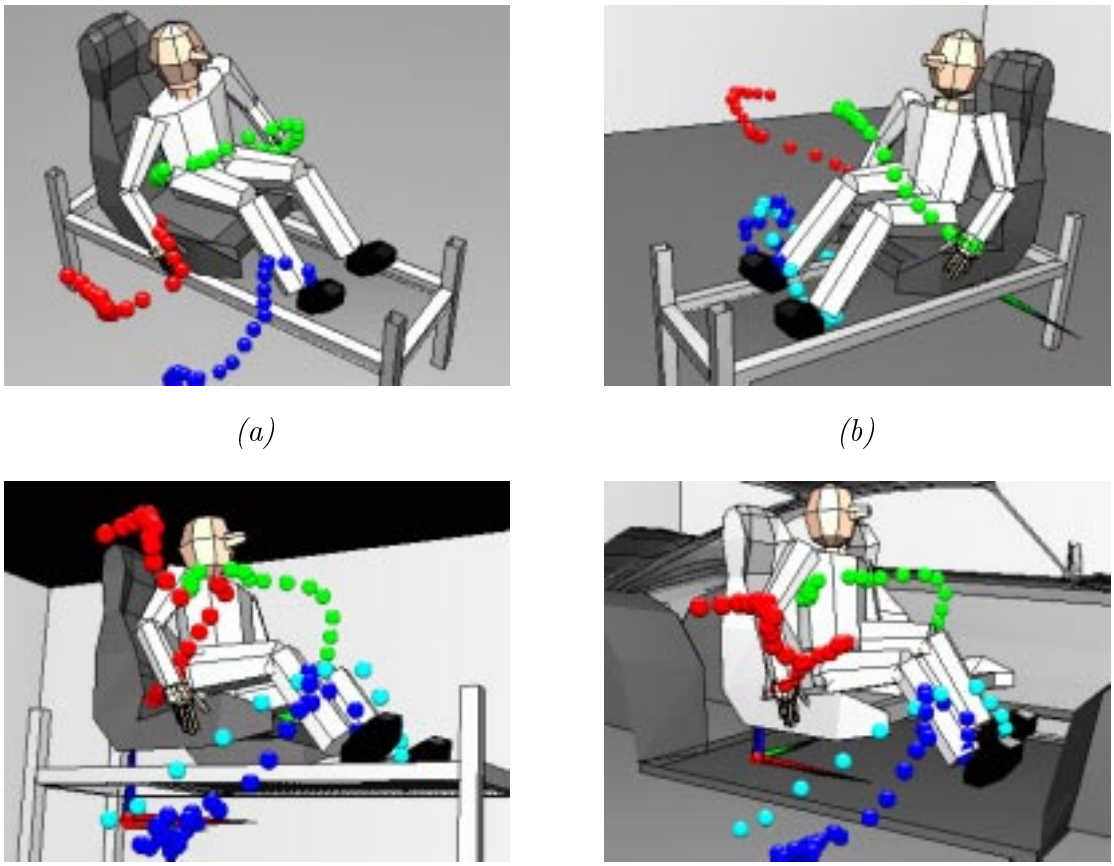


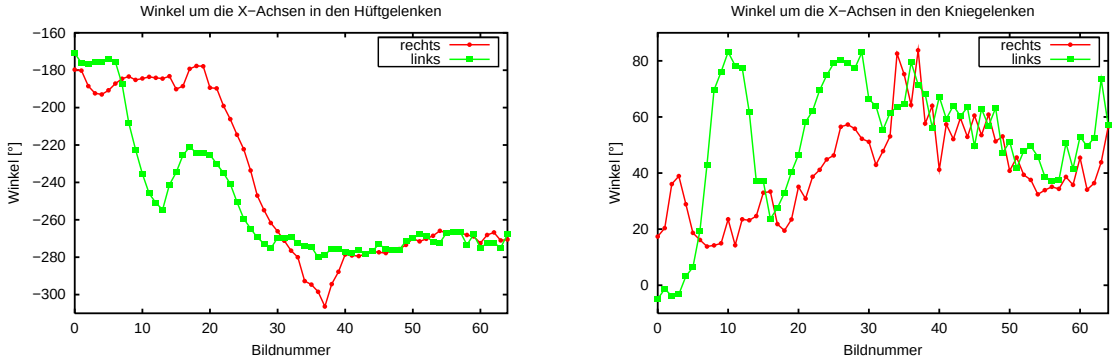
Abbildung 4.26: *Projektion des erfaßten Bewegungsablaufes in eine virtuelle CAD Umgebung; dargestellt jeweils aus drei beliebigen Ansichten.*



(a) (b)

(c) (d)

Abbildung 4.27: *Trajektorien von Händen und Füßen beim Einstiegsvorgang; Positionspunkte für die letzten 20 Bilder bis zum Bild mit Nummer 45.*



(a) (b)

Abbildung 4.28: *Beobachtung eines Einstiegsvorgangs: Gelenkwinkel um die X-Achsen in den (a) Hüft- und (b) Kniegelenken (b) der detektierten Modellinstanz.*

4.3.2 Anwendung in der Tiermedizin

In diesem Abschnitt soll die Modellierung eines weiteren artikularen Objektes vorgestellt werden. Hierzu ist ein Beispiel aus der Tiermedizin gewählt worden, wobei ein Kontakt zur Klinik für Rinderkrankheiten der Tierärztlichen Hochschule, Hannover grundlegend war. Dort sind Entzündungserkrankungen an den Hinterhufen von Rindern untersucht worden, die vermehrt bei der Haltung von Milchvieh in Großställen, bei der die Rinder auf Betonplanken stehen, auftreten. Werden die Erkrankungen frühzeitig erkannt, so können die Beschwerden erfolgreich behandelt werden. Da sich die Rinder in großen Milchbetrieben selbständig in dem Bereich der Außenflächen und Stallungen bewegen können, ist eine ausreichende Beobachtung der Tiere durch den Menschen nicht mehr gegeben.

Die Erkrankung kann jedoch durch die Beobachtung von abnormalen Bewegungsabläufen diagnostiziert werden, da die erkrankten Tiere lahmen. Das Ziel ist daher eine Methode zu finden, mit der die Bewegungsabläufe der Tiere automatisch erfaßt werden. Hierzu könnten einzelne Rinder auf dem Weg vom Stall zum Freigelände mit Videokameras aufgenommen, detektiert und deren Bewegung erfaßt werden. Aufnahmen von einem lahmenden Rind sind in der Abb. 4.29 gezeigt.[12] Für erste Tests sind zunächst einige ausgewählte Gelenkpunkte mit Landmarken markiert worden, die im Bild als Bildmerkmal zu extrahieren sind. Vgl. hierzu auch die Beschreibung des entsprechenden Bildmerkmals in STABIL^{++} , Kap. 3.4.3. Kann sichergestellt werden, daß die Rinder aufgrund der Umgebung in einem bestimmten Abstand vor der Kamera vorbeilaufen, kann diese Information für einen 2D/3D Übergang von Bild zu Szenenmerkmalen als Tiefenschätzung in einem monokularen Ansatz verwendet werden.

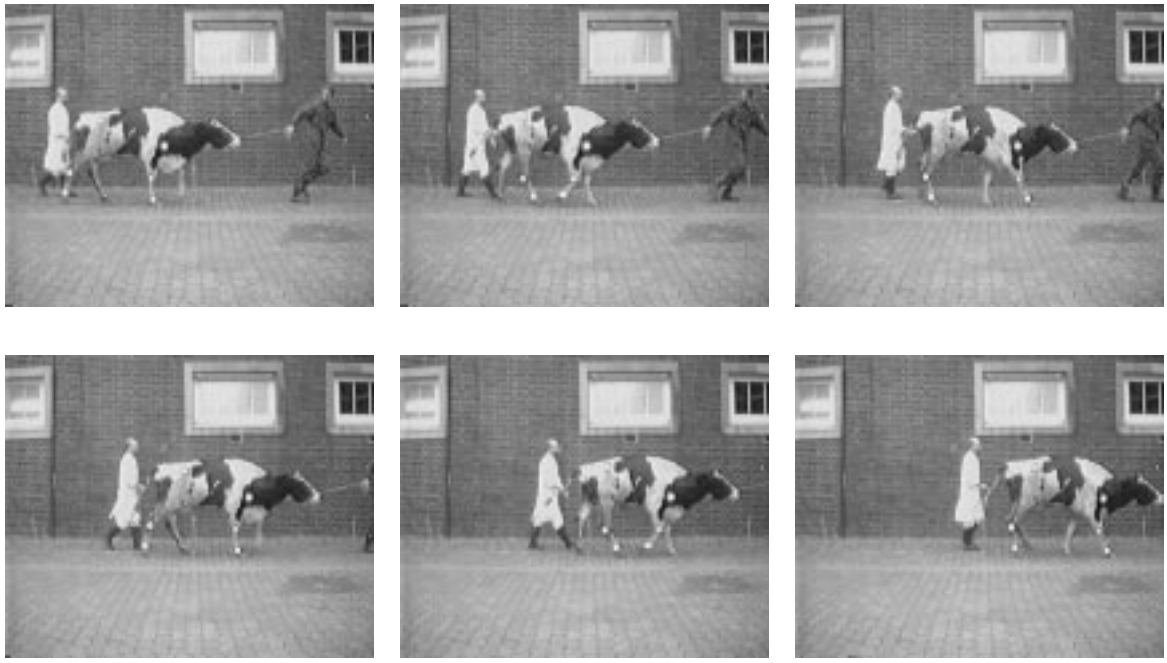


Abbildung 4.29: *Bildsequenz zur Beobachtung der Bewegung von lahmenden Rindern; Markierung der Gelenke durch Landmarken.*

[12] Die Aufnahmen sind aus einem Kontakt zur Tierärztlichen Hochschule, in Zusammenarbeit mit Herrn Dr. Moritz Metzner entstanden.

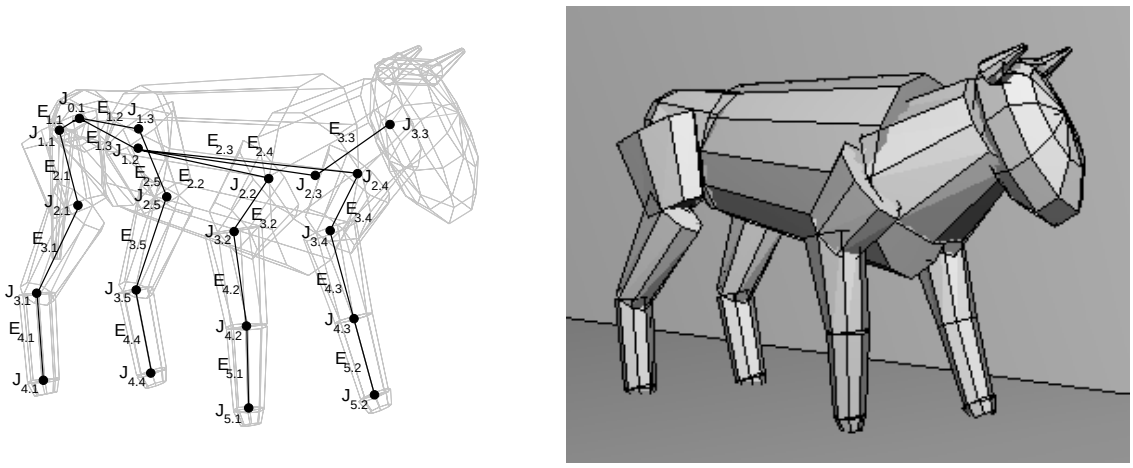


Abbildung 4.30: *Objektmodell für Rinder: (a) innere Objektmodellstruktur und (b) Objektmodellstruktur mit Volumenkörpern.*

Basierend auf der Anatomie der Rinder läßt sich ein Objektmodell erstellen. Abb. 4.30 zeigt eine beispielhafte Modellierung: (a) innere, hierarchische Objektmodellstruktur und (b) Darstellung der äußeren Modellstruktur anhand der Projektion der Volumenkörper. Für die Gliedmaßen sind hier jeweils ein Objektmodellteil mehr definiert worden als bei der vorgestellten Standardmodellierung des menschlichen Körpers. Dies begründet sich darin, daß bei einer Gegenüberstellung der Skelettstrukturen von Rindern und Menschen die Hufe der Rinder mit den Fingern und Zehen des menschlichen Körpers zu vergleichen sind. Die Hufe sind in der gewählten Modellierung jedoch nicht weiter in dreigliedrige Zehen eingeteilt worden, sondern als ein Objektmodellteil modelliert worden. Die Bezeichnungen der einzelnen Objektmodellteile und deren Numerierung sind für die dargestellte Modellierung in der Tab. 4.8 aufgelistet.

Hüfte	0.1	Rumpf	1.2
Hals	2.3	Kopf	3.3
rechter Oberschenkel	1.1	linker Oberschenkel	1.3
rechter Unterschenkel	2.1	linker Unterschenkel	2.5
rechter Hintermittelfuß	3.1	linker Hintermittelfuß	3.5
rechter Hinterhuf	4.1	linker Hinterhuf	4.4
rechter Oberarm	2.2	linker Oberarm	2.4
rechter Unterarm	3.2	linker Unterarm	3.4
rechter Vordermittelfuß	4.2	linker Vordermittelfuß	4.3
rechter Vorderhuf	5.1	linker Vorderhuf	5.2

Tabelle 4.8: *Numerierung der Objektmodellteile entsprechend Abb. 4.30.*

5 Schlußbemerkungen

5.1 Zusammenfassung

Mit dem vorgestellten Gesamtkonzept liegt die Darstellung eines modellbasierten Bildinterpretationssystems zur Beobachtung artikularer Bewegung vor. Eine entsprechende Umsetzung realisiert das System STABIL^{++}. Das Konzept umfaßt hierzu eine Modellierung der zu detektierenden Objekte, eine hinreichende Beschreibung der Umgebung und eine Modellierung der Abbildungseigenschaften der Kameras. Hierzu sind ein Szenen-, ein Kamera- und ein Objektmodell definiert worden, wobei eine durchgängige 3D Modellierung verwendet wird. Darüber hinaus umfaßt das Konzept den eigentlichen Interpretationsprozeß, der sich auf die Gesamtmodellierung stützt, jedoch direkt durch das Objektmodell gesteuert und initiiert wird.

Zusammenfassend sollen im folgenden das Objektmodell, der Interpretationsprozeß und zwei Anwendungsbereiche kurz beschrieben werden.

Objektmodell

Zur Beschreibung der zu beobachtenden artikularen Objekte ist eine dreischichtige Modellierung gewählt worden. Diese setzt sich aus folgenden Strukturen zusammen:

Innere Objektmodellstruktur zur Beschreibung der hierarchischen Ordnung der Objektmodellteile, aus der sich das zu beschreibende artikulare Objekt zusammensetzt

Geometrische Objektmodellstruktur zur Beschreibung der Lage der einzelnen Objektmodellteile zueinander durch Angabe von Translation und Rotation zwischen lokalen 3D Koordinatensystemen

Äußere Objektmodellstruktur zur Beschreibung der 3D Ausdehnung der Objektmodellteile und deren charakteristischen Merkmale

Die vorgestellte Modellierung ist nicht starr an einen Objekttyp gebunden, sondern kann, je nach zu detektierendem Objekt und Anwendung flexibel konfiguriert werden. Dies wird durch die Dreiteilung der Modellstruktur unterstützt: Mit der inneren Struktur wird zunächst festgelegt, welche Anzahl von Objektmodellteilen zu erfassen sind. Dies hängt zum einen von dem Objekt, zum anderen auch von der Anwendung ab. Es werden hierzu nur die Objektteile modelliert, zu denen Informationen in den Videobildern extrahiert werden können. In einem weiteren Schritt wird darüber hinaus mit der geometrischen Struktur festgelegt, wie die Lage der einzelnen Objektmodellteile und der dazwischen liegenden Knoten / Gelenke zueinander ist. Obwohl bei der Definition artikularer Objekte von starren Objektmodellteilen ausgegangen wird, ist bei der Festlegung der geometrischen Struktur zunächst nur die mittlere Größe der zu erwartenden Objekte festzulegen. Dies gilt auch für den dritten Schritt der Modellierung, bei der in der äußeren Objektmodellstruktur die Ausdehnung der Objektmodellteile anhand von Volumenkörpern

zu beschreiben ist. Schließlich sind noch die Merkmale zu definieren, mit denen die Erscheinung eines Objektes beschrieben wird.

Die Merkmale, mit denen die Erscheinung der Objektmodelle charakterisiert werden, werden als die Modellmerkmale bezeichnet und den Objektmodellteilen zugeordnet. Zu jedem Modellmerkmal wird angegeben, wie es aus den Bilddaten zu extrahieren ist. Hierzu werden den Modellmerkmalen Bildverarbeitungsoperatoren zugeordnet. Man spricht daher auch von zu extrahierenden Bildmerkmalen. Je nach Objekt, Anwendung, Kamera oder Umgebungsbedingung können die Attribute der Merkmale und die zuzuordnenden Bildverarbeitungsoperatoren flexibel gewählt und konfiguriert werden.

Interpretation

Basierend auf der Dreiteilung des Objektmodells werden die verschiedenen Aspekte der Modellierung auch beim Prozeß der Interpretation genutzt. In einem ersten Schritt werden zunächst plausible Hypothesen unter Berücksichtigung der hierarchischen, inneren Struktur und der Translationsanteile der geometrischen Struktur erzeugt. Erst in einem weiteren Schritt der Hypothesenbewertung und -auswahl wird die Konfiguration der Objektmodelle[1] anhand der Rotationsanteile der geometrischen Struktur und der Volumenkörper der äußeren Modellstruktur berücksichtigt. Hierdurch wird zum einen erreicht, daß über die Merkmale der einzelnen Objektmodellteile nur die 3D Position der Ursprünge der lokalen Koordinatensysteme bestimmt werden müssen. Damit können auch einfache Punktmerkmale Verwendung finden. Zum anderen müssen bei der Korrespondenzfindung der Merkmale nur die algorithmisch einfacheren Restriktionen aus der inneren Modellstruktur und aus den Translationen zwischen den Objektmodellteilen berücksichtigt werden. Die Rotationen in den Gelenken der Objektmodellteile und somit die Haltung und Konfiguration des Objektmodells wird daher nicht für jede mögliche Korrespondenz bestimmt, sondern nur für die plausiblen Hypothesen. Das gleiche gilt für die Berücksichtigung der Restriktionen, die sich aus der äußeren Objektmodellstruktur ergeben.

Für die eigentliche Interpretation sind Modellmerkmale mit Merkmalen zu vergleichen, die für die zu beobachtende Szene bestimmt werden können. Man unterscheidet für die Korrespondenzfindung zwischen Ansätzen, bei denen Struktur-Vergleichsverfahren Merkmale im 2D Raum des Bildes oder im 3D Raum der Szene vergleichen. Zum Vergleich von 2D Merkmalen muß das 3D Modell in die Bildebene projiziert werden, so daß 2D Abbilder der Modellmerkmale zur Verfügung stehen. Hierzu ist jeweils die Haltung und Konfiguration des artikularen Objektmodells vorauszusetzen. Zumindest für eine initiale Detektion ist die Haltung des zu detektierenden Objektes als bekannt vorzugeben. Für jeden weiteren Interpretationsschritt kann eine Bewegungsvorhersage verwendet werden. Viele Ansätze stützen sich jedoch zusätzlich auf Bewegungsmodelle, mit denen ein Bewegungsmuster für die zu detektierenden Objekte vorausgesetzt wird.

In STABIL^{++} wird der Struktur-Vergleich im 3D Raum durchgeführt, bei dem neben den 3D Modellmerkmalen und 2D Bildmerkmalen noch sog. 3D Szenenmerkmale eingeführt worden sind. Szenenmerkmale werden aus Bildmerkmalen erzeugt, die in einem oder mehreren Kamerabildern extrahiert werden. Die 3D Position eines Szenenmerkmals wird dabei entweder über einen (Mehrfach-) Stereoansatz ermittelt oder basierend auf Modellwissen in einem monokularen Ansatz geschätzt. Grundlegend für die Bestimmung der 3D Positionen ist dabei das Kameramodell, das bedeutet, daß jede im System zu verwendende Kamera vor der Verwendung zu kalibrieren ist.

[1] Bei Personenmodellen kann von der Haltung des Modells gesprochen werden.

Für den Struktur-Vergleich der Merkmale zur Generierung der Hypothesen wird eine erschöpfende Suche in einem Interpretationsbaum durchgeführt. Die Knoten des Baumes setzen sich aus Zuordnungen von 3D Modellmerkmalen zu 3D Szenenmerkmalen zusammen. Die Struktur des Interpretationsbaums ist aus der inneren Objektmodellstruktur heraus gegeben. Um die Suchraumgröße für die Korrespondenzsuche zu beschränken, werden schon beim Aufbau des Baumes Restriktionen berücksichtigt. Dies sind Restriktionen auf der Ebene der Merkmale und solche, die sich aus der inneren und geometrischen Modellstruktur ergeben.

Eine wichtige Restriktion auf der Ebene der Merkmale ergibt sich aus 3D Suchräumen, die bei der Verfolgung von Objekten für einzelne Objektmodellteile bestimmt werden. Die Suchräume werden anhand von 3D Positionsvorhersagen für die einzelnen Objektmodellteile generiert. Hierzu ist jedem Objektmodellteil mit der geometrischen Objektmodellstruktur ein Vorhersagefilter zugeordnet. Aus den vorhergesagten Positionen können zudem auch Positionsschätzungen generiert werden. Mit diesen werden Merkmalspositionen geschätzt, so daß der Struktur-Vergleich auch dann nicht fehlschlägt, wenn einzelne Merkmale zeitweilig verdeckt sind.

Dem Problem der Verdeckung wird in STABIL^{++} zusätzlich dadurch begegnet, daß die Anzahl der zu verwendenden Kameras nicht beschränkt ist. Ein zu beobachtendes Objekt kann also von mehreren Ansichten betrachtet werden. Damit wird sichergestellt, daß Objektmerkmale aus mindestens zwei Ansichten sichtbar sind, falls anwendungsbedingt für den 2D/3D Übergang von Bildmerkmalen zu Szenenmerkmalen der Stereoansatz verwendet werden muß. Bei der Verwendung von mehreren Kameras müssen sich die Sichtbereiche der Kameras nicht zwangsläufig überlappen; der zu observierende Bereich kann ebenso durch weitere Kameras erweitert werden. Zu verfolgende Objekte werden hierbei implizit aufgrund der Bestimmung der 3D Suchräume über mehrere Kameras verfolgt. Für die Interpretation werden immer nur von den Kameras Bilder eingezogen, in denen aufgrund der 3D Suchbereiche Objekte zu erwarten sind. Man kann daher auch von einer “Übergabe” der zu verfolgenden Objekte sprechen. Auf der Grundlage der 3D Information der Suchbereiche kann darüber hinaus bei der Verwendung von aktiven Kamerasystemen, wie z.B. Schwenk- / Neigekameras die Sichtbereiche der Kameras optimal auf die Suchbereiche automatisch positioniert werden.

Anwendungen

Die Beobachtung artikularer Objekte wird meist mit der Beobachtung des menschlichen Körpers gleichgesetzt. Dem entsprechend ist die Anwendung von STABIL^{++} auf Personen gezeigt worden. Aufgrund der Flexibilität des Konzeptes und der Objektmodellierung konnten zwei Anwendungen mit unterschiedlichem Ziel gewählt werden: Zum einen eine Personendetektion und -verfolgung für Anwendungen in der Videoüberwachung für die Sicherheitstechnik. Zum anderen die Erfassung der Bewegung einer Person für eine Auswertung und Analyse unter ergonomischen Gesichtspunkten.

Für die erste Anwendung in der Sicherheitstechnik sollen Personen detektiert und lokalisiert werden, wobei jedoch die Haltung der Personen nicht weiter von Interesse ist. Es sind von dem System lediglich die 3D Positionen der Personen und bei Verfolgung die 3D Trajektorien der Personenbewegung im Raum zu bestimmen. Entsprechend ist ein Objektmodell zu konfigurieren, bei dem ein Modellmerkmal zur Lokalisation ausreicht. Für die Überwachung in Innenräumen wird hierzu als Merkmal die “hautfarbene” Ellipse des Gesichts von bekleideten Personen genutzt. Zur Extraktion zugehöriger Bildmerkmale wird eine adaptive Farbklassifikation verwendet, um “hautfarbene” Bildregionen zu segmentieren. Zusätzlich werden Formmerkmale zur Auswahl von Ellipsen angewendet. Neben dem Objektmodellteil des Kopf-

Anwendungs-beispiel	Anzahl primärer Merkmale	Kamera-anzahl	Bildgröße	SUN [ms]	PC [ms]
Personendetektion und -verfolgung	1	1	384×288	450	250
Bewegungserfassung	14	3	384×288	4.000	1.500

Tabelle 5.1: *Mittlere zeitliche Länge eines Interpretationszyklus; Angaben für SUN Sparc ULTRA / 170 MHz (Solaris) und PC Intel Pentium II / 400 MHz (Windows NT).*

es werden auch Modellteile für den Rumpf und die Beine definiert. Diese Modellteile werden nicht für die Lokalisation verwendet, sondern zur Verifikation von Hypothesen, die aufgrund des Merkmals des Kopfes im Interpretationsprozeß aufgestellt wurden. Zur Verifikation werden Heuristiken angewandt, die als sekundäre Merkmale bei der Modellierung für jedes Objektmodellteil definiert werden können. Im Gegensatz hierzu werden die Modellmerkmale, die zur 3D Lokalisation verwendet werden, als primäre Merkmale bezeichnet.

Für den Übergang von den 2D Bildmerkmalen zu 3D Modellmerkmalen kann mit einem monokularen Ansatz zur Tiefenschätzung eine für diese Anwendung hinreichende Genauigkeit erreicht werden. Wird ein zu überwachender Bereich von mehreren Kameras abgedeckt, wobei sich diese an den Rändern der Sichtbereiche überlappen, so wird automatisch, falls eine Person in zwei Kameras sichtbar ist, die 3D Position für den Kopf mit einem Stereoansatz ermittelt.

Für die zweite Anwendung ist ein Beispiel gezeigt worden, in dem der Einstiegsvorgang in einen PKW beobachtet wird. Für eine weitere Analyse der Bewegung sind, neben den 3D Positionen der Gelenke, die Winkel der Gelenke zu erfassen. Dementsprechend ist für die Interpretation ein wesentlich detaillierteres Objektmodell verwendet worden als bei der ersten Anwendung. Es kann für diese Anwendung unter Laborbedingungen gearbeitet werden, was das Anbringen von künstlichen Markierungen der Gelenke erlaubt. Es sind hierzu farbige Bänder in vier verschiedenen Farben gewählt worden. Die Bildregionen der zugehörigen Bildmerkmale werden auch hier über einen Farbklassifikator extrahiert. Aufgrund der Verwendung von primären Merkmalen für jedes Objektmodellteil und der Mehrfachverwendung jeder Markierungsfarbe steigt bei dieser Anwendung der Suchraum für die Korrespondenzsuche gegenüber dem ersten Beispiel erheblich an.

Zur Erfassung des exakten Bewegungsablaufes werden auch andere Anforderungen an die Genauigkeit gestellt. Es wird daher für den 2D/3D Übergang nur der Stereoansatz zugelassen. In dem vorgestellten Beispiel sind Aufnahmen von drei Kameras verwendet worden, wobei die Anordnung so gewählt war, daß nach Möglichkeit zu jedem Zeitpunkt die Markierungen in mindestens zwei Kameras sichtbar sind. Die Verwendung mehrerer Kameras, in denen Objekte aufgrund der 3D Suchbereiche erwartet werden, beeinflussen neben dem Detaillierungsgrad des Objektmodells die Zykluszeit für einen Interpretationsschritt. Dies zeigt sich auch in der Gegenüberstellung der durchschnittlichen Verarbeitungszeit für einen Interpretationszyklus der beiden vorgestellten Anwendungen in Tab. 5.1. Für die erste Anwendung wurde die Verarbeitungszeit für das erste Beispiel aus dem Kap. 4.2 ermittelt, wobei auch hier auf zuvor aufgezeichneten Bildern gearbeitet wurde. Die Zeiten für die zweite Anwendung entsprechen der Verarbeitung der in Kap. 4.3.1 vorgestellten Bewegungserfassung.

Für Anwendungen in der Sicherheitstechnik ist die Echtzeitfähigkeit des Ansatzes von Bedeutung, wenn, basierend auf der 3D Lokalisation der detektierten Personen, Alarme ausgelöst oder Aufzeichnungen gestartet werden müssen. Die Anzahl der hierzu pro Sekunde zwingend zu verarbeitenden Bilder hängt von der zu erwartenden Objektgeschwindigkeit ab. Bei der Über-

wachung in Innenräumen können ab einer Verarbeitungsgeschwindigkeit von drei Interpretationszyklen pro Sekunde realistisch verwendbare Detektionsergebnisse erzielt werden. Dies kann mit dem vorgestellten Konzept mit Standard-PC-Komponenten erreicht werden.

Die Anforderungen der Anwendungen zur Bewegungserfassung sind anders gelagert: Bewegungsabläufe können zunächst aufgezeichnet und anschließend in einer *off-line* Verarbeitung interpretiert werden. Darüber hinaus muß das System hierzu nicht vollautomatisch arbeiten, denn der Benutzer kann in den Interpretationsprozeß eingreifen. Dies ist z.B. erforderlich, wenn es aufgrund der Personenbewegung zu Verdeckungen kommt und einzelne Marken nicht zu extrahieren sind. Wird aufgrund einer geforderten hohen Genauigkeit an die 3D Positionen und Gelenkwinkel eine Schätzung der Positionen verdeckter Objektmodellteile nicht zugelassen, so kann vom Benutzer die Position in den aufgenommenen Bildern manuell ausgewählt werden, vgl. hierzu das entsprechende Bildmerkmal in Kap. 3.4.3.

5.2 Ausblick

Das vorgestellte Konzept ermöglicht aufgrund der Flexibilität des Objektmodells und der Strukturierung des Interpretationsprozeßes, Weiterentwicklungen in Teilen der Interpretation und Anpassung des Prozeßes an verschiedenste Anwendungen. In den folgenden Abschnitten wird daher ein kurzer Ausblick auf mögliche und wünschenswerte Weiterentwicklungen gegeben.

Merkmale

Die Sicherheit und Güte der Detektion ist entscheidend davon abhängig, wie robust die den Modellmerkmalen zugehörigen Bildmerkmale zu extrahieren sind. In den vorgestellten Anwendungen wurden Farbmerkmale verwendet, wodurch eine ausreichende Beleuchtung der zu beobachtenden Szene vorausgesetzt wird. Für die Anwendung im Bereich der Sicherheitstechnik ist mit dem Merkmal der “hautfarbenen” Ellipse der Einsatz auf Überwachungen im Innenraum begrenzt. Für den Außenbereich muß zum einen mit stark schwankenden Beleuchtungsänderungen und damit auch Farbverschiebungen gerechnet werden. Zum anderen sind größere Bereiche zu observieren, so daß mit zunehmender Entfernung von zu detektierenden Objekten zur Kamera das Segmentierungsproblem steigt.

Eine Möglichkeit der Erweiterung ist die Verwendung des Umrißes / der Silhouette von zu detektierenden Personen als deren Merkmal. Hierzu kann der Mittelpunkt des Kopfes aus dem Umriß von Kopf und Schulteransatz bestimmt und als primäres Merkmal für das Objektmodellteil des Kopfes verwendet werden, vgl. [HE95] und [MK96]. Für Anwendungen im Innenbereich kann der Umriß als zusätzliches Merkmal sinnvoll sein, denn das “hautfarbene” Gesicht einer Person kann von der Kamera nicht erfaßt werden, wenn diese der Kamera den Rücken zukehrt. Es ist daher als Weiterentwicklung eine Kombination von primären Merkmalen denkbar: Zur initialen Detektion wird weiterhin nach “hautfarbenen” Bereichen als Bildmerkmal gesucht. Hiermit wird weiterhin angenommen, daß Personen, die den zu überwachenden Bereich betreten, von vorne aufgenommen werden. Für jede weitere Detektion wird zusätzlich im Bild nach (Differenzbild-) Kanten gesucht, die durch die Silhouette der zu verfolgenden Person hervorgerufen werden. Ist das Farbmerkmal nicht mehr zu detektieren, so kann die Position weiterhin über das zweite Merkmal bestimmt werden. Die Kombination von “hautfarbenem” Merkmal und Umriß wird auch in [IB98] vorgeschlagen. Dort wird zunächst anhand der Hautfarbe eine grobe Lokalisation von Händen im Bild vorgenommen und anschließend über einen Condensation-Algorithmus die Lokalisation verfeinert und eine Verfolgung realisiert. Eine An-

wendung des Condensation-Algorithmus zur Detektion von Oberkörperhaltungen ist in [OG99] vorgestellt.[2]

Ebenso wie eine Nutzung von mehreren primären Merkmalen für ein Objektmodellteil ist es denkbar, daß zusätzliche primäre oder sekundäre Merkmale während der Verfolgung für eine Objektmodellinstanz definiert werden. Hierüber kann ein weiteres Merkmal zur eindeutigen Charakterisierung eines Objektes geschaffen werden. Für die Anwendung zur Personendetektion kann nach der initialen Detektion einer Person anhand des primären Merkmals des Objektmodellteils des Kopfes z.B. noch ein weiteres Merkmal für das Objektmodellteil des Rumpfes definiert werden. Hierzu eignet sich ein Farb- oder ein Texturmerkmal, mit dem die Erscheinung des Rumpfes charakterisiert wird. Ein entsprechender Klassifikator kann direkt aus der Bildregion angelernt werden, in die das entsprechende Objektmodellteil der detektierten und lokalisierten Objektmodellinstanz projiziert wird.

Werden in das System neue Merkmale integriert, so sind die entsprechenden Restriktionen und Gütefunktionen, die bei der Generierung und Bewertung der Hypothesen verwendet werden, neu zu definieren oder anzupassen. Darüber hinaus kann für eine Optimierung des Interpretationsprozeßes für konkrete Anwendungsfälle eine Abänderung der vorgeschlagenen Bewertungsfunktionen notwendig werden.

Positionsvorhersage

Ein weiteres Gebiet für Weiterentwicklungen bieten die Vorhersagefilter $pred$, die den einzelnen Objektmodellteilen zugeordnet sind, um bei der Re-Detektion 3D Positionen vorherzusagen. Diese Positionen werden zum einen für die Bestimmung der 3D Suchräume verwendet und bestimmen zum anderen 3D Positionsschätzungen von Szenenmerkmalen, falls Merkmale nicht extrahiert werden können. Die vorgestellte Positionsvorhersage über einfache Extrapolation kann durch den Einsatz von Kalmanfiltern erweitert werden. Hierbei kann die "Beweglichkeit" eines Objektmodellteiles genauer berücksichtigt werden. Anhand einer von dem Filter verwendeten Modellvorstellung zur 3D Bewegung eines Objektmodellteils wird zunächst eine Position vorhergesagt, die zur Bestimmung der 3D Suchräume verwendet werden kann. Die aufgrund des Aufbaus des Kalmanfilters berücksichtigten Unsicherheiten können darüber hinaus genutzt werden, um die Form der Suchräume zu bestimmen. So muß sich der Suchraum in eine Richtung mit großer Unsicherheit weiter ausdehnen als in eine Richtung, in der eine gesicherte Vorhersage getroffen werden konnte.

Der Kalmanfilter ermöglicht weiterhin, aufgrund der Vorhersage und einer Meßgröße[3], eine gesicherte Schätzung zu erstellen, wobei die Unsicherheiten in den Meßgrößen und in den Vorhersagen berücksichtigt werden. Für die Anwendung in STABIL^{++} bedeutet dies, daß bei der Zuordnung von 3D Szenenmerkmalen zu dem 3D Modellmerkmal des primären Merkmals eines Objektmodellteils nicht die 3D Position des Szenenmerkmals verwendet wird. Statt dessen kann eine durch den Filter geschätzte 3D Position verwendet werden, die mit großer Wahrscheinlichkeit näher an der tatsächlichen Merkmalsposition liegt als die aus den Bilddaten bestimmte Position. Hierdurch werden Fehler bei der Segmentierung der Bildmerkmale und bei dem 2D/3D Übergang zu den Szenenmerkmalen ausgeglichen. Ein Ansatz und erste Tests zur Integration von Kalmanfiltern zur 3D Positionsvorhersage und -schätzung in STABIL^{++} ist in

[2] Für den Condensation-Algorithmus wird ein mehrdimensionaler Zustandsraum aufgespannt; mit den einzelnen Dimensionen sind Parameter der Form und der Lage von möglichen Objektsilhouetten zu beschreiben; die Verfolgung wird über die Verteilung / Population im Zustandsraum vorgenommen; eine Erweiterung der Ansätze auf 3D Positionen ist denkbar.

[3] 3D Position der aus den Bildmerkmalen ermittelten Szenenmerkmale.

[Rau99] beschrieben, vgl. auch die Verwendung von Kalmanfiltern in [HOW96] und [JW97].

Weiterhin ist es möglich, über die 3D Positionsvorhersage hinaus für die einzelnen Objektmodellteile einen Vorhersagefilter zu definieren, mit dem Gelenkwinkelverläufe vorhergesagt werden. Dies ist insbesondere dann sinnvoll, wenn aufgrund der zu erwartenden Objektbewegung mit den Gelenkwinkeln sich kontinuierlich ändernde Größen beschrieben werden können. Entsprechende 3D Positionsvorhersagen für die Objektmodellteile ergeben sich dann aus den Winkelvorhersagen und den Modellstrukturen. Man erreicht hierbei, daß diese innerhalb der Restriktionen auf der Modellebene bleiben, die bei der Generierung der Hypothesen verwendet werden.

Zweistufige Interpretation

Die Notwendigkeit der Erweiterung des Interpretationsprozeßes in eine Stufe zur Bestimmung einer groben Lagehypothese und eine Stufe der Verifikation und Feinerkennung soll an einer Anwendung zur Bewegungserfassung verdeutlicht werden. Die Erkennung und Deutung der Bewegung von Personen aus Videobildfolgen kann auf Bewegungsdaten aufsetzen, die mit STABIL^{++} erfaßt werden. In [CC98] ist ein Ansatz zur Erkennung von Bewegungsmustern gezeigt, der sich auf 3D Positionsdaten von Objektmodellteilen abstützt.

Das Ziel dieser Anwendung ist es, die non-verbale Kommunikation in Gesprächen zu erfassen und zu klassifizieren, vgl. auch [Nak98]. Unter der Annahme, daß viele Botschaften und Signale in Gesprächen zwischen Gesprächspartnern auf dieser Ebene unbewußt ausgetauscht werden, gilt es, das non-verbale Gesprächsverhalten zu deuten und zu nutzen. Hierzu werden z.B. Untersuchungen zur Arzt-Patient-Kommunikation an der Universität Regensburg durchgeführt.[4] Für die Analyse werden Gespräche mit Videokameras aufgezeichnet und nach Bewegungsmustern durchsucht. Sind die Muster mit der dazugehörigen Deutung bekannt, so kann in einem weiteren Schritt ein Kommunikations-Training für Ärzte entwickelt werden.

Für die Analyse und für ein Trainingssystem sind die Bewegungen in videobasierten Systemen automatisch zu erfassen und zu bewerten. Eine Erfassung der Bewegung anhand von Gelenkmarkierungen ist in dem Umfeld jedoch nicht realisierbar. Geht man davon aus, daß bei der Beobachtung von solchen Gesprächen die Bewegung von Kopf, Oberkörper und Armen von sitzenden Personen erfaßt werden soll, so können mit STABIL^{++} zunächst anhand des "hautfarbenen" Merkmals des Gesichts und der Hände die Objektmodellteile für den Kopf und die Hände lokalisiert werden. Betrachtet man diese Positionen als Groblokalisation des Objektes, so kann STABIL^{++} dahin gehend weiterentwickelt werden, daß in einem zweiten Schritt eine Feinlokalisation vorgenommen wird. Hierzu dient die Groblokalisation als Lagehypothese, mit der ein Teil des artikularen Objektmodells fixiert ist. Anschließend können verschiedene Konfigurationen für das Objektmodell angenommen werden, wobei die Restriktionen berücksichtigt werden, die sich aus den Modellstrukturen ergeben. Projiziert man das Modell entsprechend der verschiedenen Konfigurationen in die Bildebene, so kann in einem zweiten Schritt der Feinlokalisation ein 2D/2D Vergleich von projizierten Modellmerkmalen mit Bildmerkmalen vorgenommen werden, um die Lage der weiteren Objektmodellteile zu bestimmen.

Große Überwachungsbereiche

Für Anwendungen zur Personendetektion und -verfolgung in der Sicherheitstechnik ist es wünschenswert, das System mit einer beliebigen Anzahl von Kameras auszustatten, um damit den zu observierenden Bereich beliebig zu erweitern. Generell ist dies mit dem vorgestell-

[4] Einheit Medizinische Soziologie an der Universität Regensburg, Prof. Dr D. v. Schmädel.

ten Konzept möglich, jedoch werden bei der Realisierung mit einem Rechnersystem aufgrund der Echtzeitanforderung Grenzen erreicht. Dies ist zum einen durch Begrenzung der maximal anzusteuernden Digitalisierungskarten (framegrabber) gegeben und zum anderen durch die notwendige Rechenleistung zur Anwendung der Bildverarbeitungsoperatoren.

Das vorgestellte Konzept des Interpretationsprozeßes kann jedoch auf mehrere Rechner verteilt angewendet werden. Hierzu wird in einem zentralen System das Szenenmodell und somit auch die Objektmodellinstanzen verwaltet. Für die Digitalisierung der Kamerasignale werden dezentrale Kamerarechner in den zu überwachenden Bereichen aufgestellt (z.B. für jeden Raum) an die eine oder mehrere Kameras angeschlossen werden. Entsprechend des beschriebenen Interpretationsprozeßes wird vom zentralen System für jede zu detektierende Objektmodellinstanz ein 3D Suchraum vorhergesagt. Diese Informationen werden mit der Beschreibung der Bildverarbeitungsoperatoren für die Merkmalsextraktion an die dezentralen Kamerarechner übermittelt. Dort wird entschieden, ob der 3D Suchbereich im Sichtbereich der Kamera liegt; wenn ja, werden Bilder von den entsprechenden Kameras eingezogen, auf diesen die Vorverarbeitung durchgeführt und die Bildmerkmale extrahiert. Für den weiteren Interpretationsprozeß werden die Informationen der extrahierten Bildmerkmale dem zentralen System zur Verfügung gestellt.

Durch eine Kaskadierung von verteilten Systemen ist die Observierung in Gebäuden über mehrere Etagen realisierbar. Für das Gesamtsystem ist in einem zentralen Szenenmodell ein Weltkoordinatensystem für das Gebäude festzulegen, auf das sich jedes einzelne System bezieht. Es muß nur für jede Etage zusätzlich eine xy-Ebene definiert werden, mit der die Höhe des Etagenboden angegeben wird. Diese Ebene wird benötigt, um für die Tiefenschätzung des monokularen Ansatzes eine Bezugsebene zu haben.

Fazit

Mit der vorgestellten Modellierung von artikularen Objekten und dem Konzept des zugehörigen Interpretationsprozeßes sind Anwendungen von STABIL^{++} in zwei unterschiedlichen Bereichen zur Beobachtung von Bewegungen des menschlichen Körpers gezeigt worden. Erst durch die freie Konfigurierbarkeit des generischen Objektmodells sind hierbei die vielfältigen Einsatzmöglichkeiten gegeben. Darüber hinaus wird dies noch durch die Flexibilität des Interpretationsprozeßes in Bezug auf die quasi unbeschränkte Anzahl von Kameras und der damit zusammenhängenden Variabilität in den Ansätzen zur Bestimmung der 3D Positionen der Szenenmerkmale unterstützt. STABIL^{++} kann daher als Grundsystem für viele weitere Anwendungen zur Beobachtung artikularer Objekte verwendet werden, wobei die denkbaren Erweiterungen über die aufgezeigten Weiterentwicklungen hinausgehen.

Anhang

A Rotationen zwischen Objektmodellteilen

A.1 Grundlagen

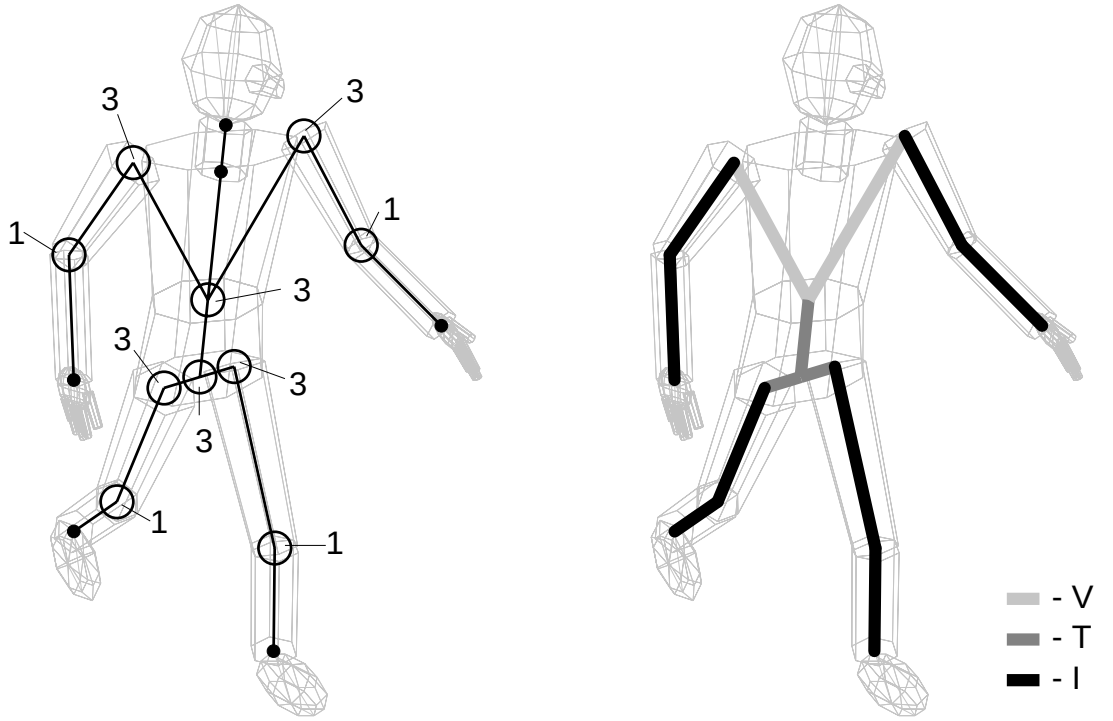
In diesem Abschnitt wird am Beispiel der Modellierung des menschlichen Körpers aufgezeigt, wie für die im Interpretationsprozeß bestimmten Hypothesen die Rotationsanteile in den Transformationen zwischen den lokalen Koordinatensystemen der Objektmodellteile ermittelt werden. Bei der Bewertung und der Auswahl der Hypothesen werden diese Gelenkwinkel mit einer entsprechenden Bewertungsregel berücksichtigt, vgl. Kap. 3.6 und 3.7.

Nachdem in STABIL^{++} Punkt- oder Flächenmerkmale verwendet werden, ist mit einer gültigen Zuordnung der 3D Szenenmerkmale $\mathbf{s}_i$ zu den 3D Modellmerkmalen $\mathbf{m}_j$ im Interpretationsprozeß zunächst nur die hierarchische, innere Objektmodellstruktur des zu detektierenden Objektes bestimmt. Es sind mit den 3D Punkten der Gelenke die Ursprünge der lokalen Koordinatensysteme in den Objektmodellteilen bestimmt worden, jedoch nicht die Orientierung der Objektmodellteile. Somit sind für die geometrische Objektmodellstruktur die Translationsanteile der Transformationsmatrizen bestimmt. Um jedoch die geometrische Struktur komplett und anschließend auch die äußere Objektmodellstruktur nutzen zu können, müssen die Rotationen zwischen den Objektmodellteilen bestimmt werden.

Bei der eingeführten Beispielmodellierung des menschlichen Körpers mit 16 Objektmodellteilen entsprechend Tab. 2.1 ergeben sich 16×3 Freiheitsgrade für die Rotationen zwischen den Objektmodellteilen. Es stehen jedoch nur maximal 16 3D Punkte für die Lösung dieses Problems zur Verfügung. Daher müssen Annahmen gemacht oder Heuristiken vorausgesetzt werden. Zunächst wird die Anzahl der zu bestimmenden Gelenkwinkel auf die bestimmbaren Gelenkwinkel eingeschränkt. Im weiteren wird angenommen, daß die Ursprünge der lokalen Koordinatensysteme bestimmter Objektmodellteile in der gleichen Ebene liegen. Diese Objektmodellteile werden für die Gelenkwinkelbestimmung gruppiert und bilden sog. *Kompositionen*. Die verschiedenen Gruppierungen der Objektmodellteile des Objektmodells wird im folgenden Abschnitt beschrieben.

Die Beschreibung der Kompositionen stützt sich in den folgenden Abschnitten auf die Beispielmodellierung des menschlichen Körpers, wie sie in Kap. 2.3 vorgestellte wurde. Die Verwendung der Kompositionen lassen sich jedoch auf die Modellierungen anderer Objekte übertragen. Hierzu ist zu beachten, daß die, zur Beschränkung der rotatorischen Freiheitsgrade in den Gelenken, zu verwendenden Heuristiken sinnvoll angewendet werden können.

Hierzu werden zunächst die Gelenkwinkel der Blätter der hierarchischen, inneren Objektmodellstruktur von der Gelenkwinkelbestimmung ausgeschlossen. Dies sind die Objektmodellteile für den Kopf, die Hände und die Füße. Somit bleiben die Winkelstellungen in den entsprechenden Objektmodellteilen immer entsprechend den Winkeln im initialen Objektmodell bestehen. Die Orientierung dieser Objektmodellteile im Raum ist somit von den jeweiligen Vorgängerobjektmodellteilen abhängig. Desweiteren wird auch das Objektmodellteil des Halses von der Gelenkwinkelbestimmung ausgenommen. Somit ist die Orientierung der lokalen Koordinatensysteme in den Objektmodellteilen für Hals und Kopf starr an das Objektmodellteil



(a) *(b)*

Abbildung A.1: *(a) Anzahl der Freiheitsgrade bei der Gelenkwinkelbestimmung und (b) Notwendige Kompositionen von Objektmodellteilen für die Gelenkwinkelbestimmung.*

des Rumpfes gebunden. Wird jedoch eine weitreichendere Modellierung gewählt, bei der am Kopf z.B. die beiden Augen oder die beiden Ohren berücksichtigt sind, kann auch die Orientierung des Kopfes / Halses bestimmt werden. Die Problematik der starren Gelenkwinkel verlagert sich dann in der hierarchischen Objektmodellstruktur weiter nach außen, so daß dann für die Augen oder für die Ohren nur die Position und keine weitere Orientierung bestimmt werden kann. Die notwendige Feinheit der Modellierung ist von der Anwendung abhängig, vgl. Kap. 4.

In den Gelenken der verbleibenden Objektmodellteile sind dann die rotatorischen Freiheitsgrade entsprechend der Anatomie des menschlichen Körpers einzuschränken. So sind die Ellbogen- und Kniegelenke Scharniergelenke, d.h. es findet nur eine Rotation um eine Achse statt. Vgl. hierzu die Ausführungen zur Anatomie der Gelenke in [Kap92a] und [Kap92c]. Die Rotationen, die z.B. nur im gebeugten Kniegelenk um die y-Achse auftreten können, werden hier vernachlässigt werden. In Abb. A.1 (a) sind die verbleibenden Freiheitsgrade an den Gelenkpunkten zwischen den Objektmodellteilen verzeichnet.

A.2 Kompositionen von Objektmodellteilen

Die für die Bestimmung der Gelenkwinkel notwendige Zusammenfassung / Gruppierung von Objektmodellteilen, deren Ursprünge ihrer lokalen Koordinatensysteme in einer Ebene liegen, sind in Abb. A.1 (b) zu sehen. Dort ist zu erkennen, daß die Objektmodellteile der Hüfte, des Rumpfes, des linken und rechten Oberschenkels zu einer Komposition zusammengefaßt sind.

#	Typ	Objektmodellteile	omp_b	omp_v	${}^{omp_v}\mathbf{T}_{wcs}$
1	T	0.1, 1.1, 1.3, 1.2	$omp_{0.1}$	omp_{obj}	${}^{omp_{obj}}\mathbf{T}_{wcs}$
2	V	1.2, 2.2, 2.4	$omp_{1.2}$	$omp_{0.1}$	${}^{omp_{0.1}}\mathbf{T}_{wcs}$
3	I	1.1, 2.1, 3.1	$omp_{1.1}$	$omp_{0.1}$	${}^{omp_{0.1}}\mathbf{T}_{wcs}$
4	I	2.2, 3.2, 4.1	$omp_{2.2}$	$omp_{1.2}$	${}^{omp_{1.2}}\mathbf{T}_{wcs}$
5	I	1.3, 2.5, 3.5	$omp_{1.3}$	$omp_{0.1}$	${}^{omp_{0.1}}\mathbf{T}_{wcs}$
6	I	2.4, 3.4, 4.2	$omp_{2.4}$	$omp_{1.2}$	${}^{omp_{1.2}}\mathbf{T}_{wcs}$

Tabelle A.1: *Kompositionen von Objektmodellteilen für die Gelenkwinkelbestimmung mit Angabe des Bezugsobjektmodellteiles omp_b, jeweiligem Vorgängerobjektmodellteil omp_v und Transformationsmatrix ${}^{omp_v}\mathbf{T}_{wcs}$; Numerierung der Objektmodellteile entsprechend der Abb. 2.1.*

Ebenso das Objektmodellteil des Rumpfes und des linken und rechten Oberarms. Ansonsten bilden noch jeweils die drei Objektmodellteile, die die Gliedmaßen repräsentieren, eine Komposition. Es wird zwischen drei verschiedenen Typen von Kompositionen unterschieden. Die unterschiedlichen Kompositionen sind entsprechend der in Abb. A.1 (b) erkennbaren Form benannt: es wird zwischen *V-Kompositionen*, *T-Kompositionen* und *I-Kompositionen* unterschieden. So bilden die Objektmodellteile für Rumpf, den linken und rechten Oberarm eine V-Komposition, die Objektmodellteile für die Hüfte, den Rumpf und den linken und rechten Oberschenkel eine T-Komposition. Die Objektmodellteile der Gliedmaßen bilden jeweils eine I-Komposition.

Somit beinhaltet die Liste der Kompositionen $COMPO = \{compo_1, \ldots, compo_n\}$ der Definition des Objektmodells bei der verwendeten Modellierung sechs Kompositionen, vgl. Glg. 2.2. Eine Komposition ist dabei definiert als:

$$compo_\nu = \langle type, OMP \rangle \tag{A.1}$$

Hierbei ist der Typ der Komposition $type \in \{`V`, `T`, `I`\}$ und $OMP = \{omp_1, \ldots, omp_n\}$ die Liste der zusammengefaßten Objektmodellteile. Für die verwendete Modellierung ergeben sich die Kompositionen entsprechend Tab. A.1.

Entsprechend der Angabe der Objektmodellteillisten in der Tab. A.1 und der Darstellung der Freiheitsgrade in der Abb. A.1 werden bei den V- und T-Kompositionen jeweils die Winkel für das erste und bei den I-Kompositionen für das erste und zweite Objektmodellteil in der Liste bestimmt. Dabei stellt immer das erste Objektmodellteil der Listen das Bezugsobjektmodellteil omp_b einer Komposition dar. Bei der Gelenkwinkelbestimmung wird für die Transformationsmatrix des jeweiligen omp_b der Rotationsanteil bestimmt. Die Transformationsmatrix hat, entsprechend der Definition in Glg. 2.3, ihren Bezug zum lokalen Koordinatensystem des jeweiligen Vorgängerobjektmodellteils. In Tab. A.1 ist daher neben der Angabe der entsprechenden omp_b auch noch das jeweilige Vorgängerobjektmodellteil omp_v der Kompositionen vermerkt. Die Vorgängerobjektmodellteile für die zweiten Objektmodellteile einer I-Komposition sind die jeweils ersten Objektmodellteile der entsprechenden Liste der Objektmodellteile; vgl. hierzu auch den Abschn. A.6 zu den I-Kompositionen.

Die Annahme, daß für die Bestimmung der Rotationen die Objektmodellteile einer Komposition in einer Ebene liegen, beschränkt sicherlich die Freiheitsgrade des Modells, läßt sich aber mit der Art der Modellierung vereinbaren. Das Objektmodell ist, entsprechend der zu beschreibenden artikularen Objekte ein, aus einzelnen starren Objektmodellteilen zusammengesetztes Modell, bei dem keine dynamische Translation zwischen den Objektmodellteilen zulässig ist und somit sind Rotationen nur an den modellierten Gelenkpunkten möglich. Für das Beispiel

der V-Komposition der verwendeten Standardmodellierung des menschlichen Körpers, die aus den drei Objektmodellteilen des Rumpfes, des linken und des rechten Oberarmes gebildet ist, bedeutet dies, daß die Lage des oberen Teils des Objektmodellteils des Rumpfes durch die Ebene bestimmt wird, die durch die Gelenkpunkte in den Schultern und einem Rumpf-Punkt definiert ist. Damit wird vernachlässigt, daß die Bewegung in den Schultern nicht nur durch Rotationen in einem Kugelgelenk bestimmt ist und der Mensch die Schulter gegenüber dem Oberkörper nach vorne nehmen kann. Somit wird bei dem "Nachvornenehmen" beider Schultern dies als "Kippen" des Rumpfes nach vorn gedeutet und ein "Zurücknehmen" oder "Nachvornenehmen" einer Schulter wird als Drehung des Oberkörpers gedeutet. In dem CAD-Menschmodell RAMSIS werden die Bewegungen der Schulter durch die Modellierung der Schlüsselbeine berücksichtigt. Auf die Modellierung mit einem weiteren Gelenk wurde in STABIL^{++} verzichtet, da sich bei der Bildinterpretation dann ein Segmentierungsproblem ergibt, denn das Modell sollte nur so fein modelliert werden, wie auch die entsprechenden 2D Bildmerkmale aus dem Videobild segmentiert werden können.

Für eine T-Komposition sind weitere Annahmen zu machen, denn in vier beliebige Punkte in 3D läßt sich nicht eindeutig eine Ebene legen. Daher wird die Ebene zunächst durch drei Punkte aufgespannt und dann die so ermittelten Rotationen in den vierten Punkt projiziert. Die Einzelheiten sind hierzu in Kap. A.5 beschrieben. Für die T-Komposition aus der Standardmodellierung für die Objektmodellteile der Hüfte, des Rumpfes und der beiden Oberschenkel bedeutet dies, daß angenommen wird, daß die beiden Hüftgelenke mit dem Hüftpunkt des Modells auf einer Linie liegen. Durch das starre Becken des Menschen hat diese Annahme direkten Bezug zum menschlichen Skelett, vgl. [Kap92b]. Desweiteren wird angenommen, daß die Punkte der beiden Objektmodellteile der Oberschenkel mit dem Punkt des Objektmodellteils des Rumpfes in einer Ebene liegen. Daher wird ein Drehen oder Kippen der Hüfte, das über die Lage der beiden Punkte für die Objektmodellteile der Oberschenkel zu messen ist, als ein Drehen oder Kippen des Objektmodellteils der Hüfte interpretiert. Hierin liegt auch begründet, daß in Abb. A.1 für das Objektmodellteil der Hüfte nur drei Freiheitsgrade vermerkt sind, denn die rotatorischen Freiheitsgrade zu den beiden angrenzenden Objektmodellteilen der Oberschenkel entfallen entsprechend der Annahmen.

A.3 Reihenfolge der Kompositionen

Mit der Liste $COMPO = \{compo_1, \ldots, compo_n\}$ sind bei der Modellierung einem Objektmodell Kompositionen zugeordnet, vgl. hierzu Glg. 2.2. Die Reihenfolge der Kompositionen in der Liste gibt an, in welcher Reihenfolge die Kompositionen zur Bestimmung der Gelenkwinkel verwendet werden, vgl. auch Tab. A.1. Die Reihenfolge kann nicht willkürlich gewählt werden, denn die geometrischen Beziehungen, d.h. Translation und Rotation zwischen den Objektmodellteilen sind in den lokalen Koordinatensystemen des jeweiligen Vorgängerobjektmodellteils definiert. Es stehen jedoch für die Gelenkwinkelbestimmung durch die Komposition zunächst nur 3D Punkte, die im Weltkoordinatensystem definiert sind, zur Verfügung. Dies liegt in der Zuordnung von 3D Szenenmerkmalen zu 3D Modellmerkmalen während des Interpretationsprozeßes begründet. Mit den folgenden Überlegungen wird die Notwendigkeit einer Reihenfolge der Kompositionen näher erläutert.

Entsprechend der verwendeten Modellierung des menschlichen Körpers und den in den Kompositionen verwendeten Punkten, liegen mit einer gültigen und kompletten Zuordnung von

3D Szenenmerkmalen zu 3D Modellmerkmalen die 3D Punkte der folgenden Liste vor:[1]

$$\left\{ {}^{\circ}\vec{p}_{wcs}^{\,0.1},\ {}^{\circ}\vec{p}_{wcs}^{\,1.1},\ {}^{\circ}\vec{p}_{wcs}^{\,1.2},\ {}^{\circ}\vec{p}_{wcs}^{\,1.3},\ {}^{\circ}\vec{p}_{wcs}^{\,2.1},\ {}^{\circ}\vec{p}_{wcs}^{\,2.2},\ {}^{\circ}\vec{p}_{wcs}^{\,2.4},\ {}^{\circ}\vec{p}_{wcs}^{\,2.5},\ {}^{\circ}\vec{p}_{wcs}^{\,3.1},\ {}^{\circ}\vec{p}_{wcs}^{\,3.2},\ {}^{\circ}\vec{p}_{wcs}^{\,3.4},\ {}^{\circ}\vec{p}_{wcs}^{\,4.1},\ {}^{\circ}\vec{p}_{wcs}^{\,4.2} \right\} \quad \text{(A.2)}$$

Diese 3D Punkte sind auf das Weltkoordinatensystem wcs des Szenenmodells bezogen. Mit "$\circ$" ist gekennzeichnet, daß es sich bei den Punkten um den Ursprung des Koordinatensystems handelt, der hochgestellte Index bezeichnet die Nummer des zugehörigen Objektmodellteils. Für die Bestimmung der Rotationen im lokalen Koordinatensystem des Bezugsobjektmodellteils omp_b einer Komposition müssen die entsprechenden 3D Punkte aus dem Weltkoordinatensystem in das Koordinatensystem seines Vorgängerobjektmodellteils omp_v transformiert werden. Somit gilt für den Punkt des Ursprunges des lokalen Koordinatensystems eines Objektmodellteils omp_ν, das einer Komposition mit dem Vorgängerobjektmodellteil omp_v angehört:

$${}^{\circ}\vec{p}_{omp_v}^{\,\mu} = {}^{omp_v}\mathbf{T}_{wcs} \cdot {}^{\circ}\vec{p}_{wcs}^{\,\mu} \quad \text{(A.3)}$$

Auf diese Weise werden alle Punkte, die zur Bestimmung der Gelenkwinkel in einer Komposition verwendet werden, in das Koordinatensystem des jeweiligen Vorgängerobjektmodellteils omp_v transformiert.

Um die Bestimmung der durch die Punkte einer Komposition aufzuspannenden Ebene zu vereinfachen, wird das Koordinatensystem des Vorgängerobjektmodellteils omp_v in den Ursprung des lokalen Koordinatensystems des Bezugsobjektmodellteils omp_b verschoben. Somit wird die Transformation in Glg. A.3 um eine Translation ergänzt; es gilt dann:

$${}^{\circ}\vec{p}_{omp_v}^{\,\mu\,\prime} = {}^{\circ}\vec{p}_{omp_v}^{\,\mu} + \vec{t}_{omp_v}$$

Der Translationsvektor $\vec{t}_{omp_v}$ ist durch das Bezugsobjektmodellteil omp_b und das Vorgängerobjektmodellteil omp_v der Komposition bestimmt, es gilt:

$$\vec{t}_{omp_v} = {}^{omp_v}\mathbf{T}_{wcs} \cdot \left[{}^{\circ}\vec{p}_{wcs}^{\,b} - {}^{\circ}\vec{p}_{wcs}^{\,v} \right]$$

Somit gilt entsprechend der Verschiebung des lokalen Koordinatensystems des Vorgängerobjektmodellteils in das lokale Koordinatensystem des Bezugsobjektmodellteils für den Punkt:

$${}^{\circ}\vec{p}_{omp_v}^{\,b\,\prime} = [0,0,0]^T \quad \text{(A.4)}$$

Aus diesen Überlegungen heraus ist ersichtlich, daß die Gelenkwinkelbestimmung nur in der Reihenfolge der hierarchischen, inneren Objektmodellstruktur für die einzelnen Objektmodellteile durchgeführt werden kann. Dies liegt darin begründet, daß für die Bestimmung der Rotationen einer Komposition jeweils die Transformationsmatrix des Vorgängerobjektmodellteils ${}^{omp_v}\mathbf{T}_{wcs}$ bekannt sein muß. Bei der Definition der Kompositionen eines Objektmodells sind die Kompositionen daher in einer entsprechenden Reihenfolge anzugeben. In Tab. A.1 ist diese Reihenfolge berücksichtigt. Zusätzlich ist in der letzten Spalte noch die für jede Komposition notwendige Transformationsmatrix angegeben. Zuerst wird daher der Rotationsanteil der Transformationsmatrix des Objektmodellteils der Hüfte bestimmt, daran schließt sich die Bestimmung für das Objektmodellteil des Rumpfes an. Im weiteren können dann die Gelenkwinkel in den I-Kompositionen für die vier Gliedmaßen bestimmt werden.

Eine Besonderheit stellt die Transformationsmatrix ${}^{obj}\mathbf{T}_{wcs}$ für die jeweils erste Komposition dar, denn hierzu muß zuerst die Lage des Koordinatensystems des Objektmodells bestimmt

[1] Vgl. Tab. 2.1 mit Numerierung der Objektmodellteile der Beispielmodellierung.

werden. Hierbei wird zum einen ausgenutzt, daß die Translation zwischen dem lokalen Koordinatensystem des ersten Objektmodellteiles und dem Koordinatensystem des Objektmodells fixiert ist. Zum anderen wird ausgenutzt, daß der Rotationsanteil der Transformationsmatrix ${}^{wcs}\mathbf{T}_{obj}$ vom Weltkoordinatensystem des Szenenmodells zum Objektmodell ebenfalls fix ist. Es bleibt somit nur die Position des Koordinatensystems des Objektmodells zu bestimmen, das sich damit dann aus der Position des Ursprungs des lokalen Koordinatensystems des ersten Objektmodellteils $omp_{0.1}$ bestimmen läßt. Es gilt entsprechend den Ausführungen zur homogenen Koordinatentransformation in Anh. B.4: ${}^{wcs}\mathbf{T}_{obj} = \mathcal{T} \cdot \mathcal{R}$. Die Rotationsmatrix $\mathcal{R}$ ist, wie o.a. bekannt, so daß nur die Translationsmatrix $\mathcal{T}$ bestimmt werden muß. Für den der Translationsmatrix $\mathcal{T}$ entsprechenden Translationsvektor $\vec{t}$ gilt:

$$\vec{t} = {}^{\circ}\vec{p}^{\,0.1}_{omp_v} - \mathcal{R}_{3\times3} \cdot \vec{t}^{\,0.1/obj}_{obj} \tag{A.5}$$

wobei $\mathcal{R}_{3\times3}$ die fixe Rotationsmatrix der Transformationsmatrix ${}^{wcs}\mathbf{T}_{obj}$ und $\vec{t}^{\,0.1/obj}_{obj}$ die fixe Translation zwischen dem Koordinatensystem des Objektmodells obj und des lokalen Koordinatensystems des ersten Objektmodellteils $omp_{0.1}$ ist.

Ist im Interpretationsprozeß keine komplette Zuordnung von 3D Szenenmerkmalen zu 3D Modellmerkmalen gefunden worden und somit die Liste aus Glg. A.2 nicht komplett, so können nicht alle Rotationen zwischen den Objektmodellteilen bestimmt werden. Es können die Rotationen mit einer Komposition nur bestimmt werden, wenn für alle Objektmodellteile, die zu einer Komposition gruppiert sind, die 3D Punkte für die Ursprünge der lokalen Koordinatensysteme vorhanden sind. Entsprechend der Reihenfolge der Kompositionen, können die Gelenkwinkel eines Modells nur soweit bestimmt werden, wie für die entsprechenden Kompositionen die Rotationen der jeweiligen Vorgängerobjektmodellteile bestimmbar sind. Für nicht bestimmbare Gelenkwinkel in einem Objektmodellteil omp_ν können die Winkel der Transformationsmatrix ${}^{omp_\mu}\mathbf{T}_{omp_{\nu,t-1}}$ aus der Historie $HIST$ des Objektmodellteils übernommen werden.

A.4 V-Komposition

Eine V-Komposition setzt sich generell aus drei Objektmodellteilen zusammen. Für die Bestimmung der Rotationen des Bezugsobjektmodellteils omp_b der Komposition wird angenommen, daß die Ursprünge der lokalen Koordinatensysteme der drei Objektmodellteile in einer Ebene liegen. Die entsprechenden drei 3D Punkte seien mit $\vec{o}$, $\vec{l}$ und $\vec{m}$ bezeichnet und sind in einem Koordinatensystem (X'_{omp_v}, Y'_{omp_v}, Z'_{omp_v}) definiert. Dieses entspricht in der Orientierung dem Koordinatensystem (X_{omp_v}, Y_{omp_v}, Z_{omp_v}) des Vorgängerobjektmodellteils omp_v, ist jedoch durch eine Parallelverschiebung in den Ursprung des lokalen Koordinatensystems des Bezugsobjektmodellteils omp_b verschoben. Entsprechend der verwendeten Beispielmodellierung und der entsprechenden Gruppierung der Objektmodellteile in Kompositionen gilt beispielsweise für die V-Komposition der Objektmodellteile des Rumpfes und des rechten und des linken Oberarmes:

$$\begin{aligned}
\vec{o} &= {}^{omp_{0.1}}\mathbf{T}_{wcs} \cdot {}^{\circ}\vec{p}^{\,1.2}_{wcs} + \vec{t}_{omp_{0.1}} \\
\vec{l} &= {}^{omp_{0.1}}\mathbf{T}_{wcs} \cdot {}^{\circ}\vec{p}^{\,2.2}_{wcs} + \vec{t}_{omp_{0.1}} \\
\vec{m} &= {}^{omp_{0.1}}\mathbf{T}_{wcs} \cdot {}^{\circ}\vec{p}^{\,2.4}_{wcs} + \vec{t}_{omp_{0.1}} \\
&\text{mit} \\
\vec{t}_{omp_{0.1}} &= {}^{omp_{0.1}}\mathbf{T}_{wcs} \cdot [{}^{\circ}\vec{p}^{\,1.2}_{wcs} - {}^{\circ}\vec{p}^{\,0.1}_{wcs}]
\end{aligned}$$

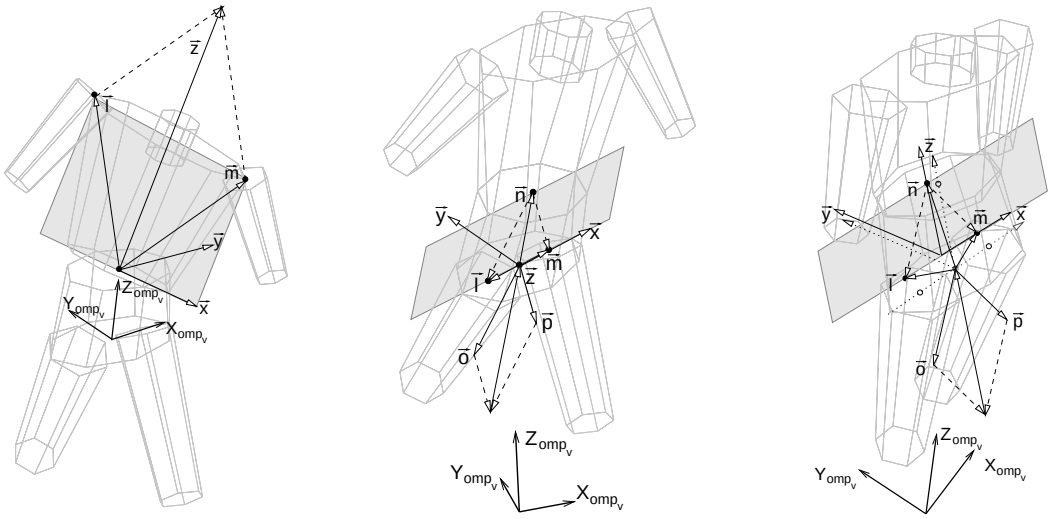


Abbildung A.2: *(a) V-Komposition zur Bestimmung der Rotationen im Objektmodellteil des Rumpfes; (b), (c) T-Komposition zur Bestimmung der Rotationen im Objektmodellteil der Hüfte; (b) idealisierte und (c) reale Bestimmung der Rotationen mit der T-Komposition.*

Mit dem Punkt $\vec{o}$ ist in der V-Komposition der Ursprung des Koordinatensystems des Bezugsobjektmodellteils omp_b definiert. $\vec{l}$ und $\vec{m}$ sind jeweils die Punkte der Ursprünge der lokalen Koordinatensysteme der dem Bezugsobjektmodellteil omp_b in der hierarchischen, inneren Modellstruktur folgenden Objektmodellteile und spannen mit $\vec{o}$ die für die Bestimmung der Rotationen notwendige Ebene auf, vgl. Abb. A.2 (a).

Auf der Basis dieser Ebene wird die Orientierung des lokalen Koordinatensystems des Bezugsobjektmodellteils bestimmt. Hierzu wird ein neues Orthonormalsystem definiert. Die z-Achse des neuen Systems verläuft von dem Punkt $\vec{o}$ in der Mitte zwischen den beiden Punkten $\vec{l}$ und $\vec{m}$. Die y-Achse verläuft senkrecht zu der durch $\vec{l}$ und $\vec{m}$ aufgespannten Ebene. Die x-Achse steht wiederum auf einer Ebene, die durch die y- und z-Achse aufgespannt wird, senkrecht und bildet mit diesen ein Rechtssystem. Durch die Reihenfolge der drei Punkte in der Liste der Objektmodellteile der V-Komposition ist die Orientierung der Ebene und somit die Orientierung des neuen Orthonormalsystems eindeutig bestimmt. Für die Richtungen der drei Achsen ergibt sich dann:

$$\begin{aligned} \vec{z} &= \vec{l} + \vec{m} \\ \vec{y} &= \vec{l} \times \vec{m} \\ \vec{x} &= \vec{y} \times \vec{z} \end{aligned}$$

Der Rotationsanteil der Transformationsmatrix des Bezugsobjektmodellteils omp_b läßt sich aus den Vektoren $\vec{x}$, $\vec{y}$ und $\vec{z}$ entsprechend der Glg. B.4 - B.9 direkt erzeugen. Hierbei werden die Vektoren auf die Länge eins normiert und ergeben somit die spaltenweisen Einträge in der Rotationsmatrix.

A.5 T-Komposition

Eine T-Komposition setzt sich generell aus vier Objektmodellteilen zusammen. Die für die Bestimmung der Rotation im Bezugsobjektmodellteil omp_b notwendigen 3D Punkte der Ursprünge der lokalen Koordinatensysteme der vier Objektmodellteile seien mit $\vec{o}$ für das Bezugsobjektmodellteil und mit $\vec{l}$, $\vec{m}$ und $\vec{n}$ für die dem Bezugsobjektmodellteil in der hierarchischen, inneren Modellstruktur folgenden Objektmodellteile bezeichnet. Diese Vektoren sind in einem Koordinatensystem $[X'_{omp_v}, Y'_{omp_v}, Z'_{omp_v}]$ definiert, das in der Orientierung dem Koordinatensystem $[X_{omp_v}, Y_{omp_v}, Z_{omp_v}]$ des Vorgängerobjektmodellteils omp_v entspricht, jedoch durch eine Parallelverschiebung in den Ursprung des lokalen Koordinatensystems des Bezugsobjektmodellteils omp_b verschoben ist.

Für die Bestimmung der Rotationen des Bezugsobjektmodellteils omp_b wird auch hier angenommen, daß die Ursprünge der lokalen Koordinatensysteme der vier Objektmodellteile in einer Ebene liegen. Jedoch ist eine Ebene von nur drei 3D Punkten eindeutig bestimmt. Daher wird die Ebene zunächst von den Punkten der drei Objektmodellteile der T-Komposition bestimmt, die nicht dem Bezugsobjektmodellteil entsprechen. Dies sind die Punkte $\vec{l}$, $\vec{m}$ und $\vec{n}$. Entsprechend der verwendeten Beispielmodellierung und der entsprechenden Gruppierung der Objektmodellteile zu Kompositionen gilt beispielsweise für die T-Komposition der Objektmodellteile der Hüfte, des Rumpfes und des rechten und des linken Oberschenkels:

$$\begin{aligned}
\vec{o} &= {}^{obj}\mathbf{T}_{wcs} \cdot {}^{\circ}\vec{p}^{\,0.1}_{wcs} + \vec{t}_{obj} \\
\vec{l} &= {}^{obj}\mathbf{T}_{wcs} \cdot {}^{\circ}\vec{p}^{\,1.1}_{wcs} + \vec{t}_{obj} \\
\vec{m} &= {}^{obj}\mathbf{T}_{wcs} \cdot {}^{\circ}\vec{p}^{\,1.3}_{wcs} + \vec{t}_{obj} \\
\vec{n} &= {}^{obj}\mathbf{T}_{wcs} \cdot {}^{\circ}\vec{p}^{\,1.2}_{wcs} + \vec{t}_{obj}
\end{aligned}$$

Wobei $\vec{t}_{obj}$ der bei der Modellierung festgelegte Versatz zwischen dem Koordinatensystem des Objektmodells und dem lokalen Koordinatensystem des ersten Objektmodellteils $omp_{0.1}$ ist.

In Abb. A.2 (b) ist zu erkennen, daß die Punkte $\vec{l}$, $\vec{m}$ und $\vec{n}$, ähnlich wie bei der V-Komposition, eine Ebene aufspannen. Daher ergibt sich für das neue Orthonormalsystem, das die Rotation dieser Ebene zum Vorgängerobjektmodellteil beschreibt:

$$\begin{aligned}
\vec{z} &= -(\vec{o} + \vec{p}) \\
\vec{y} &= \vec{o} \times \vec{p} \\
\vec{x} &= \vec{y} \times \vec{z} \\
&\text{mit} \\
\vec{o} &= \vec{l} - \vec{n} \\
\vec{p} &= \vec{m} - \vec{n}
\end{aligned}$$

In dem Beispiel in Abb. A.2 (b) ist eine ideale T-Komposition abgebildet, denn dort liegt der Punkt des Bezugsobjektmodellteils $\vec{o}$ ebenfalls in der durch $\vec{l}$, $\vec{m}$ und $\vec{n}$ aufgespannten Ebene. Somit ist durch die Orientierung der Ebene die Orientierung des lokalen Koordinatensystems des Bezugsobjektmodellteils eindeutig bestimmt. In dem Beispiel, das in Abb. A.2 (c) dargestellt ist, ist dies nicht der Fall. Dort liegt der Punkt $\vec{o}$ vor der Ebene, die durch $\vec{l}$, $\vec{m}$ und $\vec{n}$ aufgespannt ist.[2] Somit liegt das durch die T-Komposition bestimmte neue Orthonormalsystem mit dem Ursprung im Punkt $\vec{o}$ jedoch mit seiner xz-Ebene parallel zu der durch $\vec{l}$, $\vec{m}$ und $\vec{n}$

[2] Die Lage der Punkte in diesem Beispiel der T-Komposition ist der Interpretation einer realen Bildsequenz entnommen.

aufgespannten Ebene. In Abb. A.2 (c) kennzeichnen die dunkel eingezeichneten Punkte die dem Modell zugeordneten 3D Szenenmerkmale. Die als Kreise eingezeichneten Punkte sind die 3D Punkte, die sich für die Ursprünge der Objektmodellteile nach der Anwendung der T-Komposition ergeben. Diese Punkte entsprechen dann auch den Ursprüngen der gesetzten lokalen Koordinatensysteme der einzelnen Objektmodellteile. Daher liegen diese Punkte auch auf den Rotationsachsen der im Hintergrund der Abbildung heller skizzierten Objektmodellteile. Das neue Orthonormalsystem ist mit gestrichelten Linien im Punkt $\vec{o}$ eingezeichnet, zum Vergleich ist mit durchgängigen Linien das parallel verschobene Koordinatensystem in der Ebene eingezeichnet.

Es ist zu erkennen, daß die drei Punkte für die Objektmodellteile der Hüfte und der beiden Oberschenkel entsprechend der Modellierung auf einer Linie liegen, vgl. hierzu die Ausführungen in Abschn. A.2. Dementsprechend kommt auch der Ursprung des Objektmodellteiles des Rumpfes vor dem 3D Punkt des entsprechenden Szenenmerkmals zu liegen. Diese Ungenauigkeit der Modellierung wird in STABIL^{++} generell zugunsten der einfachen Realisierung der Gelenkwinkelbestimmung in Kauf genommen. Die maximal tolerierbaren Abweichungen, die hierdurch entstehen können, sind jedoch von der jeweiligen Anwendung des Systems abhängig.

Der Rotationsanteil der Transformationsmatrix des Bezugsobjektmodellteils omp_b läßt sich, wie bei der V-Komposition direkt aus den Vektoren $\vec{x}$, $\vec{y}$ und $\vec{z}$ erzeugen.

A.6 I-Komposition

Eine I-Komposition setzt sich generell aus drei Objektmodellteilen zusammen. Es werden durch diese Kompositionen die Rotationen in den drei Freiheitsgraden des Bezugsobjektmodellteils omp_b und eine weitere Rotation um eine Achse in dem jeweils zweiten Objektmodellteil der Liste der Objektmodellteile der Komposition bestimmt. Die I-Kompositionen werden im Objektmodell des Menschen zur Bestimmung der Gelenkwinkel in den Gliedmaßen verwendet. Daher handelt es sich bei der Rotation im zweiten Objektmodellteil der I-Komposition um die Rotationen entsprechend dem Knie- und dem Ellbogengelenk. Aufgrund der gewählten Modellierung ist die Rotationsachse im Knie- und Ellbogengelenk die x-Achse.

Auch bei der Bestimmung der Rotationen mit der I-Komposition wird durch die drei Punkte der Ursprünge der lokalen Koordinatensysteme der gruppierten Objektmodellteile eine Ebene aufgespannt. Die drei 3D Punkte seien mit $\vec{o}$, $\vec{l}$ und $\vec{m}$ bezeichnet und sind in einem Koordinatensystem $[X'_{omp_v}, Y'_{omp_v}, Z'_{omp_v}]$ definiert, das in der Orientierung dem Koordinatensystem $[X_{omp_v}, Y_{omp_v}, Z_{omp_v}]$ des Vorgängerobjektmodellteils omp_v entspricht, jedoch durch eine Parallelverschiebung in den Ursprung des lokalen Koordinatensystems des Bezugsobjektmodellteils omp_b verschoben ist. Entsprechend der verwendeten Beispielmodellierung und der Gruppierung der Objektmodellteile in Kompositionen gilt beispielsweise für die I-Komposition der Objektmodellteile des linken Oberschenkels, des linken Unterschenkels und des linken Knies:

$$
\begin{aligned}
\vec{o} &= {}^{omp_{0.1}}\mathbf{T}_{wcs} \cdot {}^{\circ}\vec{p}^{\,1.3}_{wcs} + \vec{t}_{omp_{0.1}} \\
\vec{l} &= {}^{omp_{0.1}}\mathbf{T}_{wcs} \cdot {}^{\circ}\vec{p}^{\,2.5}_{wcs} + \vec{t}_{omp_{0.1}} \\
\vec{m} &= {}^{omp_{0.1}}\mathbf{T}_{wcs} \cdot {}^{\circ}\vec{p}^{\,3.5}_{wcs} + \vec{t}_{omp_{0.1}} \\
\text{mit} & \\
\vec{t}_{omp_{0.1}} &= {}^{omp_{0.1}}\mathbf{T}_{wcs} \cdot [{}^{\circ}\vec{p}^{\,1.3}_{wcs} - {}^{\circ}\vec{p}^{\,0.1}_{wcs}]
\end{aligned}
$$

Für die I-Komposition der Objektmodellteile für den rechten Oberarm, den rechten Unterarm und die rechte Hand gilt entsprechend:

$$
\begin{aligned}
\vec{o} &= {}^{omp_{1.2}}\mathbf{T}_{wcs} \cdot {}^{\circ}\vec{p}^{\,2.2}_{wcs} + \vec{t}_{omp_{1.2}} \\
\vec{l} &= {}^{omp_{1.2}}\mathbf{T}_{wcs} \cdot {}^{\circ}\vec{p}^{\,3.2}_{wcs} + \vec{t}_{omp_{1.2}} \\
\vec{m} &= {}^{omp_{1.2}}\mathbf{T}_{wcs} \cdot {}^{\circ}\vec{p}^{\,4.1}_{wcs} + \vec{t}_{omp_{1.2}} \\
\text{mit} & \\
\vec{t}_{omp_{1.2}} &= {}^{omp_{1.2}}\mathbf{T}_{wcs} \cdot [{}^{\circ}\vec{p}^{\,2.2}_{wcs} - {}^{\circ}\vec{p}^{\,1.2}_{wcs}]
\end{aligned}
$$

Die für die I-Kompositionen durch die Punkte $\vec{o}$, $\vec{l}$ und $\vec{m}$ aufgespannte Ebene ist in ihrer Orientierung nicht eindeutig bestimmt. Entsprechend eines Vergleiches Abb. A.3 (a) mit (d), (b) mit (e) und (c) mit (f) ist zu erkennen, daß sich bei gleicher Position der 3D Punkte $\vec{o}$, $\vec{l}$ und $\vec{m}$ jeweils eine unterschiedliche Rotation um den Oberarmknochen oder Oberschenkelknochen ergeben kann, dies entspricht einer Drehung um die z''-Achse in xzz''-Notation.[3] In Abb. A.3 (a) und (b) ist der Fuß nach vorn und in den Abb. A.3 (d) und (e) nach hinten orientiert. Im Beispiel des Arms ist der Daumen in Abb. A.3 (c) nach unten und in Abb. A.3 (f) nach oben orientiert. Dies entspricht jeweils einer um 180° unterschiedlichen Drehung um die z''-Achse und somit um den Oberschenkelknochen oder Oberarmknochen.

Die obere Reihe in Abb. A.3 stellt für das System gültige Winkelstellungen für die Drehung um die z''-Achse dar, die untere Reihe stellt die entsprechend ungültigen Winkelstellungen dar. In den Abb. A.3 (a) und (d) ist dies auch einfach nachvollziehbar. In Abb. A.3 (a) ist das Bein nach vorn orientiert und um ca. 40° im Knie abgewinkelt. Mit der Winkelbestimmung entsprechend Abb. A.3 (d) würde sich für das Kniegelenk eine Überstreckung um ca. 40° und eine Orientierung des Beins nach hinten ergeben. Damit erhält man jeweils eine unterschiedliche Orientierung der x-Achse im Kniegelenk ($\vec{x}_2$) und der x-Achse des lokalen Koordinatensystems des Bezugsobjektmodellteils ($\vec{x}$).

Die Tatsache, daß der Mensch weder das Knie- noch das Ellbogengelenk überstrecken kann,[4] kann nicht als Heuristik zur Unterscheidung der jeweils um 180° um die z''-Achse unterschiedlichen Lösungen verwendet werden. Dies liegt darin begründet, daß durch Ungenauigkeiten bei der Bestimmung der 3D Position der Szenenmerkmale es sehr wohl passieren kann, daß das dem Objektmodellteil des Fußes zugeordnete Szenenmerkmal vor der xz-Ebene des lokalen Koordinatensystems des Objektmodellteils des Oberschenkels liegt. Dies ist in Abb. A.3 (b) für das linke Bein und die gleiche Situation in Abb. A.3 (c) für den rechten Arm dargestellt, wobei jedoch hier zur Verdeutlichung der Problematik eine extreme Überstreckung von ca. 40° dargestellt ist. Zudem ergeben sich für die Winkelstellungen, wie sie in den Abb. A.3 (e) und (f) eingezeichnet sind, keine Überstreckung für das Knie- und das Ellbogengelenk, jedoch ist der Oberschenkel und der Oberarm nach hinten verdreht.

Zur Unterscheidung der jeweils zwei Lösungen wird daher die Orientierung der x-Achse des lokalen Koordinatensystems des Bezugsobjektmodellteils, hier des Oberschenkels und des Oberarms verwendet. Für die Orientierung dieser Achse im Koordinatensystem des Vorgängerobjektmodellteils ergibt sich ein maximal erlaubter Bereich, der in Abb. A.3 jeweils als dunkel hinterlegtes Kreissegment gekennzeichnet ist. Konkret bedeutet dies, daß die Projektion der

[3] Vgl. Ausführungen zu Rotationen in Koordinatensystemen in Anh. B.5.

[4] I. A. Kapandji gibt in [Kap92a], S. 70 an, daß lediglich eine passive Streckung von 5° bis 10° über die Neutralnullstellung hinaus im Knie möglich ist. Für das Ellbogengelenk gilt entsprechend [Kap92c], S. 94, daß es definitionsgemäß keine Extension im Ellbogengelenk gibt, jedoch bei Frauen oder Kindern mit besonderer Nachgiebigkeit des Bandapparates das Ellbogengelenk um 5° bis 10° hyperextendiert werden kann.

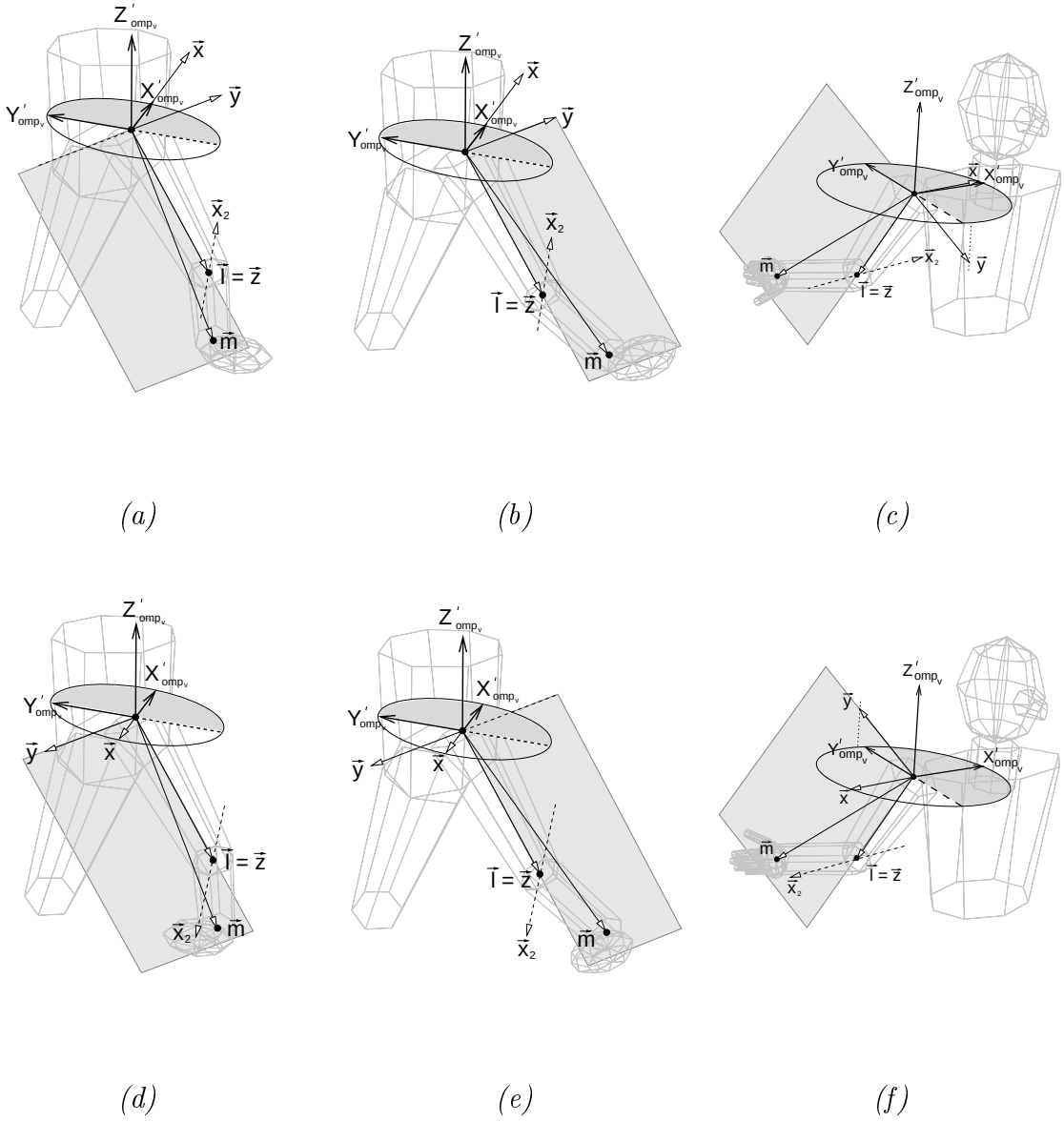


Abbildung A.3: *I-Kompositionen: In der oberen Reihe sind Bein und Arm nach vorn orientiert (erlaubte Rotation um z''-Achse); die untere Reihe zeigt die, bei gleicher Position der 3D Punkte, um 180° verdrehte Stellung von Bein und Arm (nicht erlaubte Rotation).*

x-Achse ($\vec{x}$) in die xy-Ebene des verschobenen lokalen Koordinatensystems des Vorgängerobjektmodellteils $[X'_{omp_v}, Y'_{omp_v}, Z'_{omp_v}]$ einen positiven Wert in der x-Komponente aufweisen muß. Dies gilt jedoch nur, da die Drehung im Oberschenkel und im Oberarm um die z''-Achse begrenzt ist und somit das Bein und der Arm immer nach vorn orientiert sein müssen.[5] In den Beispielen in Abb. A.3 ist diese Drehung jeweils 0°/180°, so daß die x-Achse selbst in der Ebene zum liegen kommt.

[5] I. A. Kapandji gibt die maximale Längsrotation im Hüftgelenk in [Kap92a], S.12 mit 30° nach außen und 60° nach innen an. Für die maximalen Rotationen um den Oberarmknochen gibt er in [Kap92c], S. 31 80° nach außen und 90° nach innen an.

Aufgrund dieser Vorüberlegungen ergibt sich für das neue Orthonormalsystem, das die Orientierung des lokalen Koordinatensystems des Bezugsobjektmodellteils einer I-Komposition bestimmt, folgendes:[6]

$$\begin{aligned}
\vec{z} &= \vec{l} \\
\vec{x} &= \begin{cases} \vec{x}' & , \quad \vec{x}'(x) > 0 \\ \vec{l} \times \vec{m} & , \quad \text{sonst} \end{cases} \\
\vec{y} &= \vec{z} \times \vec{x} \\
&\text{mit} \\
\vec{x}' &= \vec{m} \times \vec{l}
\end{aligned}$$

Entsprechend verläuft die z-Achse des neuen Orthonormalsystems entlang des Oberschenkel- oder Oberarmknochens. Die x-Achse steht zu der von den drei Punkten $\vec{o}$, $\vec{l}$ und $\vec{m}$ aufgespannten Ebene senkrecht, wobei die Ebene so orientiert ist, daß die x-Achse im erlaubten Drehbereich zu liegen kommt. Die y-Achse steht wiederum senkrecht zu der x- und der z-Achse und bildet mit diesen ein Rechtssystem. Somit liegt die y-Achse in der durch die Punkte $\vec{o}$, $\vec{l}$ und $\vec{m}$ aufgespannten Ebene.

Der Rotationsanteil der Transformationsmatrix des Bezugsobjektmodellteils omp_b läßt sich – wie bei der V-Komposition – direkt aus den Vektoren $\vec{x}$, $\vec{y}$ und $\vec{z}$ ablesen. Die so entstandene Rotationsmatrix ist die Grundlage für die Bestimmung des Gelenkwinkels in dem zweiten Objektmodellteil der I-Komposition. Diese Rotationsmatrix soll mit ${}^{omp_v}\mathcal{R}_{omp_b}$ bezeichnet werden, da diese den Rotationsanteil der Transformationsmatrix ${}^{omp_v}\mathbf{T}_o mp_b$ beschreibt. Die Inverse dieser Rotationsmatrix ist dementsprechend ${}^{omp_b}\mathcal{R}_{omp_v}$. Somit gilt für das neue Orthonormalsystem, das die Rotation um die x_2-Achse des lokalen Koordinatensystems des zweiten Objektmodellteils beschreibt:

$$\begin{aligned}
\vec{z}_2 &= {}^{omp_b}\mathcal{R}_{omp_v} \cdot (\vec{m} - \vec{l}) \\
\vec{x}_2 &= [1,0,0]^T \\
\vec{y}_2 &= \vec{z}_2 \times \vec{x}_2
\end{aligned}$$

Die z-Achse des neuen Systems liegt somit in dem Unterschenkel- oder Unterarmknochen. Die x-Achse liegt parallel zur x-Achse des lokalen Koordinatensystems des ersten Objektmodellteils der I-Komposition, dem Bezugsobjektmodellteil der I-Komposition und die y-Achse steht auf der x- und z-Achse senkrecht und bildet mit diesen ein Rechtssystem. Auch der Rotationsanteil der Transformationsmatrix des zweiten Objektmodellteils der I-Komposition läßt sich aus den Vektoren $\vec{x}_2$, $\vec{y}_2$ und $\vec{z}_2$ direkt ableiten.

A.7 Grenzwinkel

Zur Beurteilung der im Interpretationsprozeß aufgestellten Hypothesen werden die Gelenkwinkel zwischen den Objektmodellteilen auf ihre Zulässigkeit überprüft. Liegt ein Gelenkwinkel nicht innerhalb zulässiger Grenzen, den sog. *Grenzwinkeln*, dann kann die aufgestellte Hypothese von Zuordnungen von Szenenmerkmalen zu Modellmerkmalen nicht akzeptiert werden.

Jedem Objektmodellteil ist mit $restr_{\nu}^{\angle}$ jeweils drei minimale und drei maximale Winkel entsprechend der möglichen Freiheitsgrade in der Rotation zu seinem Vorgängerobjektmodellteil bekannt, vgl. Glg. 2.3. Die Grenzwinkel geben die Rotationsgrenzen für Rotationen an,

[6] Unter Berücksichtigung der entsprechenden Gelenkwinkelbeschränkungen.

Körperteil	#	Rotations-	α		β		γ	
		System	neg	pos	neg	pos	neg	pos
Rumpf	1.2	xyz	-60°	105°	-40°	40°	-40°	40°
rechter Oberschenkel	1.1	xyz''	-20°	100°	-45°	30°	-30°	60°
linker Oberschenkel	1.3	xyz''	-20°	100°	-30°	45°	-60°	30°
rechter Unterschenkel	2.1	xyz	-10°	120°	–	–	–	–
linker Unterschenkel	2.5	xyz	-10°	120°	–	–	–	–
rechter Oberarm	2.2	xyz''	-130°	0°	-120°	30°	-85°	80°
linker Oberarm	2.4	xyz''	-130°	0°	-30°	120°	-80°	85°
rechter Unterarm	3.2	xyz	-145°	10°	–	–	–	–
linker Unterarm	3.4	xyz	-145°	10°	–	–	–	–

Tabelle A.2: *Grenzwinkel für die Gelenke im Modell des menschlichen Körpers.*

die zwischen den lokalen Koordinatensystemen der Objektmodellteile bestimmt sind. Für das ausgezeichnete Objektmodellteil $omp_{0.1}$ beziehen sich die Grenzwinkel auf die Rotation zum Weltkoordinatensystem des Szenenmodells. Somit kann hiermit die Orientierung des Gesamtmodells in der Szene begrenzt werden.

Zur Überprüfung der Winkel zwischen einem Objektmodellteil omp_ν und seinem Vorgängerobjektmodellteil omp_μ werden die Rotationswinkel aus der Transformationsmatrix ${}^{omp_\mu}\mathbf{T}_{omp_\nu}$ bestimmt.[7] Hierbei sind die Rotations-Systeme entsprechend der gewählten Grenzwinkelangaben zu wählen. Bei dem Vergleich der Winkel mit den Grenzwinkeln sind die Mehrdeutigkeiten in den Rotationsmatrizen zu beachten, in dem jeweils für beide Lösungen die Winkel verglichen werden. Dies gilt insbesondere für die Rotationen um die z''-Achsen bei den Objektmodellteilen, deren Gelenkwinkel über I-Kompositionen bestimmt werden, da dort Überstreckungen in den Gelenken zwischen den ersten und zweiten Objektmodellteilen der Komposition zugelassen sind, vgl. hierzu Abschn. A.6.

In Tab. A.2 sind die Grenzwinkel für die Gelenke aufgelistet, deren Rotationen durch die Kompositionen für die Modellierung des menschlichen Körpers aus Tab. A.1 bestimmt werden. Die Werte stützen sich auf die Anatomie der Gelenke des menschlichen Körpers entsprechend [Kap92a], [Kap92c] und [Kap92b]. Entsprechend dieser Vorgaben sind auch verschiedene Rotations-Systeme für die Angabe der Grenzwinkel gewählt worden. Es werden im folgenden noch Anmerkungen zur Wahl der Grenzwinkel zu den einzelnen Gelenken im Modell des menschlichen Körpers gegeben.

Rumpf

Mit den Rotationen im Objektmodellteil des Rumpfes werden die Rotationen der Wirbelsäule abgebildet. Die Rotationen werden um die drei Achsen des Bezugskoordinatensystems angegeben, daher ist als Rotations-System xyz gewählt.

Die Rotation um die x-Achse entspricht hierbei der Ventral- und Dorsalflexion. Aufgrund der Bestimmung der Rotationen durch eine V-Komposition, bei der zusätzlich zu dem Objektmodellteil des Rumpfes die Objektmodellteile der beiden Oberarme verwendet werden, müssen hier die Flexion in der Lenden- sowie in der Brustwirbelsäule berücksichtigt werden.

Mit der Rotation um die y-Achse wird die Seitenneigung der Wirbelsäule berücksichtigt.

[7] Vgl. Anh. B.5.

Die Seitenneigung der Lenden- und der Brustwirbelsäule ist mit jeweils 20° angegeben.

Für die Drehamplituden der Wirbelsäule ist in [Kap92b] 5° für die Lendenwirbel- und 35° für die Brustwirbelsäule angegeben.

Oberschenkel

Die Rotationen im Objektmodellteil des Oberschenkels entsprechen den Rotationen im Hüftgelenk des menschlichen Körpers. Für die Rotationen wird eine Beugung / Streckung, eine Abduktion / Adduktion und eine Längsrotation in [Kap92c] angegeben. Beugung und Streckung entsprechen einer Rotation um die x-Achse des Objektmodellteils des Oberschenkels. Die Abduktion und Adduktion sind Rotationen um die y-Achse. Die Längsrotation ist wiederum eine Drehung um den Oberschenkelknochen und somit eine Drehung um die z''-Achse des lokalen Koordinatensystems des Objektmodellteils. Als Rotations-System ergibt sich daher xyz''.[8]

Die maximale Beugung und Streckung im Hüftgelenk des menschlichen Körpers ist von verschiedenen Faktoren abhängig, wie z.B. der Beugung des Knies. Daher sind in der Tab. A.2 mittlere Werte vermerkt, die entsprechend der in einer Anwendung auftretenden maximalen Werte für die Rotationen um die x-Achse angepaßt werden müssen.

Als maximaler Wert für die Adduktion nach außen wird für das Hüftgelenk des menschlichen Körpers 45° angegeben, jedoch kann durch Training, insbesondere bei Extremsportlern und Artisten durch passive Abduktion bei einem Spagat wesentlich größere Werte erreicht werden. Auch hier sind die Werte entsprechend der Anwendung anzupassen. Für die Adduktion nach innen wird ein Wert von 30° angegeben. Durch die Unterscheidung der Rotationen nach innen und außen ergeben sich für die Objektmodellteile des linken und des rechten Oberschenkels unterschiedliche Grenzwinkel.

Für die Längsrotation wird eine maximale Rotation nach außen von 60° und von 30° nach innen angegeben. Auch hier müssen für die positiven und negativen Grenzwinkel des Objektmodellteils des linken und des rechten Oberschenkels unterschieden werden.

Unterschenkel

Die Rotation im Objektmodellteil des Unterschenkels ist durch die Beugung und Streckung im Kniegelenk des menschlichen Körpers bestimmt. Als aktive Beugung wird ein Winkel von 120° angegeben, der auch als sinnvoller Grenzwinkel dienen kann. Werden anwendungsbedingt auch größere Beugungen erwartet, so muß der Grenzwinkel angepaßt werden. Entsprechend den Angaben zur I-Komposition muß eine Überstreckung zugelassen werden, so daß als Grenzwinkel für die Streckung 10° gewählt wird.

Oberarm

Für die Rotation im Gelenk des Oberarms müssen die Rotationen des eigentlichen Schultergelenks und aber auch die Bewegungen in der Schulter des menschlichen Körpers berücksichtigt werden. Durch die gewählte Modellierung wird dies jedoch in den Rotationen der Objektmodellteile der Oberarme zusammengefaßt.

Ähnlich wie bei den Objektmodellteilen der Oberschenkel werden die maximalen Rotationen für die Bewegungen der Ante- / Retroversion, Adduktion / Abduktion und einer Längsrotation angegeben. Daher ist das zu wählende Rotations-System xyz''.

[8] Das xyz''-Rotations-System kann auch als zxy- oder $yx'z''$-System aufgefaßt werden, es ergibt sich somit: $\mathcal{R}_{xyz''}(\alpha, \beta, \gamma) = \mathcal{R}_y(\beta) \cdot \mathcal{R}_x(\alpha) \cdot \mathcal{R}_z(\gamma)$, vgl. Anh. B.5.

Die Bewegungen der Ante- / Retroversion entsprechen somit einer Drehung um die x-Achse. Für die Retroversion ist eine maximale Rotation von 50° angegeben, was einem negativen Grenzwinkel von -130° in dem lokalen Koordinatensystem des Objektmodellteils des Oberarms entspricht. Für die Anteversion ergibt sich die größere Bewegung mit 180°, somit ist der positive Grenzwinkel im lokalen Koordinatensystem 0°.

Die Adduktion ist die Bewegung des Armes nach innen und somit eine Rotation um die y-Achse. Die maximale Rotation wird mit 30° angegeben. Die Drehung nach außen, die Abduktion, ist die entgegengesetzte Bewegung. Die maximale Drehung von 120° wird hier neben der maximalen Drehung von 60° im Schultergelenk durch die Beteiligung von Schulterblattbewegungen erreicht.[9] Durch die Unterscheidung von Rotationen nach innen und außen ergibt sich auch hier eine Unterscheidung für die Objektmodellteile des linken und des rechten Oberarms.

Auch mit der Längsrotation um die z''-Achse des lokalen Koordinatensystems werden Drehungen im Schultergelenk, sowie in der Schulter selbst berücksichtigt. Es ergibt sich eine maximale Außenrotation von 80° und eine maximale Innenrotation von 95°, bei der der Arm hinter dem Rücken zu liegen kommt. Kann anwendungsbedingt ausgeschlossen werden, daß der Arm hinter dem Rücken zu liegen kommt, so kann als Grenzwinkel für die Innenrotation 85° gewählt werden. Auch hier müssen die Grenzwinkel für die Objektmodellteile des linken und des rechten Oberarms unterschieden werden.

Unterarm

Die Rotation im Objektmodellteil des Unterarms beschreibt die Drehung im Ellbogengelenk des menschlichen Körpers. Aufgrund der Lage des lokalen Koordinatensystems entspricht diese Rotation einer Drehung um die x-Achse. Die maximale Beugung wird mit 145° angegeben. Eine Streckung muß auch hier, wie bei dem Objektmodellteil des Unterschenkels, zugelassen werden. Hierzu wird daher ein positiver Grenzwinkel von 10° gewählt.

[9] Eine Drehung von 180° kann durch Neigen des Rumpfes zur Gegenseite erreicht werden, dies ist jedoch durch das Objektmodellteil des Rumpfes berücksichtigt.

B Transformationen im 3D Raum

Die Lokalisation von artikularen Objekten und die Erfassung von deren Bewegung wird mit STABIL^{++} im 3D Raum vorgenommen. Daher ist eine konsequente 3D Modellierung der zu detektierenden Objekte, der Szene und der Abbildungseigenschaften der Kameras gewählt worden. Zur notwendigen Beschreibung der räumlichen Beziehungen sind für das System kartesische Koordinatensysteme definiert worden, die im folgenden nochmals zusammengefaßt aufgelistet sind. Zur Beschreibung der Beziehungen zwischen den Koordinatensystemen über Translation und Rotation eignen sich homogene Koordinaten und die entsprechenden Transformationen. Es schließt sich daher an die Beschreibung der Koordinatensysteme eine Darstellung der homogenen Koordinatentransformation an. Die Abschnitte beinhalten hierzu eine Zusammenfassung der Thematik mit Hinblick auf eine Erläuterung der Notation und die Verwendung zur Beschreibung der Transformationen bei der Modellierung und des Interpretationsprozeßes.

B.1 Koordinatensysteme

Zur 3D Detektion und Verfolgung von Objekten arbeitet STABIL^{++} mit dreidimensionalen Koordinaten. Hierzu sind kartesische Koordinatensysteme, wie ein Weltkoordinatensystem des Szenenmodells, Koordinatensysteme in den Kameras, im verwendeten Objektmodell und für jedes einzelne Objektmodellteil definiert. Alle definierten Koordinatensysteme sind Rechtssysteme.[1] Zusammenfassend wird im folgenden für alle definierten Koordinatensysteme deren Position und Orientierung angegeben:

Weltkoordinatensystem Das Weltkoordinatensystem $[X_{wcs}, Y_{wcs}, Z_{wcs}]$ ist in dem Szenenmodell definiert, vgl. Kap. 2.2. Das Koordinatensystem wird so gelegt, daß die xy-Ebene parallel zum Boden liegt; die z-Achse gibt somit die Höhe an. Im Interpretationsprozeß wird die Höhe einzelner Objektmodellteile überprüft, die durch diese Festlegung einfach möglich ist.
Die Lage des virtuellen Weltkoordinatensystems im realen Raum ist willkürlich. Es empfiehlt sich jedoch, die x- und y-Achse parallel zu den meist rechtwinkeligen Wänden von Gebäuden auszurichten.

Koordinatensystem des Objektmodells Jedem Objektmodell ist ein Koordinatensystem $[X_{obj}, Y_{obj}, Z_{obj}]$ zugeordnet, vgl. Kap. 2.3. Mit diesem Koordinatensystem wird jedoch generell nur die Position des Objektmodells in dem Weltkoordinatensystem beschrieben. Das bedeutet, daß die Orientierung willkürlich ist und mit der Orientierung des initialen

[1] Für ein Rechtssystem gilt $\det(\vec{x}, \vec{y}, \vec{z}) > 0$ mit jeweils einem Punkt auf den Achsen, z.B. $\vec{x} = [1, 0, 0]^T$, $\vec{y} = [0, 1, 0]^T$, $\vec{z} = [0, 0, 1]^T$. Ferner gilt die *Rechtehandregel*: Wenn man Daumen, Zeige- und Mittelfinger der rechten Hand so abspreizt, daß jeweils die nebeneinander liegenden Finger einen rechten Winkel ergeben, dann zeigt der Daumen in Richtung der positiven x-Achse, der Zeigefinger in Richtung der positiven y-Achse und der Mittelfinger in Richtung der positiven z-Achse.

Objektmodells festgelegt ist. Die Rotation des Objektmodells ergibt sich erst durch die Rotationen des ersten Objektmodellteils $omp_{0.1}$.

Koordinatensystem des Objektmodellteils In jedem Objektmodellteil ist entsprechend der geometrischen Struktur des Objektmodells ein lokales Koordinatensystem $[X_{omp}, Y_{omp}, Z_{omp}]$ definiert. Die Koordinatensysteme sind so positioniert, daß deren Ursprung in dem Gelenk / Knoten zum jeweiligen Vorgängerobjektmodellteil zu liegen kommt. Die z-Achse der Koordinatensysteme zeigt hierbei generell in Richtung des nachfolgenden Objektmodellteils. Für das Beispiel der Modellierung des menschlichen Körpers ist in Abb. 2.2 auf S. 30 die exakte Lage der einzelnen Koordinatensysteme dargestellt.

Kamerakoordinatensystem Entsprechend des verwendeten Kameramodells liegt der Ursprung des Kamerakoordinatensystems $[X_{cam}, Y_{cam}, Z_{cam}]$ im Brennpunkt der Lochkamera, vgl. Abb. 2.3 auf S. 45. Die z-Achse zeigt in Richtung der optischen Achse, die y-Achse nach unten und somit die x-Achse nach rechts, bei Blick in Richtung der positiven z-Achse, vgl. Abb. C.2 auf S. 214.
Um die Orientierung der Kamera bestimmen zu können, werden Schwenk-, Neige- und Rollwinkel verwendet, vgl. [SB96]. Die einzelnen Winkel haben folgenden Bezug zum Kamerakoordinatensystem:

Schwenkwinkel ψ Bei einer Drehung um die y-Achse des Kamerakoordinatensystems spricht man von einem Schwenken der Kamera um den Winkel ψ (engl. *yaw*).

Neigewinkel θ Bei einer Drehung um die x-Achse des Kamerakoordinatensystems spricht man von einem Neigen der Kamera um den Winkel θ (engl. *pitch*).

Rollwinkel ϕ Bei einer Drehung um die z-Achse des Kamerakoordinatensystems spricht man von einem Rollen der Kamera um den Winkel ϕ (engl. *roll*).

Es ist zu beachten, daß sich bei aktiven Kamerasystemen, z.B. Schwenk- / Neigekameras, diese Winkel nicht auf das Kamerakoordinatensystem beziehen. Dort sind die Winkel auf die Lage / Orientierung des beweglichen Kamerahalters zu einem Basissystem bezogen. Der Kamerahalter wird in Analogie zu Systemen in der Robotik auch als Manipulator bezeichnet, Es sind daher für die Handhabung der Schwenk- / Neigekameras noch ein Basis- und ein Manipulatorkoordinatensystem eingeführt worden, vgl. Abb. C.6 auf S. 219.

Bildkoordinatensystem Die Orientierung des 2D Bildkoordinatensystems $[X_{img}, Y_{img}]$ ergibt sich ebenfalls aus dem Kameramodell, vgl. Abb. 2.3 auf S. 45. Bei der Betrachtung des Bildes in Blickrichtung der Kamera liegt der Ursprung in der linken oberen Ecke. Die x-Achse läuft nach rechts in Richtung steigender Spaltennummern und die y-Achse läuft nach unten in Richtung steigender Zeilennummern. Entsprechend dem verwendeten Lochkameramodell verlaufen die Achsen $[X_{img}, Y_{img}]$ entgegen den Achsen X_{cam} und Y_{cam} des Kamerakoordinatensystems, denn im Brennpunkt des Linsensystems wird das Bild gespiegelt.

B.2 Translation und Rotation

Um einen Punkt $\vec{p}_1 = [p_{1_x}, p_{1_y}, p_{1_z}]^T$, der in einem Koordinatensystem cs_1 definiert ist, in einem weiteren Koordinatensystem cs_2 auszudrücken, muß eine Transformationsvorschrift zwischen

den beiden Koordinatensystemen bekannt sein. Diese Koordinatentransformation ist eine affine Abbildung aus Translation und Rotation. Man kann sich diese Koordinatentransformation auch als eine Parallelverschiebung und eine Drehung zwischen den beiden Koordinatensystemen vorstellen.

Bei der Parallelverschiebung wird ein Koordinatensystem gegenüber einem zweiten jeweils entlang der drei Achsen eines Koordinatensystems verschoben. Diese Verschiebung kann durch einen Verschiebungsvektor $\vec{t} = [t_x, t_y, t_z]^T$ ausgedrückt werden, der die beiden Ursprünge der Koordinatensysteme verbindet. Somit gilt für eine Translation:

$$\vec{p}_{cs_2} = \vec{p}_{cs_1} + \vec{t}$$

Bei der Drehung werden jeweils drei Winkel angegeben. Ein Drehwinkel ist dabei immer einer Achse des Koordinatensystems zugeordnet. Die Winkel sind in mathematisch positiver Richtung definiert. Mit der *Rechtsschraubenregel* kann man sich dies verdeutlichen: Wenn man in Richtung der positiven Drehachse blickt und eine Rechtsschraube, z.B. einen Korkenzieher einschraubt, dann entspricht diese Bewegung der Drehung der beiden anderen Achsen.

Die Drehung von Koordinatensystemen läßt sich durch Rotationsmatrizen einfach ausdrücken. Für die Rotationen um die drei Achsen x, y, z gelten die elementaren Rotationsmatrizen $\mathcal{R}_x(\alpha)$, $\mathcal{R}_y(\beta)$ und $\mathcal{R}_z(\gamma)$. Diese sind wie folgt definiert:

$$\mathcal{R}_x(\alpha) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha & -\sin\alpha \\ 0 & \sin\alpha & \cos\alpha \end{pmatrix} \tag{B.1}$$

$$\mathcal{R}_y(\beta) = \begin{pmatrix} \cos\beta & 0 & \sin\beta \\ 0 & 1 & 0 \\ -\sin\beta & 0 & \cos\beta \end{pmatrix} \tag{B.2}$$

$$\mathcal{R}_z(\gamma) = \begin{pmatrix} \cos\gamma & -\sin\gamma & 0 \\ \sin\gamma & \cos\gamma & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{B.3}$$

Beispielhaft gilt für die Anwendung der Rotationsmatrix $\mathcal{R}_x(\alpha)$ auf einen Punkt, der im Koordinatensystem cs_1 definiert ist:

$$\vec{p}_{cs_2} = \mathcal{R}_x(\alpha) \cdot \vec{p}_{cs_1}$$

Die drei einzelnen Rotationen lassen sich durch Matrixmultiplikationen zu einer Rotationsmatrix $\mathcal{R}_{3\times3}$[2] kombinieren. Hierbei ist jedoch die Reihenfolge der Multiplikation und somit die Drehreihenfolge zu beachten. Auf diese Problematik wird in Abschn. B.5 eingegangen. Eine Herleitung der allgemeinen Rotationsmatrix $\mathcal{R}_{3\times3}$ findet sich in [SB96]. Entsprechend der dort aufgeführten Definition der Rotationsmatrizen über die Einheitsvektoren in den beiden Koordinatensystemen kann die Rotationsmatrix zwischen zwei Koordinatensystemen bestimmt werden, wenn man jeweils einen Punkt auf einer der drei Koordinatensystemachsen kennt. Nehmen wir an, daß die Punkte im Koordinatensystem cs_2 bekannt sind und der Punkt auf der x-Achse mit $\vec{x}_{cs_1}$, der Punkt auf der y-Achse mit $\vec{y}_{cs_1}$ und der Punkt auf der z-Achse mit $\vec{z}_{cs_1}$ bezeichnet

[2] Der Index "3×3" kennzeichnet die Rotationsmatrix als $\in \mathbb{R}^{3\times3}$, im Gegensatz zu homogenen Rotationsmatrizen $\mathcal{R} \in \mathbb{R}^{4\times4}$; Für die elementaren Rotationsmatrizen gilt: $\mathcal{R}_z(\alpha), \mathcal{R}_z(\beta), \mathcal{R}_z(\gamma) \in \mathbb{R}^{3\times3}$.

sind, so gilt:

$$\frac{\vec{x}_{cs_1}}{|\vec{x}_{cs_1}|} = \mathcal{R}_{3\times3} \cdot \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \tag{B.4}$$

$$\frac{\vec{y}_{cs_1}}{|\vec{y}_{cs_1}|} = \mathcal{R}_{3\times3} \cdot \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \tag{B.5}$$

$$\frac{\vec{z}_{cs_1}}{|\vec{z}_{cs_1}|} = \mathcal{R}_{3\times3} \cdot \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \tag{B.6}$$

Hieraus ergibt sich direkt für die Elemente der Rotationsmatrix $\mathcal{R}_{3\times3}$:

$$\begin{pmatrix} r_{11} \\ r_{21} \\ r_{31} \end{pmatrix} = \frac{\vec{x}_{cs_1}}{|\vec{x}_{cs_1}|} \tag{B.7}$$

$$\begin{pmatrix} r_{12} \\ r_{22} \\ r_{32} \end{pmatrix} = \frac{\vec{y}_{cs_1}}{|\vec{y}_{cs_1}|} \tag{B.8}$$

$$\begin{pmatrix} r_{13} \\ r_{23} \\ r_{33} \end{pmatrix} = \frac{\vec{z}_{cs_1}}{|\vec{z}_{cs_1}|} \tag{B.9}$$

Es sei noch erwähnt, daß aufgrund ihrer Eigenschaften die invertierten Rotationsmatrizen den transponierten Rotationsmatrizen entsprechen, somit kann die Rotation durch ein einfaches Vertauschen der Matrixelemente umgekehrt werden:

$$\mathcal{R}_{3\times3}^{-1} = \begin{pmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{pmatrix}^{-1} = \mathcal{R}_{3\times3}^{T} = \begin{pmatrix} r_{11} & r_{21} & r_{31} \\ r_{12} & r_{22} & r_{32} \\ r_{13} & r_{23} & r_{33} \end{pmatrix} \tag{B.10}$$

B.3 Homogene Koordinatentransformation

Zur einfachen Handhabung der Transformationen, die Translation und Rotation von Koordinatensystemen vereinen, kann man homogene Koordinaten und entsprechend homogene Transformationsmatrizen einführen. Hierzu sind die Vektoren der 3D Punkte um eine Zeile zu erweitern. In der neuen Zeile steht ein Skalierungsfaktor, der jedoch hier auf eins gesetzt wird. Man erhält somit einen Vektor $\vec{p} \in \mathbb{R}^4$:

$$\vec{p} = \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix}$$

Dementsprechend werden die Rotationsmatrizen $\mathcal{R}_{3\times3}$ um eine Zeile und Spalte erweitert. Man erhält somit Matrizen $\mathcal{R} \in \mathbb{R}^{4\times4}$:

$$\mathcal{R} = \begin{pmatrix} r_{11} & r_{12} & r_{13} & 0 \\ r_{21} & r_{22} & r_{23} & 0 \\ r_{31} & r_{32} & r_{33} & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Um eine Translation auch als homogene Matrixmultiplikation ausdrücken zu können, wird aus dem Translationsvektor $\vec{t} = [x, y, z]^T$ eine Translationsmatrix $\mathcal{T} \in \mathbb{R}^{4\times4}$, entsprechend

$$\mathcal{T} = \begin{pmatrix} 1 & 0 & 0 & x \\ 0 & 1 & 0 & y \\ 0 & 0 & 1 & z \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

gebildet.

Die Transformation aus Rotation und Translation kann somit zu einer homogenen Transformationsmatrix $\mathbf{T} \in \mathbb{R}^{4\times4}$

$$\mathbf{T} = \begin{pmatrix} t_{11} & t_{12} & t_{13} & t_{14} \\ t_{21} & t_{22} & t_{23} & t_{24} \\ t_{31} & t_{32} & t_{33} & t_{34} \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

zusammengefaßt werden. Der eigentliche Aufbau der Transformationsmatrix durch Matrixmultiplikation von Rotationsmatrix und Translationsmatrix wird im Abschn. B.4 unter Beachtung der Reihenfolge der beiden Operatoren erläutert.

Für die Anwendung der homogenen Transformationsmatrix auf einen Punkt $\vec{p}_1$ gilt:

$$\begin{pmatrix} x_2 \\ y_2 \\ z_2 \\ 1 \end{pmatrix} = \begin{pmatrix} t_{11} & t_{12} & t_{13} & t_{14} \\ t_{21} & t_{22} & t_{23} & t_{24} \\ t_{31} & t_{32} & t_{33} & t_{34} \\ 0 & 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ y_1 \\ z_1 \\ 1 \end{pmatrix}$$

Aufgrund der Eigenschaften der homogenen Transformationsmatrix läßt sich die Transformation in

$$\begin{pmatrix} x_2 \\ y_2 \\ z_2 \end{pmatrix} = \begin{pmatrix} t_{11} & t_{12} & t_{13} \\ t_{21} & t_{22} & t_{23} \\ t_{31} & t_{32} & t_{33} \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ y_1 \\ z_1 \end{pmatrix} + \begin{pmatrix} t_{14} \\ t_{24} \\ t_{34} \end{pmatrix} \tag{B.11}$$

zerlegen. Es liegt nahe, dies als

$$\vec{p}_{base} = \mathcal{R}_{3\times3} \cdot \vec{p}_{trans} + \vec{t} \tag{B.12}$$

zu deuten[3], womit angenommen werden kann, daß bei der Transformation des Basiskoordinatensystems in das transformierte Koordinatensystem immer zuerst die Translation ausgeführt wird. Beim Aufbau einer homogenen Transformationsmatrix aus Rotationsmatrix und Translationsmatrix kann die Reihenfolge jedoch auch umgekehrt festgelegt werden, vgl. Abschn. B.5.

Weiterhin ist zu beachten, welche Transformationsrichtung durch die Transformationsmatrix beschrieben wird. Im allgemeinen wird man durch eine Translation und Rotation beschreiben, wie ein Basis- oder Ausgangskoordinatensystem cs_{base} in ein weiteres, transformiertes Koordinatensystem cs_{trans} überführt wird. Hierdurch ist eine Richtung der Transformation festgelegt. Bei der Anwendung der entsprechenden Transformationsmatrix wird ein Punkt, der im transformierten Koordinatensystem cs_{trans} definiert ist, in das Basiskoordinatensystem cs_{base} transformiert, es gilt:

$$\vec{p}_{base} = \mathbf{T} \cdot \vec{p}_{trans} \tag{B.13}$$

[3] Hier sind $\vec{p}_{base}, \vec{p}_{trans} \in \mathbb{R}^3$.

Um die Richtung der Koordinatentransformation, die eine Transformationsmatrix beschreibt, eindeutig zu kennzeichnen, hat sich eine in [HLZ97] verwendete Notation als nützlich erwiesen. Hierbei wird eine Kennung des Ausgangskoordinatensystems dem Symbol der Transformationsmatrix vorangestellt und eine Kennung des transformierten Koordinatensystems als Index angehängt. In der Formel der Transformation eines Punktes in Glg. B.13 treffen dann jeweils die Kennzeichnungen der Koordinatensysteme von Punkten und Transformationsmatrizen aufeinander:

$$\vec{p}_{base} = {}^{base}\mathbf{T}_{trans} \cdot \vec{p}_{trans}$$

Oftmals soll jedoch ein Punkt, der im Ausgangskoordinatensystem definiert ist, in das transformierte Koordinatensystem überführt werden. Entsprechend der eingeführten Notation gilt dann:

$$\begin{aligned} &\vec{p}_{trans} = {}^{trans}\mathbf{T}_{base} \cdot \vec{p}_{base} \\ &\text{mit} \\ &{}^{trans}\mathbf{T}_{base} = {}^{base}\mathbf{T}_{trans}^{-1} \end{aligned}$$

${}^{trans}\mathbf{T}_{base}$ ist somit die Inverse der Transformationsmatrix ${}^{base}\mathbf{T}_{trans}$. Entsprechend der umgekehrten Transformation beschreibt die Rotation und die Translation von ${}^{trans}\mathbf{T}_{base}$ die Überführung des transformierten Koordinatensystems cs_{trans} in das Ausgangs-/ Basiskoordinatensystem cs_{base}.

Sind mehrere kaskadierte Transformationen, wie z.B.

$$\begin{aligned} \vec{p}_{cs_1} &= {}^{cs_1}\mathbf{T}_{cs_2} \cdot \vec{p}_{cs_2} \\ \vec{p}_{cs_2} &= {}^{cs_2}\mathbf{T}_{cs_3} \cdot \vec{p}_{cs_3} \\ &\dots \\ \vec{p}_{cs_{n-1}} &= {}^{cs_{n-1}}\mathbf{T}_{cs_n} \cdot \vec{p}_{cs_n} \end{aligned} \tag{B.14}$$

gegeben, so ergibt sich für die Transformation eines Punktes, der im Koordinatensystem cs_n definiert ist, in das Koordinatensystem cs_1:

$$\begin{aligned} &\vec{p}_{cs_1} = {}^{cs_1}\mathbf{T}_{cs_n} \cdot \vec{p}_{cs_n} \\ &\text{mit} \\ &{}^{cs_1}\mathbf{T}_{cs_n} = {}^{cs_1}\mathbf{T}_{cs_2} \cdot {}^{cs_2}\mathbf{T}_{cs_3} \cdots {}^{cs_{n-2}}\mathbf{T}_{cs_{n-1}} \cdot {}^{cs_{n-1}}\mathbf{T}_{cs_n}. \end{aligned} \tag{B.15}$$

Durch die verwendete Kennzeichnung der Transformationen kann die Reihenfolge der Matrixmultiplikationen einfach nachvollzogen werden. Wie bei der Transformation eines Punktes treffen hier immer passende Kennzeichnungen aufeinander.

In den folgenden Abschnitten wird der Aufbau der Transformationsmatrizen beschrieben. Hierbei wird nach der Reihenfolge der Transformation und Rotationssystemen unterschieden.

B.4 Transformationsreihenfolge

Bei der Beschreibung einer Transformation von Koordinatensystemen aus Rotation und Translation ist die Reihenfolge der beiden Operationen entscheidend, d.h. ist die angegebene Translation auf das Basiskoordinatensystem cs_{base} bezogen oder auf ein Koordinatensystem cs_{rot}, das nach der zuerst ausgeführten Rotation entsteht. Beim Aufbau der homogenen Transformationsmatrix ist dies durch die Richtung der Matrixmultiplikation von Rotationsmatrix $\mathcal{R}$ und

Translationsmatrix $\mathcal{T}$ zu berücksichtigen. Es ist zwischen $\mathbf{T}_{(\mathcal{TR})} = \mathcal{T} \cdot \mathcal{R}$ und $\mathbf{T}_{(\mathcal{RT})} = \mathcal{R} \cdot \mathcal{T}$ zu unterscheiden. Dabei ist zu beachten, daß eine homogene Transformationsmatrix immer, und somit unabhängig von der Richtung der Matrixmultiplikation beim Erzeugen, entsprechend der Glg. B.11 und B.12 auf einen Punkt angewendet wird. Im folgenden wird nur betrachtet, wie eine Transformationsmatrix aufgebaut ist und wie sie gedeutet werden kann.

Bei $\mathbf{T}_{(\mathcal{TR})}$ beziehen sich die Angaben der Translation (x, y, z) auf das Basiskoordinatensystem cs_{base}. Man kann es sich so vorstellen, daß bei der Transformation des Basiskoordinatensystems in das transformierte Koordinatensystem zunächst die drei Translationen entlang der Achsen des Basiskoordinatensystems durchgeführt werden und dann auf dem verschobenen Koordinatensystem die Rotation ausgeführt wird. Es ergibt sich:

$$\mathbf{T}_{(\mathcal{TR})} = \mathcal{T} \cdot \mathcal{R} = \begin{pmatrix} r_{11} & r_{12} & r_{13} & x \\ r_{21} & r_{22} & r_{23} & y \\ r_{31} & r_{32} & r_{33} & z \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Mit diesem Aufbau der homogenen Transformationsmatrix wird die Transformation eines Punktes aus dem transformierten Koordinatensystem in das Basiskoordinatensystem entsprechend

$$\begin{aligned} \vec{p}_{base} &= \mathbf{T}_{(\mathcal{TR})} \cdot \vec{p}_{trans} \\ &= (\mathcal{R}_{3\times 3} \cdot \vec{p}_{trans}) + \vec{t} \end{aligned} \tag{B.16}$$

ausgeführt.

Bei $\mathbf{T}_{(\mathcal{RT})}$ wird zunächst das Basiskoordinatensystem cs_{base} durch die Rotation in ein Koordinatensystem cs_{rot} transformiert. Anschließend wird entlang der Achsen des neuen Koordinatensystems cs_{rot} die Translation (x, y, z) angewendet. Es ergibt sich:

$$\mathbf{T}_{(\mathcal{RT})} = \mathcal{R} \cdot \mathcal{T} = \begin{pmatrix} r_{11} & r_{12} & r_{13} & r_{11} \cdot x + r_{12} \cdot y + r_{13} \cdot z \\ r_{21} & r_{22} & r_{23} & r_{21} \cdot x + r_{22} \cdot y + r_{23} \cdot z \\ r_{31} & r_{32} & r_{33} & r_{31} \cdot x + r_{32} \cdot y + r_{33} \cdot z \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Mit diesem Aufbau der homogenen Transformationsmatrix wird die Transformation eines Punktes aus dem transformierten Koordinatensystem in das Basiskoordinatensystem entsprechend

$$\begin{aligned} \vec{p}_{base} &= \mathbf{T}_{(\mathcal{RT})} \cdot \vec{p}_{trans} \\ &= \mathcal{R}_{3\times 3} \cdot \left(\vec{p}_{trans} + \vec{t}\right) \end{aligned} \tag{B.17}$$

ausgeführt.

Aufgrund der Eigenschaften der homogenen Transformationsmatrix lassen sich die inversen Transformationsmatrizen einfach bilden. Für $\mathbf{T}^{-1}_{(\mathcal{TR})}$ gilt:

$$\begin{aligned} \mathbf{T}^{-1}_{(\mathcal{TR})} &= (\mathcal{T} \cdot \mathcal{R})^{-1} = \mathcal{R}^{-1} \cdot \mathcal{T}^{-1} \\ &= \begin{pmatrix} r_{11} & r_{21} & r_{31} & -r_{11} \cdot x - r_{21} \cdot y - r_{31} \cdot z \\ r_{12} & r_{22} & r_{32} & -r_{12} \cdot x - r_{22} \cdot y - r_{32} \cdot z \\ r_{13} & r_{23} & r_{33} & -r_{13} \cdot x - r_{23} \cdot y - r_{33} \cdot z \\ 0 & 0 & 0 & 1 \end{pmatrix} \end{aligned} \tag{B.18}$$

Für die Anwendung der inversen Transformation auf einen Punkt, der im Basiskoordinatensystem definiert ist, gilt:

$$\begin{aligned} \vec{p}_{trans} &= \mathbf{T}^{-1}_{(\mathcal{TR})} \cdot \vec{p}_{base} \\ &= \mathcal{R}^{-1}_{3\times 3} \cdot \left(\vec{p}_{base} - \vec{t}\right) \end{aligned} \tag{B.19}$$

Entsprechend gilt für invertierte Transformation bei umgekehrter Reihenfolge von Rotation und Translation:

$$\begin{aligned} \mathbf{T}^{-1}_{(\mathcal{RT})} &= (\mathcal{R}\cdot\mathcal{T})^{-1} = \mathcal{T}^{-1}\cdot\mathcal{R}^{-1} \\ &= \begin{pmatrix} r_{11} & r_{21} & r_{31} & -x \\ r_{12} & r_{22} & r_{32} & -y \\ r_{13} & r_{23} & r_{33} & -z \\ 0 & 0 & 0 & 1 \end{pmatrix} \end{aligned} \tag{B.20}$$

Für die Anwendung der inversen Transformation auf einen Punkt, der im Basiskoordinatensystem definiert ist, gilt:

$$\begin{aligned} \vec{p}_{trans} &= \mathbf{T}^{-1}_{(\mathcal{RT})} \cdot \vec{p}_{base} \\ &= \left(\mathcal{R}^{-1}_{3\times 3} \cdot \vec{p}_{base}\right) - \vec{t} \end{aligned} \tag{B.21}$$

Die Ähnlichkeiten der Glg. B.17 und B.19 und der Glg. B.16 und B.21 lassen sich einfach erklären: Durch die Umkehrung der Transformation, also bei der Transformation des transformierten Koordinatensystems cs_{trans} in das Ausgangskoordinatensystem cs_{base}, wird zum einen die Rotation umgekehrt, zum anderen aber auch die Reihenfolge von Rotation und Translation vertauscht. Die Reihenfolge und auch die Umkehrung der Rotationen bei der Transformation von Koordinatensystemen wird im folgenden Abschnitt erläutert.

B.5 Rotationsreihenfolge

Die Rotation von Koordinatensystemen kann man mittels Matrixmultiplikation aus den elementaren Rotationen aus Glg. B.1, B.2 und B.3 zusammensetzen, so daß man eine Rotationsmatrix $\mathcal{R}$ enthält.[4] Bei komplexen Drehbewegungen, die sich aus mehr als einer Drehung zusammensetzen, werden mehrere elementare Rotationen kombiniert. So können elementare Drehungen um die x-, y- und z-Achse z.B. wie folgt kombiniert werden:

$$\begin{aligned} \mathcal{R}_{(1)}(\alpha,\beta,\gamma) &= \mathcal{R}_x(\alpha)\cdot\mathcal{R}_y(\beta)\cdot\mathcal{R}_z(\gamma) \\ \mathcal{R}_{(2)}(\alpha',\beta',\gamma') &= \mathcal{R}_z(\gamma')\cdot\mathcal{R}_y(\beta')\cdot\mathcal{R}_x(\alpha') \end{aligned}$$

Kennzeichnung / Notation

Entsprechend den Multiplikationsregeln in [SB96] bedeutet ein Anfügen einer elementaren Rotationsmatrix rechts an das Matrixprodukt d.h. bei einer Multiplikation von rechts, daß sich die hinzugefügte Rotation auf das schon gedrehte Koordinatensystem bezieht. Das Anfügen einer elementaren Rotationsmatrix von links an das Matrixprodukt, d.h. die Multiplikation von links bedeutet, daß die hinzugefügte Rotation sich auf das Ausgangskoordinatensystem bezieht.

[4] In diesem Abschnitt werden nur Rotationsmatrizen $\in \mathbb{R}^{3\times 3}$ verwendet, daher wird auf die Kennzeichnung $_{3\times 3}$ der Rotationsmatrizen im Index verzichtet.

Dementsprechend kann die durch die Rotationsmatrix $\mathcal{R}_{(1)}(\alpha, \beta, \gamma)$ beschriebene Rotation des Basiskoordinatensystems in das transformierte Koordinatensystem von rechts und von links gedeutet werden. Zum einen wird eine Drehung beschrieben, bei der zunächst eine Drehung mit dem Winkel γ um die z-Achse des Basiskoordinatensystems ausgeführt wird. Daran schließt sich eine Drehung mit dem Winkel β um die y-Achse des Ausgangskoordinatensystems, also dem Basiskoordinatensystem,an. Die dritte Drehung wird um die x-Achse des Basiskoordinatensystems mit dem Winkel α ausgeführt.

Zum anderen kann die Rotationsmatrix $\mathcal{R}_{(1)}(\alpha, \beta, \gamma)$ als eine Rotation des Basiskoordinatensystems in das transformierte Koordinatensystem aufgefaßt werden, bei der zunächst eine Drehung mit dem Winkel α um die x-Achse des Basiskoordinatensystems durchgeführt wird. Daran schließt sich eine Drehung mit dem Winkel β um die neue y-Achse des gedrehten Koordinatensystems an. Diese Drehachse wird mit y' bezeichnet. Die dritte Drehung wird mit dem Winkel γ um die neue z-Achse des bisher zweimal gedrehten Koordinatensystems ausgeführt. Diese Drehachse wird mit z'' bezeichnet.

Um die Drehreihenfolge und die Bezugskoordinatensysteme der Drehwinkel zu kennzeichnen, werden die Rotationsmatrizen entsprechend der in [SB96] verwendeten Notation im Index mit den Bezeichnungen der Rotationsachsen gekennzeichnet. Für die Rotationsmatrix $\mathcal{R}_{(1)}(\alpha, \beta, \gamma)$ gilt dann entsprechend der ersten Deutung

$$\mathcal{R}_{(1)}(\alpha, \beta, \gamma) = \mathcal{R}_{(zyx)}(\alpha, \beta, \gamma)$$

und entsprechend der zweiten Deutung

$$\mathcal{R}_{(1)}(\alpha, \beta, \gamma) = \mathcal{R}_{(xy'z'')}(\alpha, \beta, \gamma)$$

Entsprechend gilt für $\mathcal{R}_{(2)}(\alpha', \beta', \gamma')$

$$\begin{aligned} \mathcal{R}_{(2)}(\alpha', \beta', \gamma') &= \mathcal{R}_{(xyz)}(\alpha', \beta', \gamma') \\ &= \mathcal{R}_{(zy'x'')}(\alpha', \beta', \gamma') \end{aligned}$$

Durch entsprechende Kombination von elementaren Rotationen lassen sich die verschiedensten Drehsysteme beschreiben. In der Robotik werden üblicherweise die Drehachsen entsprechend der realen Achsen der Manipulatoren gewählt. Somit muß sich ein Drehsystem nicht zwangsweise aus jeweils einer der drei Elementarrotationen $\mathcal{R}_x(\alpha)$, $\mathcal{R}_y(\beta)$ und $\mathcal{R}_z(\gamma)$ zusammensetzen. Vielmehr kann auch mehrmals die gleiche Elementarrotation verwendet werden, wie auch bei der aus der Mathematik bekannten Beschreibung der Rotation zwischen Koordinatensystemen durch die Eulerwinkel:

$$\mathcal{R}_{(Euler)}(\alpha, \gamma_1, \gamma_2) = \mathcal{R}_z(\gamma_1) \cdot \mathcal{R}_x(\alpha) \cdot \mathcal{R}_z(\gamma_2)$$

Entsprechend der verwendeten Notation für die Drehreihenfolge und der Bezugssysteme der Drehwinkel ergibt sich direkt:

$$\begin{aligned} \mathcal{R}_{(Euler)}(\alpha, \gamma_1, \gamma_2) &= \mathcal{R}_{(zxz)}(\alpha, \gamma_1, \gamma_2) \\ &= \mathcal{R}_{(zx'z'')}(\alpha, \gamma_1, \gamma_2) \end{aligned}$$

Jedoch kann man die Rotation des Basiskoordinatensystems in das transformierte Koordinatensystem auch wie folgt auffassen: Zunächst eine Drehung mit dem Winkel α um die x-Achse des Basiskoordinatensystems, gefolgt von einer Drehung mit dem Winkel γ_1 um die z-Achse des Basiskoordinatensystems. Daran schließt sich eine Drehung mit dem Winkel γ_2 um die neue z-Achse des bisher zweimal gedrehten Koordinatensystems (z'') an. Somit kann diese Rotation neben den Bezeichnungen zxz und $zx'z''$ auch als xzz''-Rotation bezeichnet werden.

Aufbau der Rotationsmatrizen

Im folgenden werden der Aufbau und die einzelnen Elemente der Rotationsmatrizen für die drei beschriebenen Rotations-Systeme dargestellt.
Für das zyz-/$xy'z''$-System gilt:

$$\begin{aligned}
\mathcal{R}_{(zyx)}(\alpha,\beta,\gamma) &= \\
\mathcal{R}_{(xy'z'')}(\alpha,\beta,\gamma) &= \mathcal{R}_x(\alpha)\cdot\mathcal{R}_y(\beta)\cdot\mathcal{R}_z(\gamma) \\
&= \begin{pmatrix} r_{11} & r_{21} & r_{31} \\ r_{12} & r_{22} & r_{32} \\ r_{13} & r_{23} & r_{33} \end{pmatrix}
\end{aligned}$$

mit

$$\begin{aligned}
r_{11} &= \cos\beta\cos\gamma \\
r_{12} &= -\cos\beta\sin\gamma \\
r_{13} &= \sin\beta \\
r_{21} &= \sin\alpha\sin\beta\cos\gamma + \cos\alpha\sin\gamma \\
r_{22} &= -\sin\alpha\sin\beta\sin\gamma + \cos\alpha\cos\gamma \\
r_{23} &= -\sin\alpha\cos\beta \\
r_{31} &= -\cos\alpha\sin\beta\cos\gamma + \sin\alpha\sin\gamma \\
r_{32} &= \cos\alpha\sin\beta\sin\gamma + \sin\alpha\cos\gamma \\
r_{33} &= \cos\alpha\cos\beta
\end{aligned}$$

Für das $zy'x''$-/xyz-System gilt:

$$\begin{aligned}
\mathcal{R}_{(zy'x'')}(\alpha,\beta,\gamma) &= \\
\mathcal{R}_{(xyz)}(\alpha,\beta,\gamma) &= \mathcal{R}_z(\gamma)\cdot\mathcal{R}_y(\beta)\cdot\mathcal{R}_x(\alpha) \\
&= \begin{pmatrix} r_{11} & r_{21} & r_{31} \\ r_{12} & r_{22} & r_{32} \\ r_{13} & r_{23} & r_{33} \end{pmatrix}
\end{aligned}$$

mit

$$\begin{aligned}
r_{11} &= \cos\gamma\cos\beta \\
r_{12} &= \cos\gamma\sin\beta\sin\alpha - \sin\gamma\cos\alpha \\
r_{13} &= \cos\gamma\sin\beta\cos\alpha + \sin\gamma\sin\alpha \\
r_{21} &= \sin\gamma\cos\beta \\
r_{22} &= \sin\gamma\sin\beta\sin\alpha + \cos\gamma\cos\alpha \\
r_{23} &= \sin\gamma\sin\beta\cos\alpha - \cos\gamma\sin\alpha \\
r_{31} &= -\sin\beta \\
r_{32} &= \cos\beta\sin\alpha \\
r_{33} &= \cos\beta\cos\alpha
\end{aligned}$$

Für das zxz-/$zx'z''$-/xzz''-System gilt:

$$\begin{aligned}
\mathcal{R}_{(zxz)}(\alpha,\gamma_1,\gamma_2) &= \\
\mathcal{R}_{(zx'z'')}(\alpha,\gamma_1,\gamma_2) &= \\
\mathcal{R}_{(xzz'')}(\alpha,\gamma_1,\gamma_2) &= \mathcal{R}_z(\gamma_1)\cdot\mathcal{R}_x(\alpha)\cdot\mathcal{R}_z(\gamma_2) \\
&= \begin{pmatrix} r_{11} & r_{21} & r_{31} \\ r_{12} & r_{22} & r_{32} \\ r_{13} & r_{23} & r_{33} \end{pmatrix}
\end{aligned}$$

mit

$$\begin{aligned}
r_{11} &= -\cos\alpha\sin\gamma_1\sin\gamma_2 + \cos\gamma_1\cos\gamma_2 \\
r_{12} &= -\cos\alpha\sin\gamma_1\cos\gamma_2 - \cos\gamma_1\sin\gamma_2 \\
r_{13} &= \sin\alpha\sin\gamma_1 \\
r_{21} &= \cos\alpha\cos\gamma_1\sin\gamma_2 + \sin\gamma_1\cos\gamma_2 \\
r_{22} &= \cos\alpha\cos\gamma_1\cos\gamma_2 - \sin\gamma_1\sin\gamma_2 \\
r_{23} &= -\sin\alpha\cos\gamma_1 \\
r_{31} &= \sin\alpha\sin\gamma_2 \\
r_{32} &= \sin\alpha\cos\gamma_2 \\
r_{33} &= \cos\alpha
\end{aligned}$$

Zusammenhang zwischen den Rotationsmatrizen

Für eine Umrechnung der Rotationswinkel[5] aus dem zyx-System in das $zy'x''$-System und umgekehrt, kann man ausnutzen, daß[6]

$$\begin{aligned}
\sin(-\theta) &= -\sin(\theta) \\
\cos(-\theta) &= \cos(\theta)
\end{aligned}$$

Damit ergibt sich bei Ausnutzung der inversen Rotationsmatrix:

$$\mathcal{R}_{(zy'x'')}(\alpha,\beta,\gamma) = \mathcal{R}^{-1}_{(zyx)}(-\alpha,-\beta,-\gamma)$$

Bestimmung der Drehwinkel

Um die Drehwinkel aus den Rotationsmatrizen herauszulesen, ist zu beachten, daß es dabei jeweils eine Mehrdeutigkeit gibt und jede Repräsentation einer Rotation mit drei Parametern eine Singularität aufweist. Zur Vermeidung der Singularität kann man die Rotationsmatrix minimal stören. Diese Störung kann in den meisten Fällen toleriert werden.

Nur mit entsprechenden Heuristiken / Winkelrestriktionen kann entschieden werden, welche der beiden Lösungen bei den Mehrdeutigkeiten im entsprechenden Winkelbereich liegt und ob die Rotationsmatrix im Falle der Singularität mit einer Rotation um einen positiven oder negativen Winkel gestört wird. Die Mehrdeutigkeiten ergeben sich aus folgenden Eigenschaften der Kreisfunktionen:

$$\begin{aligned}
\sin(\theta) &= \sin(180^\circ - \theta) \\
-\sin(\theta) &= \sin(180^\circ + \theta) \\
-\cos(\theta) &= \sin(180^\circ \pm \theta)
\end{aligned}$$

Für die Bestimmung der einer Rotationsmatrix zugrundeliegenden Drehwinkel im zyx-(/$xy'z''$)-System aus $\mathcal{R}_{(zyx)}(\alpha,\beta,\gamma)$ gilt:

$$\begin{aligned}
\cos\beta &= \sqrt{r_{11}^2 + r_{12}^2} \\
\alpha &= atan2\left(-r_{23}, r_{33}\right) \\
\beta &= atan2\left(r_{13}, \cos\beta\right) \\
\gamma &= atan2\left(-r_{12}, r_{11}\right)
\end{aligned} \tag{B.22}$$

[5] Diese Umrechnung der Rotationswinkel wird auch als Basiswechsel bezeichnet.

[6] Das Gleiche gilt auch für Basiswechsel zwischen zxy- / $yx'z''$-Systemen und xzy- / $yz'x''$-Systemen.

Die Singularität tritt hier bei $\cos\beta = 0$ auf, so daß β entweder $90°$ oder $-90°$ sein kann. Zum Stören der Matrix wird diese mit einer Rotationsmatrix um die y-Achse multipliziert. Für die Mehrdeutigkeit gilt:

$$\mathcal{R}_{(zyx)}(\alpha,\beta,\gamma) = \mathcal{R}_{(zyz)}(180° + \alpha, 180° - \beta, 180° + \gamma)$$

Für die Bestimmung der Drehwinkel im xyz-(/$zy'x''$)-System aus $\mathcal{R}_{(xyz)}(\alpha,\beta,\gamma)$ gilt:

$$\begin{aligned} \cos\beta &= \sqrt{r_{11}^2 + r_{21}^2} \\ \gamma &= atan2\,(r_{21}, r_{11}) \\ \beta &= atan2\,(-r_{31}, \cos\beta) \\ \alpha &= atan2\,(r_{32}, r_{33}) \end{aligned} \tag{B.23}$$

Die Singularität tritt auch hier bei $\cos\beta = 0$ auf, so daß β entweder $90°$ oder $-90°$ sein kann. Zum Stören der Matrix wird diese mit einer Rotationsmatrix um die y-Achse multipliziert. Für die Mehrdeutigkeit gilt auch hier:

$$\mathcal{R}_{(xyz)}(\alpha,\beta,\gamma) = \mathcal{R}_{(xyz)}(180° + \alpha, 180° - \beta, 180° + \gamma)$$

Für die Bestimmung der Drehwinkel im zxz-(/$zx'z''$-/xzz'')-System aus $\mathcal{R}_{(zxz)}(\alpha,\gamma_1,\gamma_2)$ gilt:

$$\begin{aligned} \sin\alpha &= \sqrt{r_{13}^2 + r_{23}^2} \\ \alpha &= atan2\,(\sin\alpha, r_{33}) \\ \gamma_1 &= atan2\,(r_{13}, -r_{23}) \\ \gamma_2 &= atan2\,(r_{31}, r_{32}) \end{aligned} \tag{B.24}$$

Die Singularität tritt hier bei $\sin\alpha = 0$ auf, so daß α entweder $0°$ oder $180°$ sein kann. Zum Stören der Matrix wird diese mit einer Rotationsmatrix um die x-Achse multipliziert. Für die Mehrdeutigkeit gilt hier:

$$\mathcal{R}_{(zxz)}(\alpha,\gamma_1,\gamma_2) = \mathcal{R}_{(zxz)}(-\alpha, 180° + \gamma_1, 180° + \gamma_2)$$

Um die Mehrdeutigkeit bei der Angabe der Euler-Winkel / Rotationen im zxz-System zu vermeiden, gilt entsprechend der Definition der Euler-Winkel für den Winkel $\alpha \in [0°, 180°]$.

Bei den Umrechnungen in Glg. B.22, B.23 und B.24 entspricht $atan2(y,x)$ der Funktion $arctan\left(\frac{y}{x}\right)$ unter Berücksichtigung des richtigen Quadranten.

C Kamerakalibrierung

C.1 Grundlagen

Um die Abbildungseigenschaften der in STABIL^{++} verwendeten Kameras zu modellieren, wird ein Lochkameramodell mit radialer Verzerrung verwendet. Dieses Modell ist in Kap. 2.5.4 beschrieben. Die Parameter des Modells teilen sich in die sog. inneren und äußeren Kameraparameter. Die inneren Kameraparameter $camPar$ beschreiben die eigentlichen Abbildungseigenschaften, wobei mit den äußeren Kameraparametern $camPose$ die Lage (Position und Orientierung) der Kamera in dem Weltkoordinatensystem bestimmt ist, vgl. Glg. 2.15 und Glg. 2.14.

Das verwendete Kameramodell beschreibt die komplette Abbildungskette von Objektiv, eigentlicher Kamera mit CCD-Chip, Signalübertragung bis hin zur Digitalisierung. Daher sind die Modellparameter von all diesen Komponenten abhängig. Aufgrund der Modellvorstellung für die Abbildungskette sind die Parameter nicht aus den technischen Daten der einzelnen Komponenten zu bestimmen. Diese sind vielmehr in einem Kalibrierungsprozeß zu ermitteln.

An einen Kalibrierungsprozeß sind neben der Anforderung an die Genauigkeit auch noch Anforderungen bezüglich einer einfachen Handhabbarkeit und einer geringen Komplexität und einfachen Herstellung des Eichkörpers zu stellen. Ein entsprechendes Kamerakalibrierwerkzeug ist in das verwendete Bildverarbeitungssystem HALCON integriert.[1] In den folgenden Abschnitten wird im wesentlichen die Handhabung der Kalibrierung im Bezug auf die Anforderungen für STABIL^{++} erläutert. Es soll daher hier keine exakte Herleitung der Kalibrierung gegeben werden, hierzu sei auf die Ausführungen in [LZB95] und [Lan98] verwiesen.[2]

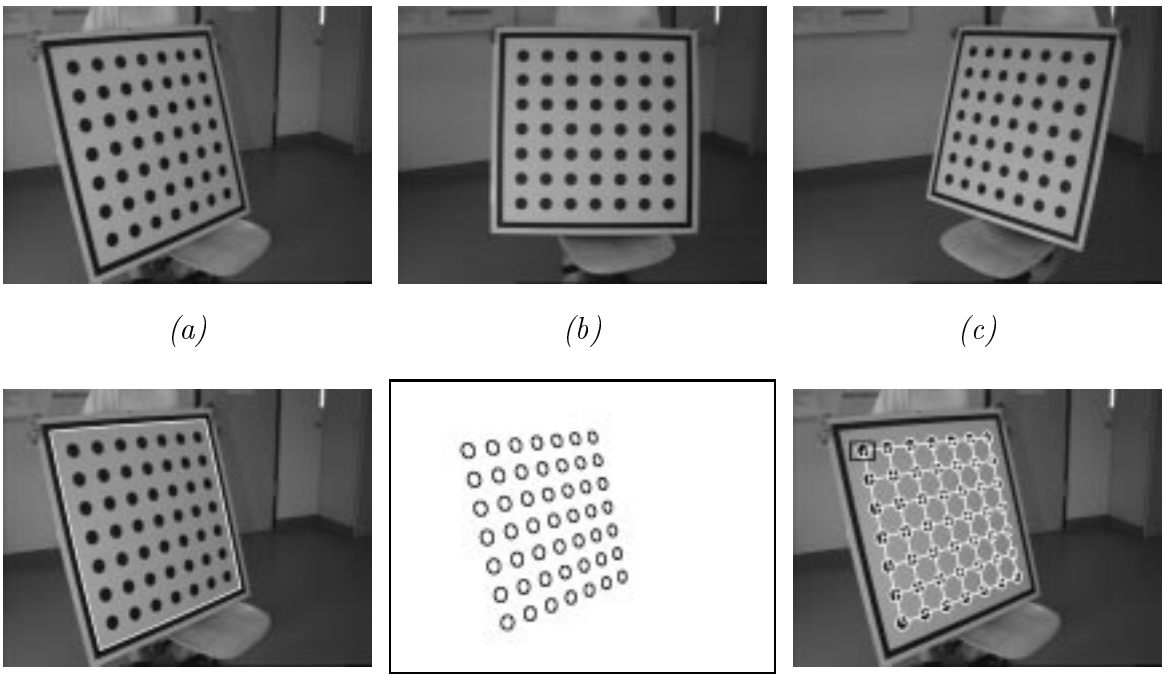
C.2 Innere Kameraparameter

Das Kameramodell beschreibt die Abbildung von 3D Weltpunkten in 2D Bildpunkte, daher sind für die Kalibrierung der Modellparameter vermessene dreidimensionale Eichpunkte notwendig. Jedoch hat ein 3D Eichkörper drei Nachteile: Erstens ist die hochgenaue Herstellung oder 3D Vermessung des Eichkörpers sehr aufwendig. Zweitens ist die Handhabung und der Transport aufgrund der 3D Ausdehnung schwierig. Drittens ist es nicht einfach, alle Meßpunkte des Eichkörpers immer ohne Eigenverdeckung im Bild zu sehen.

Daher wird von dem Kalibrierungswerkzeug ein zweidimensionaler Eichkörper verwendet. Der Eichkörper wird aus vielen verschiedenen Blickwinkeln aufgenommen, so daß mehrere Videoaufnahmen in einem Kalibriervorgang ausgewertet werden. Es entsteht somit eine Multibildkalibrierung bei der ein *Bündelblockausgleichsverfahren* angewendet wird. Da die Lage des Eichkörpers zur Kamera bei jeder Kalibrieraufnahme unterschiedlich ist, wird quasi ein 3D Eichkörper simuliert.

[1] Vgl. Kap. 3.4.

[2] Vgl. hierzu auch die Anm. in Kap. 2.5.4.



(a) *(b)* *(c)*

(d) *(e)* *(f)*

Abbildung C.1: *Obere Reihe - drei beispielhafte Aufnahmen des verwendeten Eichkörpers für die Multibildkalibrierung: Untere Reihe - (d) der in Bild (a) detektierte Eichkörper; (e) die detektierten Eichkörpermarken; (f) Rückprojektion des Eichkörpermodells; die Aufnahmen sind [Lan98] entnommen.*

Als Eichkörper wird eine Kalibrierplatte verwendet, auf der 49 dunkle kreisförmige Eichkörpermarken auf hellem Untergrund zu sehen sind. Die Kalibrierplatte ist hochgenau vermessen, so daß die Lage und Größe der Marken auf der Platte bekannt sind. Der verwendete Eichkörper und drei beispielhafte Kalibrieraufnahmen sind in Abb. C.1 zu sehen.

Zur automatischen Detektion des Eichkörpers wird im Bild nach einer hellen zusammenhängenden Region mit 49 dunklen "Löchern" gesucht, vgl. Abb. C.1(d). Anschließend werden durch eine subpixelgenaue Kontursuche die Umrisse der Markenpunkte ermittelt, vgl. Abb. C.1(e). Die Markenmittelpunkte werden dann durch eine Ellipsenanpassung bestimmt. Somit erhält man für jede Kalibrieraufnahme 49 subpixelgenaue Bildpunkte für die Markenmittelpunkte.

Im weiteren werden für die inneren Kameraparameter $camPar$ Startwerte angenommen. Hierbei

- wird für die Kammerkonstante b die nominelle Brennweite des Objektivs angenommen.
- wird keine radiale Verzerrung angenommen, daher wird der Verzerrungskoeffizient $\kappa = 0$ gesetzt.
- werden für die Skalierungsfaktoren S_x und S_y entsprechend der Größe des CCD-Chips und der Bildgröße gesetzt. Beispielsweise hat ein $1/2$"-CCD-Chip eine Breite von 6,4 mm und eine Höhe von 4,8 mm[3]. Es ergibt sich dann bei einer Bildauflösung von 768 $\times$ 576

[3]Eine Aufstellung der Maße der gängigen CCD-Chip-Größen sind in [Gwo97] auf S. 58 zu finden.

Bildpunkten (PAL-Norm):

$$S_x = \frac{6{,}4}{768} \left[\frac{mm}{Bildpunkte}\right] = 0.00833 \left[\frac{mm}{Bildpunkt}\right]$$
$$S_x = \frac{4{,}8}{576} \left[\frac{mm}{Bildpunkte}\right] = 0.00833 \left[\frac{mm}{Bildpunkt}\right]$$

- für den Hauptpunkt der Verzerrung (C_x, C_y) ist die Bildmitte angenommen.

Aufgrund dieser Startwerte, dem Wissen über die Anordnung und Größe der Eichkörperpunkte und der Position, der Größe und der Orientierung der detektierten Ellipsen in den Kalibrieraufnahmen läßt sich die Lage der Kalibrierplatte zur Kamera bestimmen. Definiert man ein 3D Koordinatensystem auf der Kalibrierplatte, für das alle Marken in einer Ebene liegen,[4] dann ist eine Transformation zwischen diesem Eichkörperkoordinatensystem und dem Kamerakoordinatensystem bestimmt. Mit diesen Transformationen für jede Kalibrieraufnahme erhält man Startwerte für die äußeren Kameraparameter. Daher kann auf eine Vermessung des Kalibrieraufbaus verzichtet werden, womit eine einfache Handhabbarkeit der Kalibrierung gegeben ist.

Für jeden in den Kalibrieraufnahmen lokalisierten Eichkörperpunkt erhält man eine Abbildungsgleichung. Entsprechend den vier Projektionsschritten des Kameramodells ergibt sich die Projektion für die Punkte aus den Glg. 2.13, 2.16, 2.17 und 2.18. In die Gleichungen gehen für alle lokalisierten Eichkörperpunkte deren Bildpunkte, die entsprechende Position im Eichkörperkoordinatensystem und die Startwerte für die inneren und äußeren Kameraparameter ein. Hierbei sind die Startwerte für die äußeren Kameraparameter für die einzelnen Aufnahmen verschieden. Über eine nichtlineare Ausgleichsrechnung werden die Kameraparameter über alle aufgestellten Gleichungen optimiert. Hierbei werden die Fehler in den Abbildungen der 3D Eichkörperpunkte zu den Bildpunkten minimiert.[5] Für eine Beschreibung der Extraktion der Eichkörpermarken und Untersuchungen zur Genauigkeit der Kamerakalibrierung sei auf [LZB95], [Lan98] verwiesen. Mit der Kalibrierung ist für jede Kalibrieraufnahme auch die Translation des Eichkörperkoordinatensystems zum Kamerakoordinatensystem bestimmt, womit sich das Modell des Eichkörpers in das entsprechende Bild projizieren läßt, vgl. Abb. C.1 (f). Anhand dieser Projektion kann die Genauigkeit der Kalibrierung überprüft werden.

C.3 Äußere Parameter

Mit dem im vorhergehenden Abschnitt beschriebenen Verfahren werden auch die äußeren Kameraparameter bestimmt. Bei den unterschiedlichen Aufnahmen der Kalibrierplatte ist somit jeweils eine Transformation des Kamerakoordinatensystems zu dem Eichkörperkoordinatensystem auf der Kalibrierplatte bestimmt worden.

Für die 3D Detektion und -verfolgung von Objekten müssen die Kameras jedoch zum Weltkoordinatensystem wcs des Szenenmodells $scene$ vermessen sein. Hierzu ist es möglich, die Lage der Kalibrierplatte in einer der Kalibrieraufnahmen zum Weltkoordinatensystem zu vermessen und die ermittelten Daten für die Transformation um die entsprechende Lage zu korrigieren.

Eine weitere Möglichkeit besteht darin, die Kalibrierung der inneren und äußeren Kameraparameter zu trennen. Damit kann die Kalibrierung der inneren Kameraparameter in einer

[4] In dem verwendeten Kalibrierwerkzeug liegen die Punkte in der xy-Ebene des Eichkörperkoordinatensystems; der Ursprung des Koordinatensystems liegt in der Mitte der mittleren Marke; die x-Achse zeigt beim Blick auf die Kalibrierplatte nach rechts, die y-Achse nach unten, somit zeigt die z-Achse in die Platte hinein.

[5] Methode der kleinsten Quadrate; (engl.) *least-squares*.

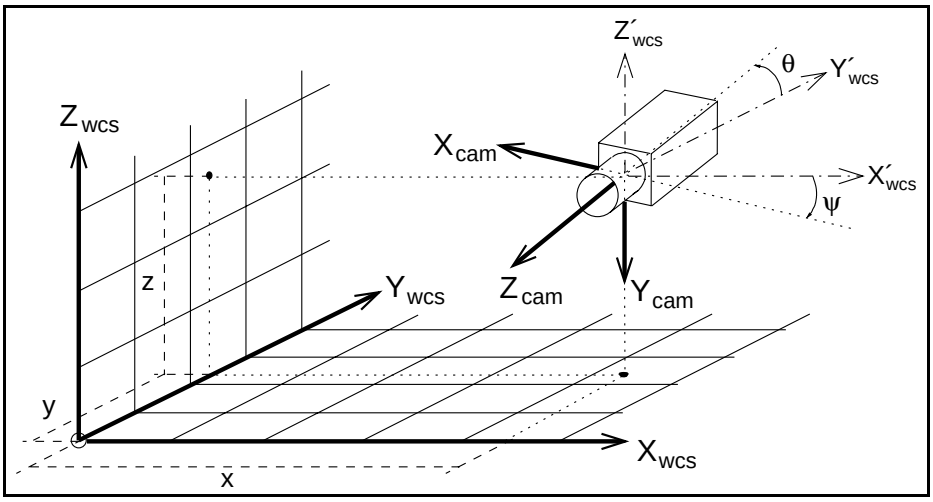


Abbildung C.2: *Lage des Kamerakoordinatensystems* $[X_{cam}, Y_{cam}, Z_{cam}]$ *im Weltkoordinatensystem* $[X_{wcs}, Y_{wcs}, Z_{wcs}]$ *mit Translation* $[x, y, z]^T$ *und Rotationen um einen Neigewinkel* θ *und einen Schwenkwinkel* ψ*.*

Laborumgebung vorgenommen werden, wobei die Kalibrierung der äußeren Kameraparameter vor Ort, am Einsatzort des Systems, vorgenommen werden kann. Hierzu werden als Meßpunkte natürliche oder künstliche Marken in dem durch das Weltkoordinatensystem aufgespannten Raum verwendet. Die Koordinaten dieser Punkte müssen, bezogen auf das Weltkoordinatensystem, vermessen werden. Durch Lokalisation der Marken im Bild oder einfaches Markieren im Bild erhält man die zugehörigen Bildpunkte. Wie bei der Kalibrierung der inneren Kameraparameter werden für jeden Meßpunkt die Abbildungsgleichung aufgestellt und die Parameter der Gleichungen über eine nichtlineare Ausgleichsrechnung optimiert. In diese Gleichungen gehen die 3D Meßpunkte, die entsprechenden Bildpunkte, die zuvor ermittelten inneren Kameraparameter und Startwerte für die äußeren Kameraparameter ein.

Als Startwerte für die äußeren Kameraparameter muß eine ungefähre Lage des Kamerakoordinatensystems im Weltkoordinatensystem durch eine homogene Transformationsmatrix ${}^{cam}\mathbf{T}_{wcs}$ angegeben werden. Abb. C.2 zeigt dies beispielhaft. Dort befindet sich der Ursprung des Kamerakoordinatensystems $[X_{cam}, Y_{cam}, Z_{cam}]$ im Weltkoordinatensystem $[X_{wcs}, Y_{wcs}, Z_{wcs}]$ an dem Punkt $[x, y, z]^T$. Die Kamera ist um den Neigewinkel θ aus der waagerechten Orientierung nach unten geneigt und um den Schwenkwinkel ψ gegenüber einer parallelen Ausrichtung zur y-Achse des Weltkoordinatensystems verdreht. Die Kamera ist, wie in den meisten Anwendungen erwünscht, nicht um ihre Längsachse gedreht, so daß der Rollwinkel $\phi = 0$ ist.

Anstelle der eigentlichen, dem Kameramodell entsprechenden äußeren Kameraparameter ${}^{cam}\mathbf{T}_{wcs}$ ist es oftmals einfacher, die Lage der Kamera durch die Verschiebung des Weltkoordinatensystems in das Kamerakoordinatensystem zu beschreiben. Daher wird im folgenden die Bestimmung der zu ${}^{cam}\mathbf{T}_{wcs}$ inversen Transformation ${}^{wcs}\mathbf{T}_{cam}$ beschrieben.[6]

Es empfiehlt sich, die Sichtweise der Transformation so zu wählen, daß die Translation vor der Rotation ausgeführt wird. Daher kann man zur Ermittelung der Startwerte zunächst das Weltkoordinatensystem virtuell in die Position der Kamera verschieben. Man erhält somit das Koordinatensystem $[X'_{wcs}, Y'_{wcs}, Z'_{wcs}]$. Bei der Bestimmung der Rotationswinkel sind je nach Lage (Position und Orientierung) des Weltkoordinatensystems zur Kamera vier Fälle zu

[6] In dem verwendeten Kalibrierwerkzeug wird die Transformation daher auch direkt als ${}^{wcs}\mathbf{T}_{cam}$ angegeben.

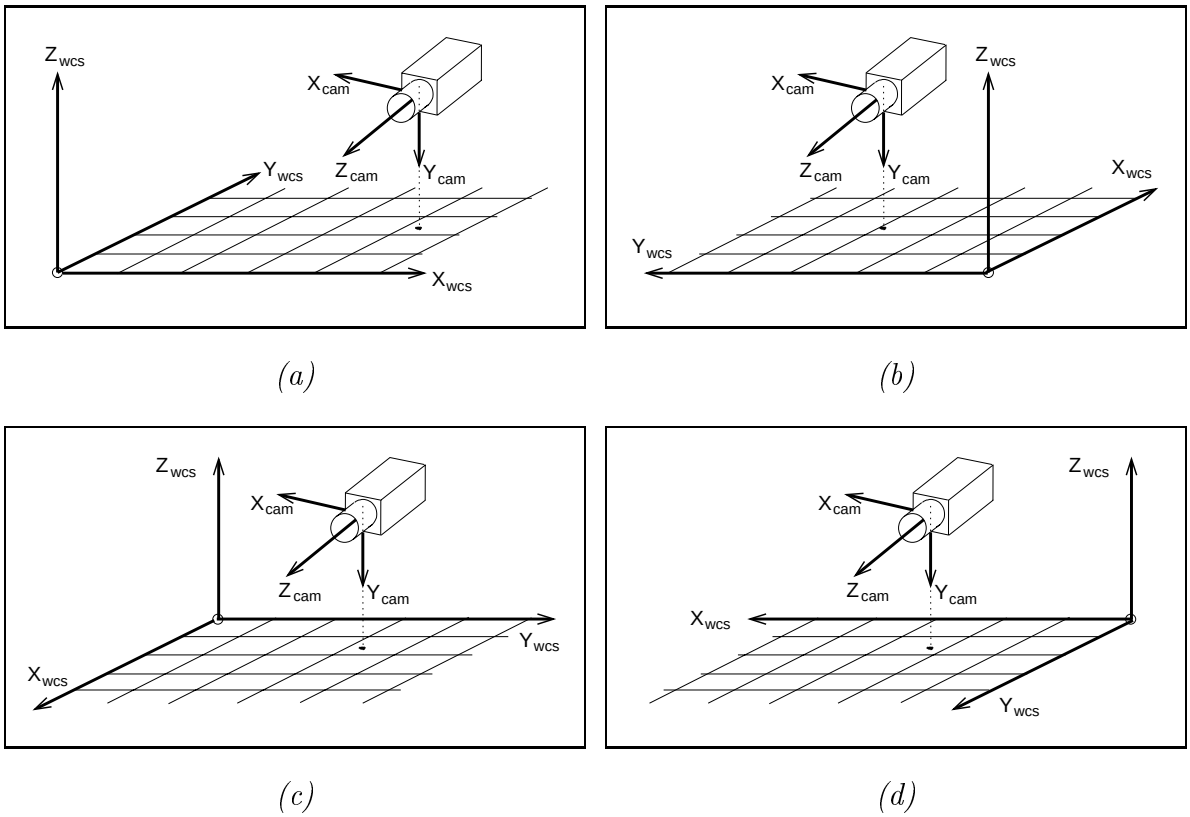


Abbildung C.3: *Vier verschiedene Lagen des Kamerakoordinatensystems* $[X_{cam}, Y_{cam}, Z_{cam}]$ *zum Weltkoordinatensystem* $[X_{wcs}, Y_{wcs}, Z_{wcs}]$*, die bei der Bestimmung der Rotationswinkel zu beachten sind.*

unterscheiden. Es muß unterschieden werden, ob der Ursprung des Weltkoordinatensystems vor oder hinter der Kamera liegt und dort jeweils links oder rechts, vgl. Abb. C.3. Hierbei sind neben dem Neige- und Schwenkwinkel noch zu beachten, daß das Kamerakoordinatensystem so orientiert ist, daß die z-Achse der optischen Achse entspricht. Somit verläuft dann die x-Achse des Kamerakoordinatensystems horizontal zum Kamerabild und die y-Achse vertikal, vgl. Abb. 2.3.

Für die Angabe der Rotationswinkel wählt man aufgrund der einfacheren Handhabbarkeit sinnvollerweise die Notation des $zy'x''$-Systems. Für die vier möglichen Positionen ergeben sich die Rotationswinkel bei vorgegebenem Neigewinkel θ und Schwenkwinkel ψ entsprechend Tab. C.1. Die vier Fälle sind in der Tabelle entsprechend der Bezeichnung in Abb. C.3 gekennzeichnet. Es ist zu beachten, daß es sich bei den Angaben nicht um die einzigen Lösungen handelt. Die Rotationen können für den Fall (a) auch mit $\gamma = -\psi$, $\beta = 180°$ und $\alpha = (90° - \theta)$ angesetzt werden.

Für aktive Kameras, z.B. Schwenk-/Neigekameras verändern sich mit jeder Bewegung die äußeren Kameraparameter. Daher muß die Kalibrierung mit der Bewegung nachgeführt werden. Auf die Kalibrierung von aktiven und insbesondere von Schwenk-/Neigekameras wird im letzten Abschnitt eingegangen.

	γ	β	α
(a)	$180° - \psi$	0	$-(90° + \theta)$
(b)	$90° - \psi$	0	$-(90° + \theta)$
(c)	$-(90° + \psi)$	0	$-(90° + \theta)$
(d)	$-\psi$	0	$-(90° + \theta)$

Tabelle C.1: *Rotationswinkel der Transformation ${}^{wcs}\mathbf{T}_{cam}$ in Abhängigkeit der Lage der Kamera im Weltkoordinatensystem; Winkelangaben im $zy'x''$-System; θ = Neigewinkel; ψ = Schwenkwinkel.*

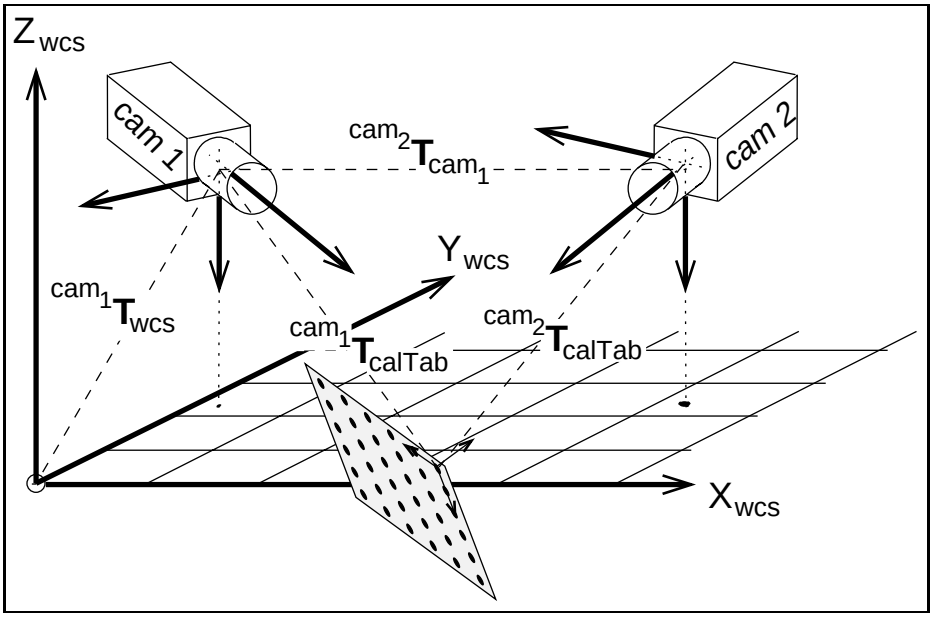


Abbildung C.4: *Bestimmung der äußeren Kameraparameter bei mehreren Kameras.*

C.4 Mehrere Kameras

Beim Einsatz von mehreren Kameras in einem System müssen diese alle zum Weltkoordinatensystem wcs kalibriert sein. Hierzu werden zunächst für alle Kameras die internen Kameraparameter bestimmt. Für eine der Kameras wird, wie im letzten Abschnitt beschrieben, die Transformation zum Weltkoordinatensystem bestimmt. Entsprechend Abb. C.4 ist dies für eine Kamera cam_1 die homogene Transformationsmatrix ${}^{cam_1}\mathbf{T}_{wcs}$, womit

$$\vec{p}_{cam_1} = {}^{cam_1}\mathbf{T}_{wcs} \cdot \vec{p}_{wcs}$$

gilt. Sofern sich die Sichtbereiche von zwei Kameras überschneiden und in beiden Kameras gleichzeitig der Eichkörper abgebildet werden kann, so kann auf eine direkte Kalibrierung der zweiten Kamera cam_2 zum Weltkoordinatensystem verzichtet werden. Stattdessen kann die Transformationsmatrix ${}^{cam_2}\mathbf{T}_{wcs}$ bestimmt werden, wenn man beide Kameras zu einem Referenzkoordinatensystem kalibriert. Hierzu eignet sich ein auf dem Eichkörper definiertes Koordinatensystem. Über das Kalibrierwerkzeug werden dann die Transformationsmatrizen ${}^{cam_1}\mathbf{T}_{calTab}$ und ${}^{cam_2}\mathbf{T}_{calTab}$ bestimmt. Mit diesen Matrizen gilt:

$$\begin{aligned} \vec{p}_{cam_1} &= {}^{cam_1}\mathbf{T}_{calTab} \cdot \vec{p}_{calTab} \\ \vec{p}_{cam_2} &= {}^{cam_2}\mathbf{T}_{calTab} \cdot \vec{p}_{calTab} \end{aligned}$$

Für die Transformation zwischen den beiden Kamerakoordinatensystemen ergibt sich hieraus:

$$\begin{aligned} {}^{cam_2}\mathbf{T}_{cam_1} &= {}^{cam_2}\mathbf{T}_{calTab} \cdot {}^{cam_1}\mathbf{T}^{-1}_{calTab} \\ &= {}^{cam_2}\mathbf{T}_{calTab} \cdot {}^{calTab}\mathbf{T}_{cam_1} \end{aligned}$$

und damit

$$\vec{p}_{cam_2} = {}^{cam_2}\mathbf{T}_{cam_1} \cdot \vec{p}_{cam_1}$$

Somit ist die Lage der beiden Kameras zueinander bestimmt worden. Nachdem die Lage der ersten Kamera im Weltkoordinatensystem bereits bekannt ist, kann durch einfache Matrizenmultiplikation die Transformationsmatrix ${}^{cam_2}\mathbf{T}_{wcs}$ bestimmt werden, es gilt:

$$\begin{aligned} {}^{cam_2}\mathbf{T}_{wcs} &= {}^{cam_2}\mathbf{T}_{cam_1} \cdot {}^{cam_1}\mathbf{T}_{wcs} \\ &= {}^{cam_2}\mathbf{T}_{calTab} \cdot {}^{cam_1}\mathbf{T}^{-1}_{calTab} \cdot {}^{cam_1}\mathbf{T}_{wcs} \end{aligned} \tag{C.1}$$

und damit

$$\vec{p}_{cam_2} = {}^{cam_2}\mathbf{T}_{wcs} \cdot \vec{p}_{wcs}$$

Für jede weitere Kamera kann genauso verfahren werden, wenn sich die Sichtbereiche überlappen und die Kalibrierplatte in den Kamerabildern von jeweils zwei Kameras sichtbar ist. Es ist hierbei jedoch zu beachten, daß sich die Restfehler des Fehlerminimierungsverfahrens aus jedem einzelnen Kalibriervorgang fortpflanzt. Daher kann bei der Kalibrierung von mehr als zwei Kameras das Bündelausgleichsverfahren über die für alle Kalibrieraufnahmen aufgestellten Abbildungsgleichungen angewendet werden; damit wird der Fehler über das Gesamtsystem aller Kameras minimiert.

Aufgrund der einfacheren Handhabbarkeit wird oft mit den invertierten Transformationen gearbeitet.[7] Damit wird die Lage der Kamera im Weltkoordinatensystem durch die Verschiebung des Weltkoordinatensystems in das Kamerakoordinatensystem durch ${}^{wcs}\mathbf{T}_{cam}$ angegeben. Zur Bestimmung der Lage der cam_2 im Weltkoordinatensystem muß Glg. C.1 entsprechend invertiert werden. Entsprechend ergibt sich:

$$\begin{aligned} {}^{wcs}\mathbf{T}_{cam_2} &= {}^{wcs}\mathbf{T}_{cam_1} \cdot {}^{cam_1}\mathbf{T}_{cam_2} \\ &= {}^{wcs}\mathbf{T}_{cam_1} \cdot {}^{calTab}\mathbf{T}^{-1}_{cam_1} \cdot {}^{calTab}\mathbf{T}_{cam_2} \\ \text{wobei} & \\ \vec{p}_{wcs} &= {}^{wcs}\mathbf{T}_{cam_2} \cdot \vec{p}_{cam_2} \end{aligned}$$

C.5 Schwenk- / Neigekameras

Aufbau und Koordinatensysteme

Schwenk- / Neigekameras haben aufgrund ihrer Konstruktion zwei Bewegungsfreiheitsgrade. Dies ist die Rotation um den Schwenkwinkel ψ und Neigewinkel θ. Eine Rotation der Kamera um die optische Achse (Rollbewegung) ist hierbei nicht vorgesehen. Die in STABIL^{++} integrierten handelsüblichen Kuppelkameras (engl. *Dome-Camera*) erlauben einen Schwenk von $0° - 360°$, damit kann die Kamera ohne Anschlag verdreht werden. Für den Neigewinkel ergibt sich eine Begrenzung ca. $20°$ über der waagerechten Blickrichtung und ca. $10°$ weiter als die senkrechte Blickrichtung nach unten.

[7] Auch das von STABIL^{++} verwendete Kalibrierwerkzeug liefert nicht die äußeren Kameraparameter ${}^{cam}\mathbf{T}_{wcs}$ entsprechend des Kameramodells, sondern die inverse Transformation ${}^{wcs}\mathbf{T}_{cam}$.

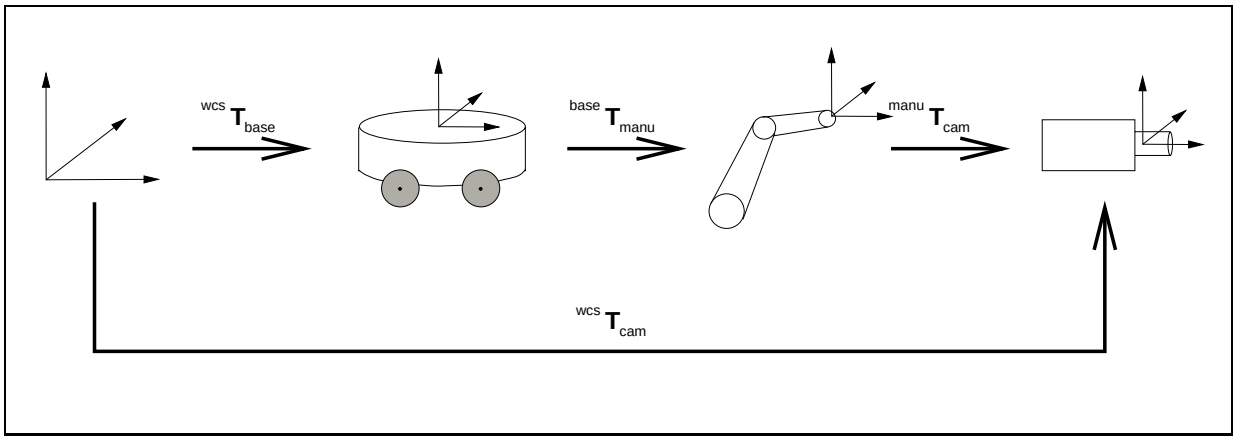


Abbildung C.5: *Transformationen zwischen den Kamerakoordinatensystemen in einem Hand-Auge-System in der Robotik.*

Durch die Rotationen wird das, mit dem verwendeten Kameramodell eingeführte, Kamerakoordinatensystem in seiner Lage zum Weltkoordinatensystem des Szenenmodells verändert. Fällt der Ursprung des Kamerakoordinatensystems mit dem Ursprung des Drehzentrums der aktiven Kamera zusammen, so könnte man für jede neue Lage des Kamerakoordinatensystems die Transformationsmatrix ${}^{wcs}\mathbf{T}_{cam}$ durch eine Multiplikation mit der entsprechenden Rotationsmatrix anpassen. Jedoch fallen baulich bedingt der Ursprung des Drehsystems und des Kamerakoordinatensystems nicht zusammen. Dieser Versatz muß entsprechend berücksichtigt werden, so daß zusätzlich zum Kamerakoordinatensystem in der aktiven Schwenk- / Neigekamera noch zwei weitere Koordinatensysteme definiert werden.

Hierzu wird die Analogie zu Kameras an Manipulatoren von Robotern gesucht. Dort ist zunächst die Lage eines Basiskoordinatensystems $base$ festzulegen. Dies kann in der festmontierten Grundplatte eines Montageroboters liegen oder auch in dem Fahrwerk eines mobilen Systems. Bezogen auf dieses Basiskoordinatensystem wird ein Koordinatensystem in dem Manipulator festgelegt. Dieses Manipulatorkoordinatensystem wird mit $manu$ bezeichnet. Der Manipulator ist der Roboterarm und wird auch als Träger bezeichnet, da an diesem die Kamera montiert ist. Insbesondere für videogesteuerte Greifaufgaben ist es wichtig, den Versatz zwischen dem Koordinatensystem des Manipulators und dem Kamerakoordinatensystem zu kennen. Man spricht daher hier von dem *Hand-Auge-Versatz*. Der Hand-Auge-Versatz wird mit der Transformationsmatrix ${}^{manu}\mathbf{T}_{cam}$ angegeben. Diese Transformation ist fix und muß zuvor vermessen / kalibriert werden, vgl. den folgenden Abschnitt. Die Bewegungen, die mit dem Manipulator ausgeführt werden können, werden in der Transformationsmatrix ${}^{base}\mathbf{T}_{manu}$ berücksichtigt, die die Lage des Manipulatorkoordinatensystems in dem Basiskoordinatensystem angibt. Mit einer Transformationsmatrix ${}^{wcs}\mathbf{T}_{base}$ wird noch angegeben, welche Lage das Basissystem zum Weltkoordinatensystem des Szenenmodells hat. Die hierdurch aufgebaute Transformationskette ist in Abb. C.5 dargestellt. Es gilt daher für die Transformation eines Punktes $\vec{p}_{wcs}$, der im Weltkoordinatensystem definiert ist, in das Kamerakoordinatensystem:

$$\begin{aligned} \vec{p}_{cam} &= {}^{cam}\mathbf{T}_{wcs} \cdot \vec{p}_{wcs} \\ &= {}^{cam}\mathbf{T}_{manu} \cdot {}^{manu}\mathbf{T}_{base} \cdot {}^{base}\mathbf{T}_{wcs} \cdot \vec{p}_{wcs} \end{aligned} \tag{C.2}$$

für die umgekehrte Transformation gilt mit den entsprechenden inversen Transformationsma-

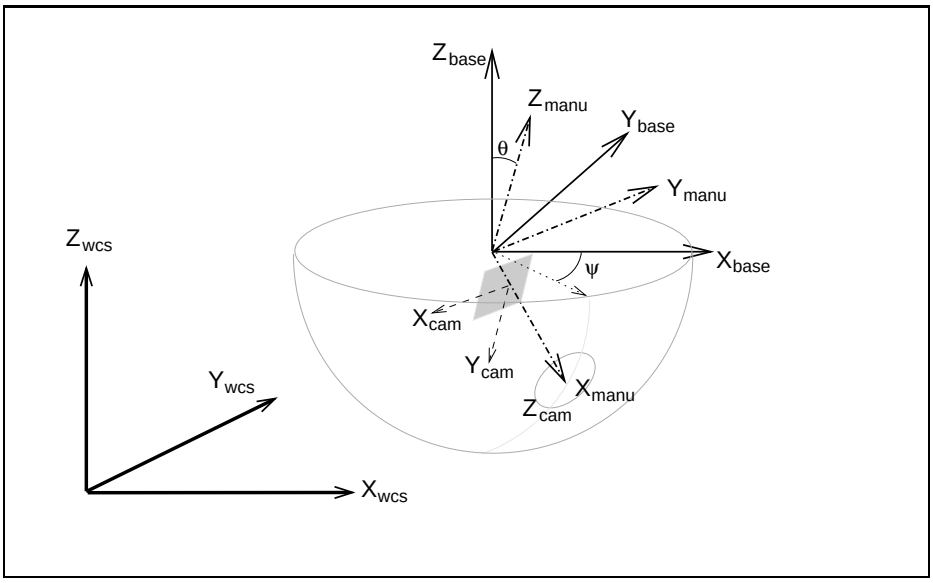


Abbildung C.6: *Koordinatensysteme in Schwenk- / Neigekameras (Kuppelkamera).*

trizen:

$$\begin{aligned} \vec{p}_{wcs} &= {}^{wcs}\mathbf{T}_{cam} \cdot \vec{p}_{cam} \\ &= {}^{wcs}\mathbf{T}_{base} \cdot {}^{base}\mathbf{T}_{manu} \cdot {}^{manu}\mathbf{T}_{cam} \cdot \vec{p}_{cam} \end{aligned} \tag{C.3}$$

Für die Schwenk- / Neigekameras ist das Basissystem durch das festmontierte Kameragehäuse bestimmt. In diesem Gehäuse kann sich ein Drehsystem bewegen, auf den die eigentliche Kamera, bestehend aus Objektiv und CCD-Chip montiert ist. Das Drehsystem ist mit dem Manipulator aus der Robotik zu vergleichen. Zur Vereinfachung wird angenommen, daß die beiden Drehachsen sich in einem Punkt treffen, in dem dann der Ursprung des Manipulatorkoordinatensystems liegt. In Abb. C.6 ist eine Kuppelkamera schematisch mit den definierten Koordinatensystemen dargestellt.

Die Transformationsmatrix ${}^{wcs}\mathbf{T}_{base}$ ist durch die Montage der Kamera bestimmt und verändert sich nicht. Die Schwenk- und Neigebewegung wird in der Transformationsmatrix ${}^{base}\mathbf{T}_{manu}$ abgebildet. Somit ist ${}^{base}\mathbf{T}_{manu}$ durch den Schwenkwinkel ψ und Neigewinkel θ bestimmt. Die als Hand-Auge-Versatz bezeichnete Transformation ${}^{manu}\mathbf{T}_{cam}$ gibt zum einen an, wie weit das Koordinatensystem des Manipulators und das Kamerakoordinatensystem gegeneinander versetzt sind.

Zum anderen wird mit dem Rotationsteil der Transformationsmatrix angegeben wie die beiden Systeme zueinander orientiert sind. Hierbei ist zu beachten, daß aufgrund des verwendeten Kameramodells das Kamerakoordinatensystem so orientiert ist, daß die z-Achse in Richtung der optischen Achse zeigt und die y-Achse im Bild nach unten zeigt; die x-Achse ergibt sich aus der Bedingung eines Rechtssystems. Das Manipulatorkoordinatensystem ist jedoch so orientiert, daß es in der Nullage bei $\theta = 0^\circ$ und $\psi = 0^\circ$ mit dem Basiskoordinatensystem zur Deckung kommt. Das Basiskoordinatensystem ist wiederum so orientiert, daß bei waagerechter Montage der Kamera die xy-Ebene parallel zur xy-Ebene des Weltkoordinatensystems liegt. Somit zeigt die optische Achse Z_{cam} ungefähr in Richtung der X_{manu}-Achse. Diese liegen aufeinander wenn zusätzlich zu den beiden Drehungen um 90° der Grundorientierung keine weitere Verdrehung zwischen den beiden Systemen vorhanden ist. Auch in Abb. C.6 ist lediglich ein Versatz zwischen Z_{cam} und X_{manu} dargestellt.

Hand-Auge-Kalibrierung

Zur Bestimmung des Hand-Auge-Versatzes ${}^{manu}\mathbf{T}_{cam}$ wird prinzipiell der gleiche Kalibrieransatz, wie zur Bestimmung der inneren und äußeren Kameraparameter verwendet. Es werden hierzu zunächst die inneren Kameraparameter der Kameraeinheit der Schwenk- / Neigekamera bestimmt.

Anschließend werden n Aufnahmen der Kalibrierplatte mit unterschiedlichen Schwenk- und Neigewinkeln vorgenommen. Für jede der n Position muß die Transformationsmatrix ${}^{base}\mathbf{T}_{manu1}$, $i = 1 \ldots n$ bekannt sein. Ferner muß die Kalibrierplatte an einer Stelle fixiert stehen und stellt ein Bezugskoordinatensystem, das als Weltkoordinatensystem wcs betrachtet werden kann, dar. Für jede dieser Aufnahmen läßt sich nun über die äußere Kamerakalibrierung die Transformation zwischen dem Weltkoordinatensystem und dem Kamerakoordinatensystem ${}^{wcs}\mathbf{T}_{cam}$ bestimmen. Entsprechend der Glg. C.3 und der Abb. C.5 erhält man daher für jede Aufnahme eine Gleichung entsprechend:

$$ {}^{wcs}\mathbf{T}_{cami} = {}^{wcs}\mathbf{T}_{base} \cdot {}^{base}\mathbf{T}_{manui} \cdot {}^{manu}\mathbf{T}_{cam}, \quad i = 1 \ldots n \tag{C.4} $$

Die Position der Kalibrierplatte soll bezüglich dem Basissystem der Schwenk- / Neigekamera nicht explizit vermessen werden, damit erhält man n Gleichungen zur Bestimmung von ${}^{wcs}\mathbf{T}_{base}$ und dem Hand-Auge-Versatz ${}^{manu}\mathbf{T}_{cam}$ aus Glg. C.4. Verwendet man zusätzlich noch jeweils die Projektionsgleichungen der 49 Marken auf dem Eichkörper, so erhält man $n \cdot 49$ Gleichungen. Zur Bestimmung der beiden Transformationsmatrizen wird auch hier wieder ein Ausgleichsverfahren angewendet, bei dem die Fehler in allen Projektionen minimiert werden, vgl. [Lan98].

D Beispiele zum Interpretationsbaum

Entsprechend der Ausführungen zur Größe der Interpretationsbäume in Kap. 3.6.5 beeinflußt die Reihenfolge der Objektmodellteile in der hierarchischen, inneren Objektmodellstruktur den Aufbau des Interpretationsbaumes. Desweiteren hängt dies auch von der Anzahl und der Reihenfolge der verschiedenen Basisattribute ab. In diesem Abschnitt soll dies bezugnehmend auf die initiale Detektion der in Kap. 3.6.3 dargestellten Beispielinterpretation aufgezeigt werden.

Bezug zur Beispielinterpretation in Kap. 3.6.3

Hierzu wird zunächst auf die dort verwendete und in Abb. 3.13 (a) aufgezeigte hierarchische, innere Objektmodellstruktur verwiesen. Die Zuordnung der einzelnen primären Merkmale mit den entsprechenden Basisattributen ist für diese Beispielinterpretation in Tab. 3.4 aufgelistet. Der entsprechende Interpretationsbaum ist in einfacher Darstellung in Abb. 3.15 gezeigt. Für diese erste Beispielinterpretation ist in diesem Abschnitt in der Abb. D.1 der komplette Interpretationsbaum in einer detaillierten Darstellung dargestellt. Für jeden Knoten, der eine Assoziation von primärem Merkmal zu Modellmerkmal darstellt, sind dort alle 3D Positionen der bis zu dieser Ebene zugeordneten Szenenmerkmale projiziert dargestellt. Hierdurch erhält man einen Eindruck der hypothetisch zugeordneten inneren Modellstruktur.[1]

Abänderung der inneren Objektmodellstruktur

Abweichend von diesem ersten Beispiel zur initialen Detektion wird nun, bei gleichbleibenden Basisattributen der primären Merkmale der einzelnen Objektmodellteile, die hierarchische, innere Objektmodellstruktur des Objektmodells verändert. Bei dieser zweiten Beispielinterpretation sind die Nachfolgeobjektmodellteile des Objektmodellteils $omp_{0.1}$ des Rumpfes nun die Objektmodellteile des rechten Oberarms gefolgt von dem Objektmodellteil des linken Oberarms. Das Objektmodellteil des Halses ist nun das letzte Nachfolgeobjektmodellteil von $omp_{0.1}$. Die Numerierung der Objektmodellteile ist daher entsprechend der Tab. D.1 abgeändert.

Rumpf	0.1	grün	Hals	1.3	grün
rechter Oberarm	1.1	cyan	linker Oberarm	1.2	gelb
rechter Unterarm	2.1	gelb	linker Unterarm	2.2	cyan
rechte Hand	3.1	cyan	linke Hand	3.2	gelb

Tabelle D.1: *Zweite Beispielinterpretation: Numerierung der Objektmodellteile und Farbattribut der primären Merkmale.*

[1] Vgl. hierzu auch die weiteren Erläuterungen zur Abb. 3.16, bei der zusätzlich ein Kamerabild in den Knoten hinterlegt ist.

Ebene	Name	Attribut	Knoten	Nachf.	max.	reduz. %	gültig
0	Wurzel		1	2	1	0.000	1
1	Rumpf	grün	2	3	2	0.000	2
2	r. O-Arm	cyan	6	3	6	0.000	6
3	r. U-Arm	gelb	18	2	18	0.000	10
4	r. Hand	cyan	20	2	36	44.444	8
5	l. O-Arm	gelb	16	1	72	77.778	7
6	l. U-Arm	cyan	7	1	72	90.278	6
7	l. Hand	gelb	6	1	72	91.667	4
8	Hals	grün	4	0	72	94.444	4
Summe			80		351	77.208	48

Tabelle D.2: *Maßzahlen des Interpretationsbaums zur initialen Detektion der zweiten Beispielinterpretation; geänderte innere Objektmodellstruktur.*

Rumpf	0.1	grün	Hals	1.1	cyan
rechter Oberarm	1.2	grün	linker Oberarm	1.3	gelb
rechter Unterarm	2.1	gelb	linker Unterarm	2.2	cyan
rechte Hand	3.1	cyan	linke Hand	3.2	gelb

Tabelle D.3: *Dritte Beispielinterpretation: Numerierung der Objektmodellteile und Farbattribut der primären Merkmale.*

Der Interpretationsbaum ergibt sich entsprechend der Abb. D.2, wobei die Maßzahlen des Baumes in Tab. D.2 aufgelistet sind. Aufgrund der Verschiebung der Zuordnung der zweiten Merkmale mit dem Basisattribut der Farbe “grün” für das Objektmodellteil des Halses in die letzte Ebene des Baumes, steigt die maximale Anzahl der möglichen Assoziationen ab der Ebene 2 an. Die Gesamtanzahl der möglichen Assoziationen von 351 übersteigt daher deutlich die Anzahl von 282 möglichen Assoziationen bei der ersten Beispielinterpretation. Zusätzlich kann die Restriktion $restr^{(siblingD)}$ zur Überprüfung des Geschwister-Abstandes erst auf der Ebene 5 angewendet werden, da erst dann Zuordnungen für das zweite Nachfolgeobjektmodellteil des Objektmodellteils des Rumpfes erzeugt werden. Bei der ersten Beispielinterpretation war dies schon auf der Ebene 3 möglich, vgl. auch die Kennzeichnungen an den nicht gültigen Knoten in den Abb. D.1 und D.2.

Abänderung der Basisattribute

In einer dritten Beispielinterpretation ist gegenüber der Beispielinterpretation aus Kap. 3.6.3 das Basisattribut der primären Merkmale der Objektmodellteile des Halses und des rechten Oberarms vertauscht worden.[2] Es ergibt sich somit eine Zuordnung der Farbattribute entsprechend der Tab. D.3. Es ist zu beachten, daß sich die hierarchische, innere Objektmodellstruktur und somit die Numerierung der Objektmodellteile nicht geändert hat.

Der entsprechende Interpretationsbaum ist in Abb. D.3 dargestellt. Die zugehörigen Maßzahlen sind in Tab. D.4 zu finden. Auch hier hat sich die maximale Anzahl der möglichen

[2]Entsprechend ist auch ein anderes Bildpaar erzeugt worden, bei denen die farbigen Markierungen vertauscht wurden, vgl. Abb. 3.14.

Ebene	Name	Attribut	Knoten	Nachf.	max.	reduz. %	gültig
0	Wurzel		1	2	1	0.000	1
1	Rumpf	grün	2	3	2	0.000	2
2	Hals	cyan	6	1	6	0.000	5
3	r. O-Arm	grün	5	3	6	16.667	4
4	r. U-Arm	gelb	12	2	18	33.333	8
5	r. Hand	cyan	16	2	36	55.556	10
6	l. O-Arm	gelb	20	1	72	72.222	10
7	l. U-Arm	cyan	10	1	72	86.111	8
8	l. Hand	gelb	8	0	72	88.889	4
Summe			80		285	71.930	52

Tabelle D.4: *Maßzahlen des Interpretationsbaums zur initialen Detektion der dritten Beispielinterpretation; geänderte Basisattribute.*

Assoziationen im Vergleich zur ersten Beispielinterpretation erhöht, jedoch lediglich von 281 auf 285. Dies begründet sich auf der Vertauschung der Basisattribute in der zweiten und der dritten Ebene des Interpretationsbaumes. Auch in diesem Beispiel kann durch die ungeänderte hierarchische, innere Objektmodellstruktur in der Ebene 3 die Restriktion $restr^{(siblingD)}$ angewendet werden. Durch die andere Verwendung der Basisattribute ergeben sich auch andere Zuordnungen von Modell- / primären Merkmalen zu Szenenmerkmalen, so daß die Restriktion $restr^{(parentD)}(.)$ zur Überprüfung des Vater-Abstandes für die erste und die dritte Beispielinterpretation mit unterschiedlichem Erfolg eingesetzt werden kann. So ergibt sich hier, daß in der Ebene 4 noch acht gültige Knoten vorhanden sind, gegenüber sechs gültigen Knoten in der ersten Beispielinterpretation. Somit erhöht sich die Anzahl der aufgestellten Assoziationen von 59 auf 80. Anzumerken ist noch, daß die Hypothese mit der besten Bewertung bei der dritten Beispielinterpretation nicht im ersten Ast, wie in den anderen beiden Beispielinterpretationen, sondern im letzten Ast des Interpretationsbaumes liegt.

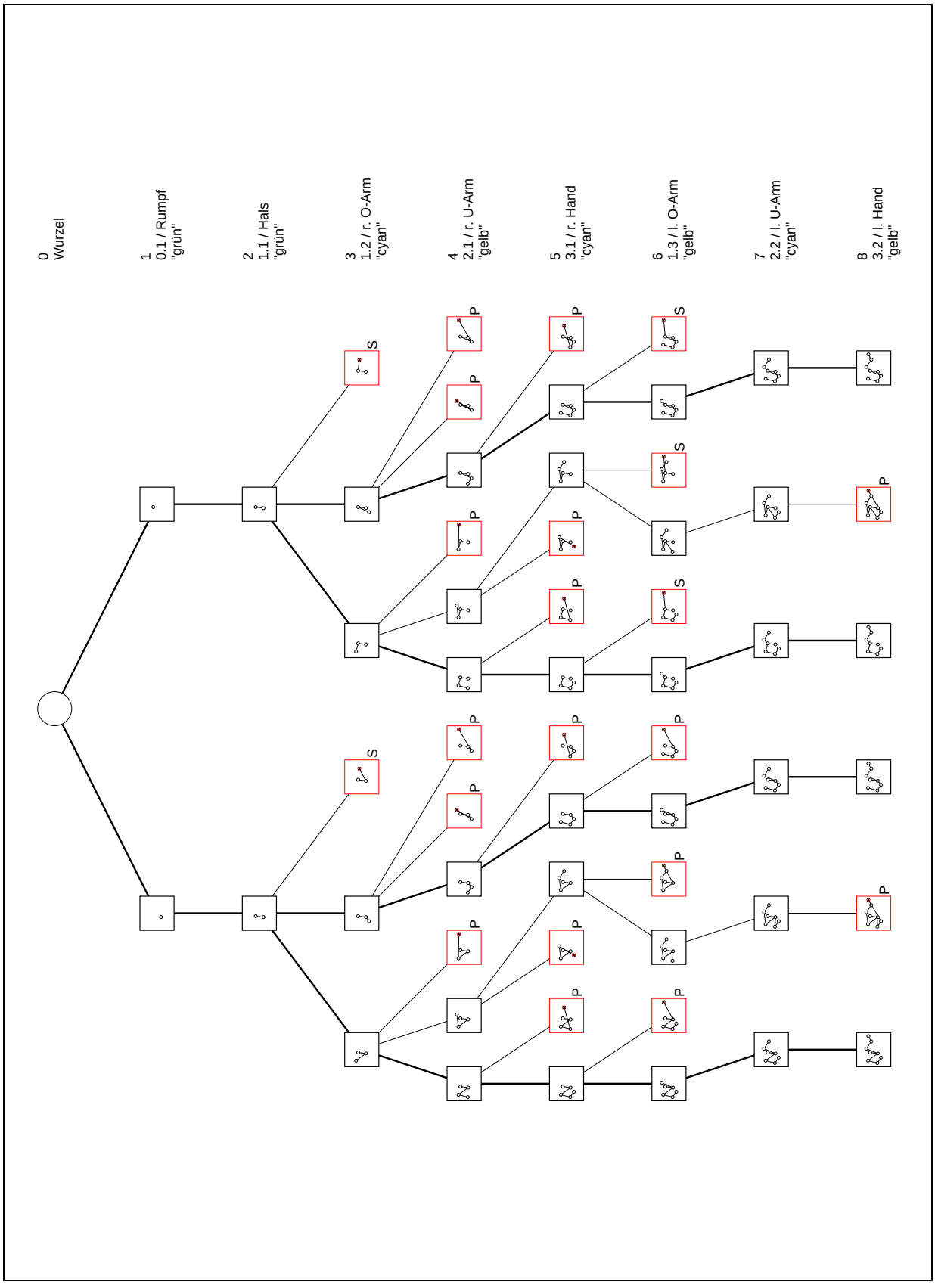


Abbildung D.1: *Detaildarstellung des Interpretationsbaums zur initialen Detektion der Beispielinterpretation aus Kap. 3.6.3.*

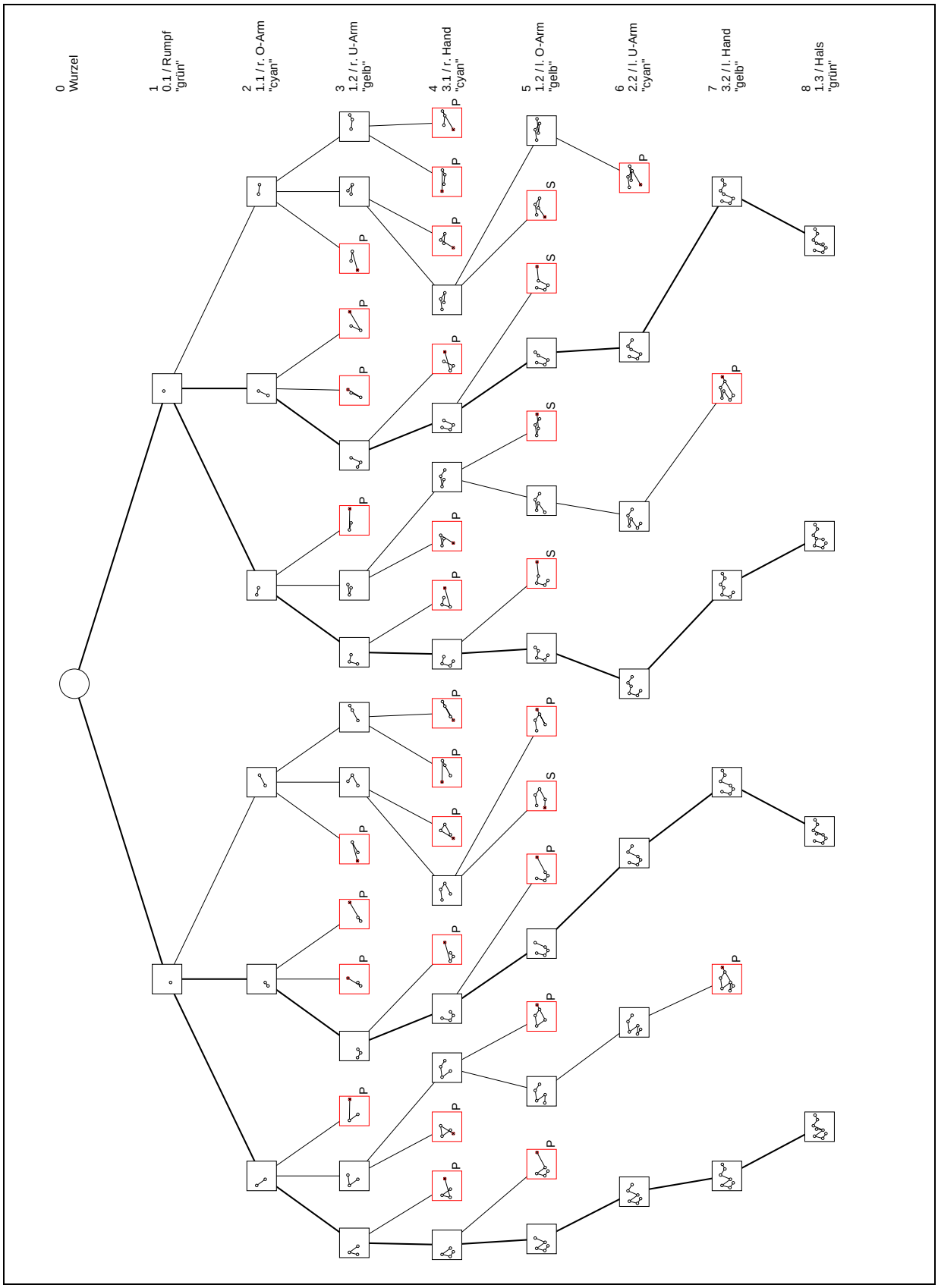


Abbildung D.2: *Detaildarstellung des Interpretationsbaums zur initialen Detektion der zweiten Beispielinterpretation; geänderte innere Objektmodellstruktur.*

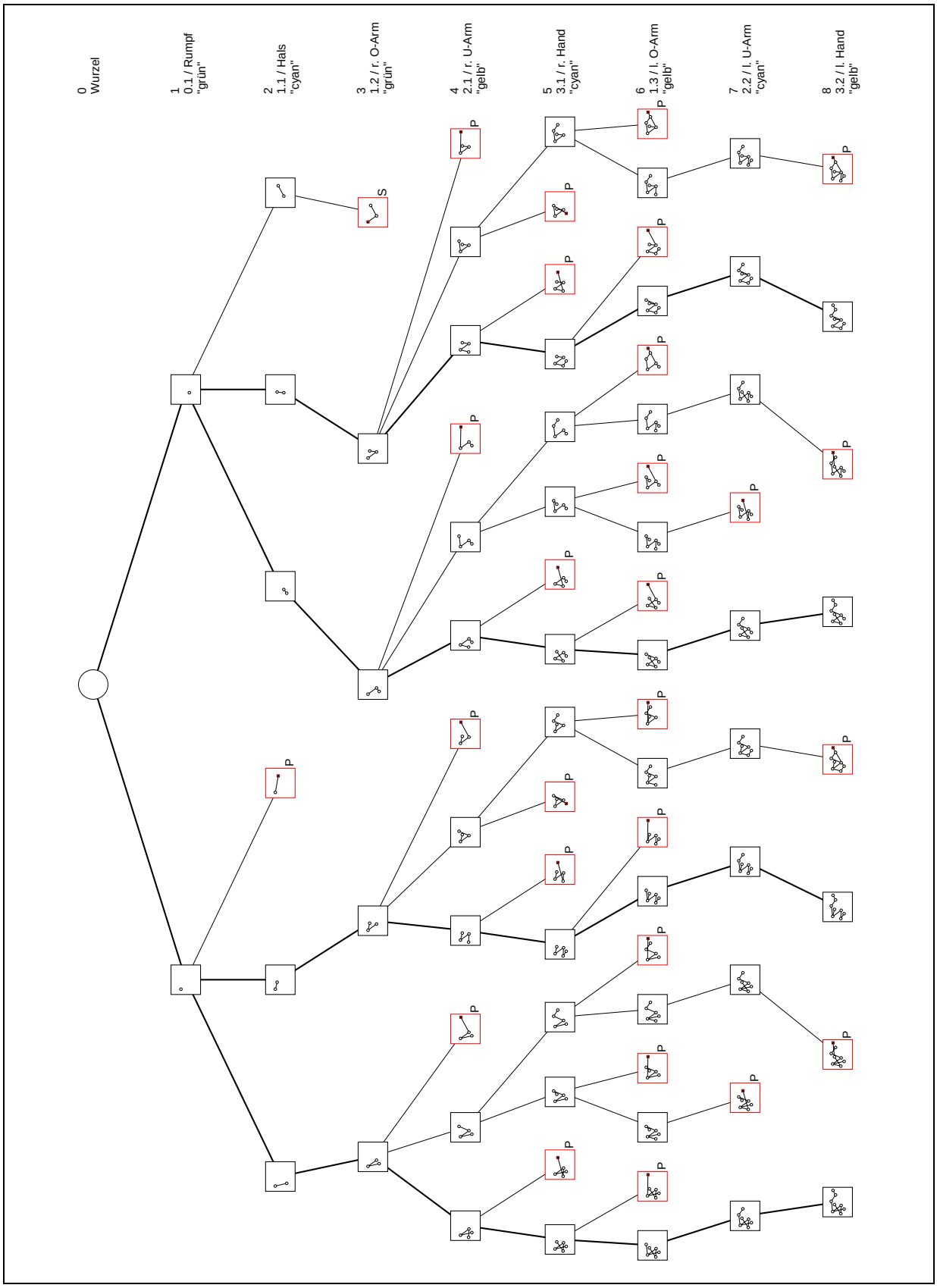


Abbildung D.3: *Detaildarstellung des Interpretationsbaums zur initialen Detektion der dritten Beispielinterpretation, geänderte Basisattribute.*

Abbildungsverzeichnis

Tabellenverzeichnis

Verzeichnis der Algorithmen

Symbolverzeichnis

In dem Symbolverzeichnis sind vorrangig die mathematischen Symbole aufgelistet, auf die sich die Modellierung stützt und die zur Erläuterung des Interpretationsprozesses verwendet werden. Alle die Symbole, deren Verwendung auf einen Abschnitt begrenzt ist, erscheinen hier nicht.

Insbesondere bei der Modellierung ist die Verwendung von Listen einzelner Komponenten notwendig, daher ist folgende Konvention eingeführt: einzelne Komponenten werden mit Kleinbuchstaben und eine Liste von Komponenten eines Typs mit den entsprechenden Großbuchstaben gekennzeichnet; z.B. ist die Liste von Objektmodellinstanzen $OBJ = \{obj_1, \ldots, obj_n\}$.

Die Notation der in der Modellierung verwendeten Merkmale erfolgt mit fett gedruckten Symbolen. Dies gilt sowohl für die einzelnen Merkmale, als auch für Listen von Merkmalen des gleichen Typs.

Die verwendeten Symbole sind oftmals aus den zugehörigen englischen Begriffen abgeleitet. Um diesen Bezug darzustellen, sind zur Erläuterung die englischen in Klammern angegeben. Desweiteren sind bei den Symbolen, bei denen eindeutig ein Verweis auf die Definition oder ersten Verwendung angegeben werden kann, die entsprechende Seite referenziert.

κ	Verzerrungskoeffizient der internen Kameraparameter, s. S. 47.
ϕ	Rollwinkel einer Kamera (roll), s. S. 200.
ψ	Schwenkwinkel einer Kamera (yaw), s. S. 200.
θ	Neigewinkel einer Kamera (pitch), s. S. 200.
act	Akteur zur Ausführung einer Aktion am Ende des Interpretationszyklus (engl. **act**or).
$assoc$	Assoziation / Zuordnung von Szenenmerkmal **s** zu primärem Merkmal **f** (engl. **assoc**iation), s. S. 98.
$attr$	Attribut (engl. **attr**ibute), s. S. 35.
b	Kammerkonstante der internen Kameraparameter, s. S. 47.
$[C_x, C_y]^T$	Hauptpunkt der internen Kameraparameter, s. S. 47.
cam	Kamera (engl. **cam**era), s. S. 42.
$camPar$	Interne / innere Kameraparameter (engl. **cam**era **par**ameter).
$camPose$	Äußere Kameraparameter (engl. **cam**era **pose**), s. S. 46.
can	Einzelner Bildkanal (engl. **c**h**an**nel), s. S. 42.
clc^{attr}	Klassifikator zur Segmentierung eines Videobildes in Regionen mit dem Attribut $attr$, s. S. 79.
$color$	Eine einem Farbklassifikator bekannte Farbe, z.B. ‘red’, s. S. 84.

$compo$	Kompositionen / Gruppierungen von Objektmodellteilen (engl. **compo**sition), s. S. 185.
$\mathbf{f}$	Primäres Merkmal eines Objektmodellteils (engl. **f**eature), s. S. 37.
$\mathbf{f}'$	Sekundäres Merkmal eines Objektmodellteils, s. S. 39.
h	Hypothese für eine Objektmodellinstanz, s. S. 99.
$HIST$	Historie des Objektmodells, s. S. 27.
$HIST_\nu$	Historie des Objektmodellteils omp_ν, s. S. 31.
$\mathbf{i}$	2D Bildmerkmal im Sensorraum (engl. **i**mage feature), s. S. 35.
$\mathbf{I}_{extr}$	Liste extrahierter Bildmerkmale, mit denen 3D Position von **s** bestimmt ist (engl. **extr**acted), s. S. 36.
img	2D Videobild, auch mehrkanalig (engl. **im**a**g**e), s. S. 42.
$ip(.)$	Bildverarbeitungsoperator (engl. **i**mage **p**rocessing operator), s. S. 76.
IP	Liste von Bildverarbeitungsoperatoren $\{ip(.)_1, \ldots, ip(.)_n\}$.
$\mathbf{m}$	3D Modellmerkmal im Modellraum (engl. **m**odel feature), s. S. 37.
$\mathbf{M}^{(attr)}$	Liste von Modellmerkmalen mit dem Basisattribut $attr$, s. S. 107.
obj	Objektmodell / Objektmodellinstanz (engl. **obj**ect), s. S. 26.
obj_0	initiales Objektmodell des Szenenmodells, s. S. 24.
omp	Objektmodellteil (engl. **o**bject **m**odel **p**art), s. S. 27.
$omp_{0.1}$	Erstes Objektmodellteil in hierarchischer, innerer Objektmodellstruktur (Wurzelelement).
$omp_{a.b}$	Objektmodellteil, das in der Hierarchie a Ebenen vom Objektmodellteil $omp_{0.1}$ entfernt ist und in der Ebene die Nummer b hat.
$pred_\nu$	Filter zur Vorhersage der 3D Position des Objektmodellteils omp_ν (engl. **pred**iction), s. S. 31.
$\vec{p}_\nu$	(2D oder) 3D Punkt $[x_\nu, y_\nu, z_\nu]^T$, der im Koordinatensystem ν definiert ist.
${}^\circ\vec{p}_{wcs}^{\,\mu}$	3D Position des Ursprungs des lokalen Koordinatensystems μ im Weltkoordinatensystem wcs.
$\vec{p}_{(-1)}$	3D Punkt zum Zeitpunkt $t_{(-1)}$.
$\tilde{\vec{p}}$	Vorhergesagter 3D Punkt.
$\tilde{\vec{p}}_0$	An alter Position vorhergesagter 3D Punkt, s. S. 65.
$\tilde{\vec{p}}_e$	Durch Extrapolation vorhergesagter 3D Punkt, s. S. 64.
q	Qualität $\in [0, 1]$ eines Szenenmerkmals **s**, einer Hypothese h oder als Funktionswert einer Gütefunktion $qual(.)$ (engl. **q**uality).
$qual(.)$	Gütefunktion zur Beurteilung von Hypothesen (engl. **qual**ity), s. S. 127.
$qual_{\mathbf{f}'}(.)$	Gütefunktion eines sekundären Merkmals, s. S. 134.
$reg^{(red)}$	Einzelregion, mit dem Attribut “der Klasse ‘rot’ zugehörig”, s. S. 43.

$restr(.)$	Restriktion zur Anwendung im Interpretationsbaum (engl. **restr**iction), s. S. 100.
$restr_{\nu}^{\angle}$	Restriktion für Gelenkwinkel des Objektmodellteiles omp_{ν}, s. S. 31.
$\mathcal{R}$	Rotationsmatrix.
$\mathbf{s}$	3D Szenenmerkmal im Szenenraum (engl. **s**cene feature), s. S. 36.
$\hat{\mathbf{s}}$	Geschätztes Szenenmerkmal, s. S. 115.
$\mathbf{s}_{nil}$	nil-Szenenmerkmal (engl. **n**ot **i**n **l**ist), s. S. 115.
$\mathbf{S}^{(attr)}$	Liste von Szenenmerkmalen mit dem Basisattribut $attr$, s. S. 108.
$scene$	Szenenmodell (engl. **scene**), s. S. 23.
ssp	3D Regionen allgemein, im speziellen Suchraum (**s**erach **sp**ace).
SSP^0	Suchräume für die Detektion der initialen Objektmodelle, s. S. 24.
SSP_s	Observierungsraum des Szenenmodells als Liste von Weltregionen $\{wr_1, \ldots, wr_n\}$, s. S. 24.
$\vec{s}$	Sichtstrahl, s. S. 47.
S_x, S_y	Skalierungsfaktoren der internen Kameraparameter, s. S. 47.
$t_{(-i)}$	Zeitstempel für einen Interpretationszyklus, der i Bildtakte zurückliegt (engl. **t**ime / **t**ime stamp), s. S. 27.
${}^{\mu}\mathbf{T}_{\nu}$	Homogene Transformationsmatrix (Transformation von Koordinatensystem μ in Koordinatensystem ν), s. S. 204.
$\mathcal{T}$	Translationsmatrix, s. S. 203.
$\vec{t}$	Verschiebungsvektor, der für ein Modellmerkmal die Position zum lokalen Koordinatensystem des zugehörigen Objektmodellteils angibt.
vol_{type}	Volumenkörper vom Typ $type$ (engl. **vol**ume), s. S. 32.
wcs	Bezeichnung des Weltkoordinatensystems (engl. **w**orld **c**oordinate **s**ystem).
wr	Weltregion zur Beschreibung des Observierungsraumes, s. S. 24.
$xy'z''$	Bezeichnung für Drehreihenfolge und Bezugsachsen in Rotationsmatrizen, s. S. 207.
$[X_{\mu}, Y_{\mu}, Z_{\mu}]$	Koordinatensystem μ mit x_{μ}, y_{μ} und z_{μ}-Achse, s. S. 199.

Literaturverzeichnis

[AC99] J. K. Aggarwal und Q. Cai. Human Motion Analysis: A Review. *Computer Vision and Image Understanding*, 73(3):428–440, März 1999.

[ACF99] J. Amat, A. Casals und M. Frigola. Stereoskopic System for Human Body Tracking. In *Proc. IEEE Workshop on Modelling People (MPEOPLE)*, Seiten 70–76, Korfu, Griechenland, September 1999. IEEE Computer Society Press.

[ACLS94] J. K. Aggarwal, Q. Cai, W. Liao und B. Sabata. Articulated and Elastic Non-rigid Motion: A Review. In *Proc. IEEE Workshop on Motion of Non-Rigid and Articulated Objects (NAM)*, Seiten 2–14, Austin, Texas, USA, November 1994. IEEE Computer Society Press.

[ACLS98] J.K. Aggarwal, Q. Cai, W. Liao und B. Sabata. Nonrigid Motion Analysis: Articulated and Elastic Motion. *Computer Vision and Image Understanding*, 70(2):142–156, Mai 1998.

[Aki84] K. Akita. Image Sequence Analysis of Real World Human Motion. *Pattern Recognition*, 17(1):73–83, 1984.

[Arl99] F. Arlt. *Untersuchungen zielgerichteter Bewegungen zur Simulation mit einem CAD-Menschmodell*. Dissertation, Lehrstuhl für Ergonomie der Technischen Universität München, 1999.

[Bad97] N. Badler. Virtual Humans for Animation, Ergonomics, and Simulation. In *Proc. IEEE Workshop on Motion of Non-Rigid and Articulated Objects (NAM)*, Seiten 28–36, San Juan, Puerto Rico, Juni 1997. IEEE Computer Society Press.

[BBBC98] H.-J. Böhme, U.-D. Braumann, A. Brakensiek und A. Corradini. User Localisation for Visual-based Human-Machine-Interaction. In *International Conference on Automatic Face and Gesture Recognition (FG)*, Seiten 486–491, Nara, Japan, April 1998. IEEE Computer Society Press.

[Bec96] D. A. Becker. Sensei: A Real-Time Recognition, Feedback and Training System for T'ai Chi Gestures. Technischer Report 426, Massachusetts Institute of Technology, Media Laboratory, Perceptual Computing Section, Cambridge, Massachusetts, USA, 1996.

[BL97] J.E. Boyd und J.J. Little. Global versus Structured Interpretation of Motion: Moving Light Displays. In *Proc. IEEE Workshop on Motion of Non-Rigid and Articulated Objects (NAM)*, Seiten 18–25, San Juan, Puerto Rico, Juni 1997. IEEECSP.

[Bre97] C. Bregler. Learning and Recognizing Human Dynamics in Video Sequences. In *Computer Vision and Pattern Recognition (CVPR)*, Seiten 568–574, San Juan, Puerto Rico, 1997. IEEE Computer Society Press.

[BT97a] F. Brémond und M. Thonnat. Recognition of scenarios describing human activities. In *Proc. of the International Workshop on Dynamic Scene Recognition from Sensor Data*, Toulouse, Frankreich, Juni 1997.

[BT97b] F. Brémond und M. Thonnat. Tracking multiple non-rigid objects in a cluttered scene. In *Proc. of the Scandinavian Conference in Image Analysis (SCIA'97)*, Lappeenranta, Finnland, Juni 1997.

[CA96] Q. Cai und J.K. Aggarwal. Tracking Human Motion Using Multiple Cameras. In *13th International Conference on Pattern Recognition (ICPR)*, Volume III, Track C: Applications and Robotic Systems, Seiten 68–72, Wien, Österreich, August 1996. IEEE Computer Society Press.

[CA98] Q. Cai und J.K. Aggarwal. Automatic tracking of human motion in indoot sceenes across multiple synchronized video streams. In *Sixth International Conference on Computer Vision (ICCV)*, Seiten 356–362, Bombay, Indien, Januar 1998. IEEE Computer Society Press.

[CC98] K. Cho und H. Cho. Erkennung von Körperbewegungen durch Automaten. In P. Levi, R.-J. Ahlers, F. May und M. Schanz, Hrsg., *Mustererkennung*, Informatik aktuell, Seiten 322–329. Deutsche Arbeitsgemeinschaft für Mustererkennung (DAGM), Springer-Verlag, Berlin, Heidelberg, New York, 1998.

[CHG98] Y. Cui, Q. Huang und M. Greiffhagen. Indoor Monitoring Via the Collaboration Between a Peripheral Sensor and a Foveal Sensor. In *Proc. IEEE Workshop on Visual Surveillance (VS)*, Seiten 2–9, Bombay, Indien, Januar 1998. IEEE Computer Society Press.

[D+97] A.M. DiGioia et al. HipNav: Pre-operative Planning and Intra-operative Navigational Guidance for Acetabular Implant Placement in Total Hip Replacement Surgery. Technischer Report, Center for Orthopaedic Research, Shadyside Hospital, Pittsburgh, 1997.

[Dau97] J. Daugman. Face and Gesture Recognition: Overview. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 19(7):675–678, Juli 1997.

[DF99] Q. Delamarre und O. Faugeras. 3D Articualted Models and Multi-View Tracking with Silhoutettes. In *7th International Conference on Computer Vision (ICCV)*, Band II, Seiten 716–721, Korfu, Griechenland, September 1999. IEEE Computer Society Press.

[DGW+98] T. Darrell, G. Gordon, J. Woodfill, H. Baker und M. Harville. Robust, real-time people tracking in open environments using integrated stereo, color, and face detection. In *Proc. IEEE Workshop on Visual Surveillance (VS)*, Seiten 26–32, Bombay, Indien, Januar 1998. IEEE Computer Society Press.

[Ebe99] M. Ebersberger. Hintergrundschätzer für eine aktive (Schwenk-/Neige-) Kamera. Fortgeschrittenenpraktikum, Technische Universität München, Forschungsgruppe Bildverstehen (FG BV) / Informatik IX, Februar 1999.

[Edw97] A.D.N. Edwards. Progress in Sign Language Recognition. In *Proc. Int. Gesture Workshop*, Nummer 1371 in Lecture Notes in Artificial Inteligence, Seiten 13–21. Springer-Verlag, Berlin, Heidelberg, New York, 1997.

[EKR98] St. Eickeler, A. Kosmala und G. Rigoll. Hidden Markov Model Based Continous Online Gesture Recognition. In *14th International Conference on Pattern Recognition (ICPR)*, Band 2, Seiten 1206–1208, Brisbane, Australien, August 1998. IEEE Computer Society Press.

[ELMG+93] W. Eckstein, G. Lohmann, U. Meyer-Gruhl, R. Riemer, L.A. Robles und J. Wunderwald. Benutzerfreundliche Bildanalyse mit HORUS: Gegenwärtiger Stand und Weiterentwicklungen. In S.J. Poeppl, Hrsg., *Mustererkennung*, Informatik aktuell, Seiten 332–340. Deutsche Arbeitsgemeinschaft für Mustererkennung (DAGM), Springer-Verlag, Berlin, Heidelberg, New York, 1993.

[ES96] W. Eckstein und C. Steger. Interactive Data Inspection and Program Development for Computer Vision. In *Visual Data Exploration and Analysis III*, Band 2656 der SPIE Proceedings, Seiten 296–309. SPIE - The Intern. Soc. for Optical Engineering, 1996.

[ES97] W. Eckstein und C. Steger. Architecture for Computer Vision Application Development within the HORUS System. *Journal of Electronic Imaging*, 6(2):244–261, April 1997.

[Gav99] D. M. Gavrila. The Visual Analysis of Human Movement: A Survey. *Computer Vision and Image Understanding*, 73(1):82–98, Januar 1999.

[GD95] D. M. Gavrila und L. S. Davis. 3-D model-based tracking of human upper body movement: a multi-view approach. In *Proceedings International Symposium on Computer Vision*, Seiten 253–258, Coral Gables, Florida, USA, 1995. IEEE Computer Society Press.

[Geu94] H. Geuß. *Entwicklung eines anthropometrischen Meßverfahrens für das CAD-Menschmodell RAMSIS*. Dissertation, Lehrstuhl für Ergonomie der Technischen Universität München, 1994.

[GLP84] W. E. L. Grimson und T. Lozano-Perez. Model-based recognition and localization from sparse range or tactile data. *Int. J. Robotics Research*, 3(3):3–35, 1984.

[GLP87] W. E. L. Grimson und T. Lozano-Perez. Localizing Overlapping Parts by Searching the Interpretation Tree. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 9(4):469–482, Juli 1987.

[GP99] D.M. Gavrila und V. Philomin. Real-Time Object Detection for “Smart” Vehicles. In *7th International Conference on Computer Vision (ICCV)*, Band I, Seiten 87–93, Korfu, Griechenland, September 1999. IEEE Computer Society Press.

[Gri90a] W. E. L. Grimson. *Object Recognition by Computer – The Role of Geometric Contraints*. The MIT Press, Cambridge, London, 1990.

[Gri90b] W. E. L. Grimson. Object Recognition by Constrained Search. In H. Freeman, Hrsg., *Machine Vision for Three Dimensional Scenes*, Seiten 73–108. Academic Press, San Diego, New York, Boston, London, Sydney, Tokyo, 1990.

[Gwo97] M. Gwozdek. *Lexikon der Video-Überwachungstechnik*. W-&-S-Praxiswissen. Hüthig Verlag, Heidelberg, 1997.

[Hab91] P. Haberäcker. *Digitale Bildverarbeitung*. Carl Hanser Verlag, München, Wien, 4. Auflage, 1991.

[Haf99] W. Hafner. *Segmentierung von Video-Bildfolgen durch Adaptive Farbklassifikation*. Informatik. Herbert Utz Verlag, München, 1999. Vollständiger Abdruck der von der Fakultät für Informatik der Technischen Universität München genehmigten Dissertation.

[HCHD99] I. Haritaoglu, R. Cutler, D. Harwood und L.S. Davis. Backpack: Detection of People Carrying Objects Using Silhouettes. In *7th International Conference on Computer Vision (ICCV)*, Band I, Seiten 102–107, Korfu, Griechenland, September 1999. IEEE Computer Society Press.

[HE95] Kai Huggle und Wolfgang Eckstein. Extraktion von Personen in Videobildern. In Gerhard Sagerer, Stefan Posch und Franz Kummert, Hrsg., *Mustererkennung*, Informatik aktuell, Seiten 134–144. Deutsche Arbeitsgemeinschaft für Mustererkennung (DAGM), Springer-Verlag, Berlin, Heidelberg, New York, 1995.

[HHD98] I. Haritaoglu, D. Harwood und L.S. Davis. Ghost: A Human Body Part Labeling System Using Silhouettes. In *14th International Conference on Pattern Recognition (ICPR)*, Band 1, Seiten 77–82, Brisbane, Australien, August 1998. IEEE Computer Society Press.

[HKM95] W. Hafner, H. Kirchner und O. Munkelt. Farbklassifikation im Projekt STABIL. In V. Rehrmann, Hrsg., *1. Workshop Farbbildverarbeitung*, Seiten 25–29, Koblenz, Oktober 1995. Universität Koblenz-Landau, Fachberichte Informatik 15/95.

[HLZ97] A. Hauck, St. Lanser und Ch. Zierl. Hierarchical Recognition of Articulated Objects from Single Perspective Views. In *Computer Vision and Pattern Recognition (CVPR)*, Seiten 870–876. IEEE Computer Society Press, 1997.

[HM96] W. Hafner und O. Munkelt. Using Color for Detecting Persons in Image Sequences. In *4-th Open German-Russian Workshop on Pattern Recognition and Image Understanding*, Seiten 74–75, Valday, Russia, März 1996. The russian academy of sciences.

[HM97] W. Hafner und O. Munkelt. Using Color for Detecting People in Image Sequences. *Pattern Recognition and Image Analysis*, 7(1):47–52, 1997.

[Hog83] D. Hogg. Model-based vision: a program to see a walking person. *Image and Vision Computing*, 1(1):5–20, 1983.

[Hog87] D. Hogg. Finding a known object using a generate and test strategie. In I. Page, Hrsg., *Parallel Architectures and Computer Vision*, Seiten 119–132. Oxford Science Publications, 1987.

[HOW94] Y. Hel-Or und M. Werman. Constraint-Fusion for Interpretation of Articulated Objects. In *Computer Vision and Pattern Recognition (CVPR)*, Seiten 39–45. IEEE Computer Society Press, 1994.

[HOW96] Y. Hel-Or und M. Werman. Constraint Fusion for Recognition and Localization of Articulated Objects. *Int. J. of Computer Vision*, 19(1):5–28, Juli 1996.

[HR95] G. Herzog und K. Rohr. Integrating Vision and Language: Towards Automatic Description of Human Movements. In I. Wachsmuth, C.-R. Rollinger und W. Brauer, Hrsg., *Proc. 19th Conf. on Artifical Intelligence, KI-95*, Nummer 981 in Lecture Notes in Artificial Inteligence, Seiten 259–268, Bielefeld, 1995. Springer-Verlag, Berlin, Heidelberg, New York.

[HS92] R. M. Haralick und K. G. Shapiro. *Computer and Robot Vision*. Band 1. Addison-Wesley Publishing Company, 1992.

[IB98] M. Isard und A. Blake. ICONDENSATION: Unifying low-level and high-level tracking in a stochastic framework. In H. Burghardt und B. Neumann, Hrsg., *5th European Conference on Computer Vision (ECCV)*, Nummer 1406 in Lecture Notes in Computer Science, Seiten 893–908. Springer-Verlag, Berlin, Heidelberg, New York, 1998.

[IOTS99] S. Iwasawa, J. Ohya, K. Takahashi und T. Sakaguchi. Real-time, 3D Estimation of Human Body Postures from Trinocular Images. In *Proc. IEEE Workshop on Modelling People (MPEOPLE)*, Seiten 3–10, Korfu, Griechenland, September 1999. IEEE Computer Society Press.

[Joh73] G. Johansson. Visual perception of biological motion and a model for its analysis. *Perception & Psychophysics*, 14(2):201–211, 1973.

[JW97] S. Jung und K. Wohn. Tracking and Motion Estimation of the Articulated Object: a Hierarchical Kalman Filter Approach. *Real-Time Imaging*, (3):415–432, 1997.

[Kal60] R. E. Kalman. A New Approach for Linear Filtering and Prediction Problems. *Transactins of the ASME - Journal of Basic Engineering*, Seiten 35–45, März 1960.

[Kap92a] I. A. Kapandji. *Obere Extremität*. Nummer 1 in Funktionelle Anatomie der Gelenke. Ferdinand Enke Verlag, Stuttgart, 2. Auflage, 1992.

[Kap92b] I. A. Kapandji. *Rumpf und Wirbelsäule*. Nummer 3 in Funktionelle Anatomie der Gelenke. Ferdinand Enke Verlag, Stuttgart, 2. Auflage, 1992.

[Kap92c] I. A. Kapandji. *Untere Extremität*. Nummer 2 in Funktionelle Anatomie der Gelenke. Ferdinand Enke Verlag, Stuttgart, 2. Auflage, 1992.

[KGTH94] C. Kambhamettu, D. B. Goldgof, D. Terzopoulos und T. S. Huang. Nonrigid Motion Analysis. In T. Y. Young, Hrsg., *Handbook of Pattern Recognition and Image Processing*, Band 2: Computer Vision, Kapitel 11, Seiten 406–431. Academic Press, San Diego, New York, Boston, London, Sydney, Tokyo, 1994.

[Kin94] W. Kinzel. *Präattentive und attentive Bildverarbeitungsschritte zur visuellen Erkennung von Fußgängern*. Nummer 329 in Fortschr.-Ber. VDI Reihe 10. VDI-Verlag, Düsseldorf, 1994.

[KvB90] K.-P. Karmann und A. v. Brandt. Moving Object Recognition Using an Adaptive Background Memory. In V. Cappellini, Hrsg., *Time-varying Image Processing and Moving Object Recognition, 2*, Seiten 297–307. Elsevier Science Publishers B.V., Amsterdam, The Netherlands, 1990.

[Lan98] St. Lanser. *Modellbasierte Lokalisation gestützt auf monoculare Videobilder*. Berichte aus der Informatik. Shaker Verlag, Aachen, 1998. Vollständiger Abdruck der von der Fakultät für Informatik der Technischen Universität München genehmigten Dissertation.

[Len87] R. Lenz. Linsenfehlerkorrigierte Eichung von Halbleiterkameras mit Standardobjektiven für hochgenaue 3D-Messungen in Echtzeit. In E. Paulus, Hrsg., *Mustererkennung*, Informatik-Fachberichte 149, Seiten 212–216. Deutsche Arbeitsgemeinschaft für Mustererkennung (DAGM), Springer-Verlag, Berlin, Heidelberg, New York, 1987.

[LY95] M. K. Leung und Y.-H. Yang. First Sight: A Human Body Outline Labeling System. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 17(4):359–377, April 1995.

[LZ96] S. Lanser und C. Zierl. On the Use of Topological Constraints within Object Recognition Tasks. In *13th International Conference on Pattern Recognition (ICPR)*, Volume I, Track A: computer Vision, Seiten 580–584, Wien, Österreich, August 1996. IEEE Computer Society Press.

[LZB95] St. Lanser, Ch. Zierl und R. Beutlhauser. Multibildkalibrierung einer CCD-Kamera. In G. Sagerer, S. Posch und F. Kummert, Hrsg., *Mustererkennung*, Informatik aktuell, Seiten 481–491. Deutsche Arbeitsgemeinschaft für Mustererkennung (DAGM), Springer-Verlag, Berlin, Heidelberg, New York, 1995.

[MK96] O. Munkelt und H. Kirchner. STABIL: A SYSTEM FOR MONITORING PERSONS IN IMAGE SEQUENCES. In *Image and Video Processing IV*, Nummer 2666 in SPIE Proceedings, Seiten 163–179. SPIE - The Intern. Soc. for Optical Engineering, 1996.

[MN78] D. Marr und H.K. Nishihara. Representation and recognition of the spatial organization of three-dimensional shapes. *Proceedings of the Royal Society London*, B-200:269–294, 1978.

[MRHH98] O. Munkelt, Ch. Ridder, D. Hansel und W. Hafner. A model driven 3D image interpretation system applied to person detection in video images. In *14th International Conference on Pattern Recognition (ICPR)*, Band 1, Seiten 70–73, Brisbane, Australien, August 1998. IEEE Computer Society Press.

[Mun96] O. Munkelt. *Erkennung von Objekten in Einzelvideobildern mittels Aspektbäumen*. Nummer 125 in Dissertationen zur künstlichen Intelligenz. infix, Sankt Augustin, 1996. Vollständiger Abdruck der von der Fakultät für Informatik der Technischen Universität München genehmigten Dissertation.

[Mur67] M. P. Murray. Gait as a total pattern of movement. *American J. of Physical Medicine*, 46(1):290–332, 1967.

[N+94] L.P. Nolte et al. A Novel Approach to Computer Assisted Spine Surgery. In *Proc. 1st International Symposium on Medical Robotics and Computer Assisted Surgery (MRCAS)*, Seiten 323–328, 1994.

[Nak98] R. Nakatsu. Nonverbal Information Recognition and Its Application to Communocations. In *International Conference on Automatic Face and Gesture Recognition (FG)*, Seiten 2–7, Nara, Japan, April 1998. IEEE Computer Society Press.

[NKI98] A. Nakazawa, H. Kata und S. Inokuchi. Human Tracking Using Distributed Vision Systems. In *14th International Conference on Pattern Recognition (ICPR)*, Band 1, Seiten 593–596, Brisbane, Australien, August 1998. IEEE Computer Society Press.

[OB80] J. O'Rourke und N. I. Badler. Model-Based Image Analysis of Human Motion Using Constraint Propagation. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2(6):522–535, 1980.

[OG99] E.-J. Ong und S. Gong. Tracking Hybrid 2D-3D Human Models from Multiple Views. In *Proc. IEEE Workshop on Modelling People (MPEOPLE)*, Seiten 11–18, Korfu, Griechenland, September 1999. IEEE Computer Society Press.

[OKS80] Y.-I. Ohta, T. Kanade und T. Sakai. Color Information for Region Segmentation. *Computer Graphics and Image Processing*, (13):222–241, 1980.

[OYYT98] Y. Onoe, N. Yokoya, K. Yamazawa und H. Takemura. Visual Surveillance and Monitoring System Using an Omnidirectional Video Camera. In *14th International Conference on Pattern Recognition (ICPR)*, Band 1, Seiten 588–592, Brisbane, Australien, August 1998. IEEE Computer Society Press.

[PBA98] P. Peixoto, J. Batista und H. Araújo. A Surveillance System Combining Peripheral and Foveated Motion Tracking. In *14th International Conference on Pattern Recognition (ICPR)*, Band 1, Seiten 574–577, Brisbane, Australien, August 1998. IEEE Computer Society Press.

[PFD99] R. Plänkers, P. Fua und N. D'Apuzzo. Automated Body Modeling from Video Sequences. In *Proc. IEEE Workshop on Modelling People (MPEOPLE)*, Seiten 44–52, Korfu, Griechenland, September 1999. IEEE Computer Society Press.

[PHZ96] J. Ponce, M. Hebert und A. Zisserman. Report on the 1996 International Workshop on Object Representation in Computer Vision. In *Proc. Int. Workshop on Object Representation in Computer Vision*, Nummer 1144 in Lecture Notes in Computer Science, Seiten 1–8. Springer-Verlag, Berlin, Heidelberg, New York, April 1996.

[Pop94] A.R. Pope. Model-Based Object Recognition. Technical Report TR-94-04, University of British Columbia, Januar 1994.

[Rau99] I. Rauschert. 3D Positionsvorhersage und -schätzung zur Merkmalsverfolgung. Systementwicklungsprojekt, Technische Universität München, Forschungsgruppe Bildverstehen (FG BV) / Informatik IX, November 1999.

[Ric95] S. Richter. *Ein mehrfach adaptierendes, stabiles Modell zur Analyse von Straßenszene*. Nummer 106 in Dissertationen zur künstlichen Intelligenz. infix, Sankt Augustin, 1995. Vollständiger Abdruck der von der Fakultät für Informatik der Technischen Universität München genehmigten Dissertation.

[RMK95] Ch. Ridder, O. Munkelt und H. Kirchner. Adaptive Background Estimation and Foreground Detection using Kalman-Filtering. In O. Kaynak, M. Özkan, N. Bekiroğlu und İ. Tunay, Hrsg., *Proceedings of International Conference on recent Advances in Mechatronics, ICRAM'95*, Seiten 193–199, Boğaziçi University, 80815 Bebek, Istanbul, TURKEY, August 1995. UNESCO Chair on Mechatronics.

[RMR+99] B. Radig, O. Munkelt, Ch. Ridder, D. Hansel und W. Hafner. A Model-Driven Three-Dimentional Image-Interpretation System Applied to Person Detection in Video Images. In B. Jähne, H. Haußecker und P. Geißler, Hrsg., *Handbook of computer vision and applications*, Band 3: Systems and Applications, Kapitel 22. Academic Press, San Diego, New York, Boston, London, Sydney, Tokyo, 1999.

[Roh93] K. Rohr. Incremental Recognition of Pedestrians from Image Sequences. In *Computer Vision and Pattern Recognition (CVPR)*, Seiten 8–13. IEEE Computer Society Press, 1993.

[Roh97] K. Rohr. Human Movement Analysis based on explicit Motion Models. In M. Shah und R. Jain, Hrsg., *Motion-Based Recognition*, Seiten 171–198. Kluver Academic Publishers, Dordrecht Boston, 1997.

[SB96] H.-J. Siegert und S. Bocionek. *Robotic: Programmierung intelligenter Roboter*. Springer-Verlag, Berlin, Heidelberg, New York, 1996.

[Sei94] A. Seidl. *Das Menschmodell RAMSIS - Analyse, Synthese und Simulation dreidimensionaler Körperhaltungen des Menschen*. Dissertation, Lehrstuhl für Ergonomie der Technischen Universität München, 1994.

[Sei99] P. Seitz. Solid-State Image Sensing. In B. Jähne, H. Haußecker und P. Geißler, Hrsg., *Handbook of computer vision and applications*, Band 1: Sensors and Imaging, Kapitel 7. Academic Press, San Diego, New York, Boston, London, Sydney, Tokyo, 1999.

[SF96] G. Schmidt und F. Freyberger, Hrsg. *Autonome Mobile Systeme 1996*, Informatik aktuell. Gesellschaft für Informatik (GI), Springer-Verlag, Berlin, Heidelberg, New York, 1996.

[Sta99] R. Stahl. Untersuchungen und Erweiterungen des Kalman - Hintergrundschätzers: – Modellierung der Grauwertänderungen, – Berücksichtigung benachbarter Bildpunkte, – Zeitabhängigkeit des Filterverhaltens. Systementwicklungsprojekt, Technische Universität München, Forschungsgruppe Bildverstehen (FG BV) / Informatik IX, November 1999.

[Sti96] Ch. Stimmelmayr. Hintergrundschätzung zur Bewegungsdetektion in Farbbildfolgen. Fortgeschrittenenpraktikum, Technische Universität München, Forschungsgruppe Bildverstehen (FG BV) / Informatik IX, Juni 1996.

[TKBM99] K. Toyama, J. Krumm, B. Brumitt und B. Meyers. Wallflower: Principles and Practice of Background Maintenance. In *7th International Conference on Computer Vision (ICCV)*, Band I, Seiten 255–261, Korfu, Griechenland, September 1999. IEEE Computer Society Press.

[Wac97] St. Wachter. *Verfolgung von Personen in monokularen Bildfolgen*. Vice Versa Verlag, Berlin, 1997.

[WADP97] Ch. R. Wren, A. Azarbayejani, T. Darrell und A. P. Pentland. Pfinder: Real-Time Tracking of the Human Body. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 19(7):780–785, Juli 1997.

[Wex97] A. Wexelblat. Research Challenges in Gesture: Open Issues and Unsolved Problems. In *Proc. Int. Gesture Workshop*, Nummer 1371 in Lecture Notes in Artificial Inteligence, Seiten 1–11. Springer-Verlag, Berlin, Heidelberg, New York, 1997.

[WH96] P. Wunsch und G. Hirzinger. Registration of CAD-Models to Images by Iterative Inverse Perspective Matching. In *13th International Conference on Pattern Recognition (ICPR)*, Volume III, Track C: Applications and Robotic Systems, Seiten 78–83, Wien, Österreich, August 1996. IEEE Computer Society Press.

[WM96] T. Wada und T. Matsumyama. Appearance Sphere: Background Model for Pan-Tilt-Zoom Camera. In *13th International Conference on Pattern Recognition (ICPR)*, Volume I, Track A: computer Vision, Seiten 718–722, Wien, Österreich, August 1996. IEEE Computer Society Press.

[WN97] St. Wachter und H.-H. Nagel. Tracking of Persons in Monocular Image Sequences. In *Proc. IEEE Workshop on Motion of Non-Rigid and Articulated Objects (NAM)*, Seiten 2–9, San Juan, Puerto Rico, Juni 1997. IEEE Computer Society Press.

[WP99] C.R. Wren und A.P. Pentland. Understanding Purposeful Human Motion. In *Proc. IEEE Workshop on Modelling People (MPEOPLE)*, Seiten 19–25, Korfu, Griechenland, September 1999. IEEE Computer Society Press.

[Wun98] P. Wunsch. *Modellbasierte 3-D Objektlageschätzung für visuell geregelte Greifvorgänge in der Robotik*. Berichte aus der Informatik. Shaker Verlag, Aachen, 1998. Vollständiger Abdruck der von der Fakultät für Informatik der Technischen Universität München genehmigten Dissertation.

[XZ96] G. Xu und Z. Zhang. *Epipolar Geometry in Stereo, Motion and Object Recognition*. Kluwer Academic Publishers, Dordrecht, The Netherlands, 1996.

[YTBH98] Y. Ye, J.K. Tsotsos, K. Bennet und E. Harley. Tracking a Person with Prerecorded Image Database and a Pan, Tilt, and Zoom Camera. In *Proc. IEEE Workshop on Visual Surveillance (VS)*, Seiten 10–17, Bombay, Indien, Januar 1998. IEEE Computer Society Press.

Index