

Sichere Mensch-Roboter-Kooperation durch Auswertung von Bildfolgen

zur Erlangung des akademischen Grades eines

Doktors der Ingenieurwissenschaften

der Fakultät für Informatik
des Karlsruher Instituts für Technologie (KIT)

vorgelegte

D i s s e r t a t i o n

von

Dipl.–Inform. Jürgen Graf

aus Memmingen

Tag der mündlichen Prüfung:	18. Mai 2010
Erster Gutachter:	Prof. Dr.–Ing. Heinz Wörn
Zweiter Gutachter:	Prof. Dr.–Ing. Rüdiger Dillmann

Vorwort

Die hier vorliegende Arbeit entstand während meiner wissenschaftlichen Tätigkeit als Doktorand am Institut für Prozessrechentechnik, Automation und Robotik der Fakultät für Informatik des Karlsruher Instituts für Technologie (KIT).

Mein besonderer Dank gilt Herrn Prof. Dr.–Ing. Heinz Wörn für die Betreuung der Arbeit und seinen unermüdlichen Einsatz um den Forschungsbetrieb am Lehrstuhl voran zu bringen. Außerdem danke ich Herrn Prof. Dr.–Ing. Rüdiger Dillmann für die Übernahme des Korreferats.

Meine besondere Wertschätzung möchte ich auch meinem ehemaligen Diplomarbeitsbetreuer und Lehrer Herrn Prof. Dr.rer.nat. (em.) Hans–Hellmut Nagel aussprechen, der mich durch seine leidenschaftlichen Vorlesungen über Bildfolgenauswertung und Kognitive Sichtsysteme in diese faszinierenden Forschungsgebiete eingeführt hat.

Ich bedanke mich außerdem bei meinen Studenten, die in der Vergangenheit durch ihre Teilnahme an den von mir angebotenen Aufgaben tatkräftig die Arbeit mitunterstützt haben.

Weiterhin bedanke ich mich in besonderer Weise bei Heike Schmidt für Ihre großartige Unterstützung.

Schließlich bedanke ich mich bei meinen Eltern und Großeltern, die mir mein Studium ermöglicht haben. Ganz besonderer Dank gebührt dabei meiner Mutter Anna Graf für Ihre guten und weisen Ratschläge und die Tatsache, dass Sie einer der geduldigsten Zuhörerinnen ist, die ich kenne.

Inhaltsverzeichnis

1	Einleitung	1
1.1	Motivation	1
1.2	Problemstellung und Zielsetzung	2
1.3	Gliederung der Arbeit	3
2	Stand der Technik und Forschung	5
2.1	Sensoren	5
2.1.1	Optische Systeme	6
2.1.2	Kompakte optische Systeme zur Messung von Distanzen	7
2.1.3	Nicht optische Systeme für die Mensch-Roboter-Kooperation	10
2.1.4	Das Lichtflugzeit-Prinzip und eine technologische Realisierung	12
2.2	Ausgewählte Forschungsgruppen	15
2.2.1	Industrielle Forschung und Entwicklung	16
2.2.2	Forschungseinrichtungen und universitäre Forschung	17
2.3	Normen und Vorschriften	28
2.4	Abschließende Bemerkung und Diskussion	32
3	Systemarchitektur	35
3.1	Entwurf einer ersten Systemarchitektur	35
3.2	Entwurf einer zweiten Systemarchitektur	38
3.2.1	MAROCO - Ein Softwaresystem für die sichere Mensch-Roboter-Kooperation	39
4	Bildfolgenauswertung	47
4.1	Vorverarbeitung	48
4.1.1	Anisotrope Diffusion	51
4.1.2	Effiziente bilaterale Filterung	55
4.1.3	Orientierte Medianfilterung	58
4.1.4	Experimente und Diskussion	59
4.2	Hintergrundmodellierung	63
4.2.1	Einleitung	63
4.2.2	Gaußsche Mischmodelle zur Hintergrundschätzung	63
4.2.3	Parameterbestimmung und -analyse der gaußschen Mischmodelle	67
4.2.4	Experimente und Diskussion	72
4.2.5	Zusammenfassung	78
4.3	Merkmale für die Bewegungsschätzung	78

4.3.1	Grundlagen zur Schätzung des Optischen Flusses	79
4.3.2	Ein Verfahren zur Bestimmung von dynamischen Konturen	85
4.4	Ein OF-Schätzverfahren mit anisotropem Datenterm	92
4.4.1	Grundlagen	95
4.4.2	Die XCLG-Methode	97
4.4.3	Effiziente Lösung durch Mehrgitter-Verfahren	105
4.4.4	3D Optischer Fluss	115
4.4.5	Experimente und Diskussion	118
5	Robuste Merkmale zur Kinematik- und Risikoschätzung	127
5.1	Merkmale für die Risikoquantifizierung	127
5.1.1	Lokalisation von menschlichen Körpern	128
5.1.2	Lage von Kopf und Schulter	128
5.1.3	Kinematische Parameter der Arme	131
5.2	Robuste Merkmale	136
5.2.1	Ein Kalman-Filter für Kinematikkonfigurationen	139
5.2.2	Projektion von Kugelmodellen auf die Bildebene	152
5.2.3	Experimente und Bewertung	155
6	Risikoschätzung	159
6.1	Unschärfe Logiken für die Risikoquantifizierung	159
6.2	Grundlagen	160
6.3	Eine adaptierte Realisierung für MAROCO	161
6.3.1	Diskussion	166
7	Risikominimierung	169
7.1	Arbeitsgeschwindigkeit	169
7.2	Eine Bahnplanung für die sichere MRK	169
7.2.1	Grundlagen des Verfahrens	169
7.2.2	Experimente und Diskussion	173
8	Zusammenfassung und Ausblick	183
8.1	Einordnung	183
8.2	Ausblick	185
	Abbildungsverzeichnis	187
	Tabellenverzeichnis	191
	Literaturverzeichnis	192

Kapitel 1

Einleitung

Industrierobotik ist ein interdisziplinäres Feld und spielt eine Schlüsselrolle in der Automatisierungstechnik. Zum heutigen Zeitpunkt werden Industrieroboter vorzugsweise in Bereichen eingesetzt, die eine ausdauernde, monotone und präzise Arbeitsweise voraussetzen. Typische Aufgaben sind bspw.:

- Handhabung (z.B. vereinzeln, sortieren, be- und entladen).
- Montage (z.B. schrauben, einfädeln).
- Bearbeitung (z.B. schweißen, kleben, polieren).
- Qualitätskontrolle (z.B. Schweißnahtkontrolle, Oberflächeninspektion).

Eine Herausforderung ist die Integration von Sensoren in die Roboterzelle und die Implementierung der dazugehörigen Auswerteverfahren. Durch die Integration von Sensorsystemen soll es den Manipulatoren ermöglicht werden, weitestgehend autonom Probleme zu lösen. Ein wichtiger Trend in der Forschung und Entwicklung komplexer Robotersysteme liegt darin, eine zuverlässige kognitive Industrierobotik zu entwickeln.

In der Vergangenheit haben sich aus den Anforderungen der Industrierobotik eigenständige Forschungsgebiete, wie bspw. das Maschinen- bzw. das Robotersehen, etabliert.

Die hohen Erwartungen an die Auswertung und Interpretation von Sensorsignalen, gewonnen aus in den Arbeitsbereich der Roboter integrierten Sensoren, haben bis heute nicht in allgemeiner und umfassender Form zu befriedigenden Ergebnissen geführt. Dies gilt insbesondere für die sichere Mensch-Roboter-Kooperation.

Erwartet wird ein einfach zu installierender und flexibel anpassbarer Aufbau von kommerziell nutzbaren Zellen, um damit z.B. eine Kooperation von Mensch und Maschine möglich zu machen.

Das daraus resultierende Forschungsgebiet wird als Mensch-Roboter-Kooperation bezeichnet. Eine wichtige Anforderung an die Kooperation liegt darin, die Sicherheit für den Menschen während der Kooperation zu gewährleisten. Die hier vorliegende Arbeit will einen Beitrag leisten, damit dieses Ziel in naher Zukunft sicher und effizient realisiert wird.

1.1 Motivation

Möchte man kognitive Fähigkeiten in die Industrierobotik integrieren, um damit das Risiko für den Menschen zu minimieren, muss man sich Gedanken darüber machen, welche Art von

Sensorik für den Aufbau verwendet werden soll. Ein derartiges System muss Signale aus der Roboterzelle wahrnehmen und idealerweise interpretieren können, deshalb sollte mindestens die Auswertung der durch Sensoren aufgezeichneten Signale gewährleistet sein.

Der Aufbau eines solchen Systems kann realisiert werden, indem ein Auswertezyklus entworfen wird, welcher alle beteiligten Entitäten (Mensch, Roboter, etc.) berücksichtigt.

Die hier vorliegende Arbeit will sich auf ein vergleichsweise neuartiges, 3D optisches Sensorsystem konzentrieren. Das optische System basiert auf dem Lichtflugprinzip und ist in der Lage, neben einer 2D Abbildung der Szene, für einzelne Bildpunkte den Abstand zur Szene zu bestimmen.

Der Vorteil bei der Integration eines einzelnen Sensors zur Arbeitsraumüberwachung liegt in der Platzersparnis gegenüber den Schutzzäunen und der daraus resultierenden Aufhebung der räumlichen und zeitlichen Trennung von Mensch und Roboter. Durch die Aufhebung wird eine Grundlage für eine effiziente Mensch–Roboter–Kooperation geschaffen. Dabei setzen sich Synergieeffekte frei, welche die Stärken der Menschen und die Fähigkeiten der Roboter besser gemeinsam nutzen lassen. Die möglichen Anwendungsszenarien sind dabei nicht auf rein industrielle Anwendungsgebiete beschränkt. Auch im Bereich der Medizinrobotik können Systeme mit optischer Arbeitsraumüberwachung helfen, zukünftig optimale Ergebnisse durch robotergestützte Chirurgie zu erzielen.

Die markerlose Erkennung von kinematischen Parametern des Menschen ist ebenfalls eine große Herausforderung, welche es ermöglicht, genauer und umfassender das Risiko zu bestimmen, welches sich im industriellen Umfeld, z.B. durch die Dynamik der eingesetzten Industrieroboter, ergibt. Es wird von vornherein gefordert, keine markerbasierten Systeme für diese Aufgabe zu verwenden, da dafür ein zu großer technischer Aufwand für die Hardwareintegration entsteht und damit auch ein großes Stück an Flexibilität beim Aufbau entsprechender Roboterzellen verloren geht. Die Kenntnis über die Dynamik, mindestens jedoch der Kinematik der Aktoren erlaubt es, eine Grundlage für eine effiziente und sichere Kooperation zwischen den Aktoren einzurichten. Im industriellen Umfeld gibt es derzeit kein System, welches diesem Paradigma folgt.

Die Kenntnisse der kinematischen und dynamischen Konfigurationen der Aktoren bietet die Möglichkeit, eine neuartige Sicht auf die Bahnplanung zu werfen, insbesondere, wenn es um die Sicherheit des Menschen geht. Es ergeben sich multikriterielle Optimierungsprobleme, die effizient gelöst werden müssen. Ein System für die sichere Mensch–Roboter–Kooperation muss mindestens die beiden Anforderungen Sicherheit und Effizienz erfüllen, um sich zukünftig im wirtschaftlichen Wettbewerb etablieren zu können.

Ein weiterer wichtiger Punkt ist die Möglichkeit, die reale Szene virtuell im Rechner abzubilden. Geschieht dies mindestens in Echtzeit, d.h. im Takt des Sensorsystems, lassen sich weitere Anwendungsgebiete erschließen, die für eine Kooperation zwischen Mensch und Maschine wichtig sind. Man kann dadurch bspw. schon frühzeitig bei der Planung von Roboterzellen unterschiedliche Probleme identifizieren und verhindern. Denkbar sind aber auch Anwendungen im Bereich der Telekooperation.

1.2 Problemstellung und Zielsetzung

Das hier vorgestellte System zur Realisierung einer sicheren Mensch–Roboter–Kooperation und Interaktion behandelt insbesondere die folgenden Teilprobleme, welche sich aus den speziellen Herausforderungen von Robotikanwendungen im industriellen Umfeld ergeben:

- Identifikation und Lösung von Anforderungen.
- Markerlose Schätzung von kinematischen Parametern des Menschen.
- Bewegungsschätzung für bestimmte Körperteile.
- Entwicklung und Bewertung von Methoden zur Schätzung des Risikos.
- Entwicklung und Bewertung von risikoreduzierenden Maßnahmen.
- Entwicklung eines im Takt der Kamera laufenden Software-Systems mit Simulationsumgebung.
- Aufbau eines realen Demonstrators für die experimentelle Bewertung.
- Bewertung des 3D-Sensorsystems für das Forschungsgebiet.

1.3 Gliederung der Arbeit

In Kapitel 2 werden zuerst Sensoren vorgestellt, die für den Bereich der sicheren Mensch-Roboter-Kooperation üblicherweise eingesetzt werden. Es kommt zu einer Unterscheidung von allgemeinen optischen Systemen, speziellen optischen Systemen für die Distanzmessung und nicht optischen Systemen. Anschließend wird das Lichtflugzeit-Prinzip und die Photonen-Mischtechnik vorgestellt.

Der danach folgende Abschnitt stellt ausgewählte nationale und internationale Forschungsgruppen vor, die in teils unterschiedlicher Weise Beiträge zur Realisierung sicherer Mensch-Roboter-Systeme geleistet haben.

Danach werden einige wichtige Fakten im Hinblick auf Normen und Vorschriften vorgestellt. Abschließend wird die PMD-Technologie eingeführt und die technischen Grundlagen vorgestellt. Zuletzt werden kurz die Vor- und Nachteile dieser Technologie diskutiert.

In Kapitel 3 wird zuerst eine Systemarchitektur vorgestellt, die einen rein auf Bildfolgenauswertung basierenden Ansatz verfolgt. Dazu wird ein vollautomatisches Verfahren zur Verfolgung von sich bewegenden Entitäten in der Roboterzelle vorgestellt. Das Verfahren ist theoretisch motiviert und basiert auf einem Variationsansatz.

Danach wird die für diese Arbeit wichtige Systemarchitektur vorgestellt. Diese Systemarchitektur basiert auf der Verarbeitung von Merkmalen und verwendet dazu neuartige Modelle für die markerlose Verfolgung von Personen, z.B. durch Integration 3D Optischer Fluss Schätzverfahren in statistische Schätzverfahren. Die Schätzverfahren liefern wichtige Parameter zur Konfiguration eines kinematischen Menschmodells. Darüber hinaus wird eine Bahnplanung integriert. Beide Architekturen können echtzeitfähig, i.S. des Kamerataktes, betrieben werden.

In Kapitel 4 werden die für die Architekturen wichtigen Grundlagen der Bildfolgenauswertung vorgestellt und diskutiert. Aufgrund der Fülle, der im Rahmen dieser Arbeit entwickelten Methoden, wird auf ein Kapitel für die experimentelle Auswertung verzichtet und stattdessen werden direkt nach der Vorstellung der jeweiligen Grundlagen die Experimente diskutiert.

Das Kapitel stellt in den beiden ersten Abschnitten Techniken zur Vorverarbeitung und Hintergrundmodellierung von Tiefenbildfolgen vor. Anschließend werden Merkmale für die Szenerekonstruktion diskutiert. Insbesondere ein Verfahren für die dynamische Verfolgung von Konturzügen und Verfahren zur Schätzung des Optischen Flusses prägen die entsprechenden Unterabschnitte. Im darauf folgenden Abschnitt wird ein anisotroper Datenterm für die

Schätzung des Optischen Flusses eingeführt und in einen Variationsansatz eingebettet. Anschließend werden die zum Variationsansatz korrespondierenden Euler–Lagrange–Gleichungen effizient durch ein an das Verfahren adaptiertes Mehrgitter–Verfahren gelöst. Die Besonderheit liegt darin, dass das Mehrgitter–Verfahren auf modernen Grafikkartenprozessoren angewandt wird, wodurch eine signifikante Reduktion der Berechnungszeit erzielt wurde. Abschließend wird eine Reihe von Experimenten dargestellt und diskutiert.

In Kapitel 5 werden Merkmale für die Rekonstruktion einer menschlichen Kinematik eingeführt. Die Komplexität des intern repräsentierten Menschmodells umfasst dabei 13 Parameter. Anschließend wird ein Rahmenwerk zur systematischen Schätzung von menschlichen Kinematiken vorgestellt. Das Rahmenwerk basiert auf einer im Rahmen dieser Arbeit adaptierten Kalman–Filter Variante. Für die experimentelle Bewertung der Verfahren wurde ein spezielles Hüllkörpermodell entwickelt, welches explizit die Geometrie der Rückprojektion auf die Bildebene durch das Kamerasystem berücksichtigt und zusätzlich für eine effiziente Abstandsberechnung zwischen Mensch und Robotern verwendet werden kann.

In Kapitel 6 wird ein adaptiertes Modell für eine zweisträngige, unscharfe Logik vorgestellt und diskutiert. Die Quantifizierung des Risikos geschieht dabei in Abhängigkeit der Konfiguration der Kinematik des Menschen und der Konfiguration der Roboter.

In Kapitel 7 werden zwei Strategien vorgestellt. Da die Reduktion der Arbeitsgeschwindigkeit ein weit verbreiteter Ansatz ist, wird nur kurz auf die hier verwendete Realisierung eingegangen. Das Hauptaugenmerk liegt auf einer proaktiven Bahnplanung, welche im Rahmen dieser Arbeit entwickelt wurde und ein Risikopotential in jedem Zeitschritt der Kamera auf der Grundlage der Risikoschätzung generiert. Eine experimentelle Bewertung des Verfahrens schließt das Kapitel ab.

Kapitel 2

Stand der Technik und Forschung

In Deutschland müssen technische Systeme genormte Sicherheitsbestimmungen erfüllen. Die hier vorliegende Arbeit berücksichtigt nicht alle heutzutage gültigen Normen für den realen Einsatz. Es wird viel mehr anhand eines Demonstrators eine Vorgehensweise gezeigt, wie eine sichere Mensch–Roboter–Kooperation zu realisieren ist, wobei die Sicherheit des Menschen auf verschiedenartige Weise realisiert wird.

Es existieren unterschiedliche nationale und internationale Forschungsgruppen, Forschungseinrichtungen und Forschungszentren, die an Systemen für die sicheren Mensch–Roboter–Kooperation im industriellen Umfeld arbeiten und damit zur Weiterentwicklung dieser Domäne beitragen.

Die im Folgenden dargestellte Übersicht kann keinen Anspruch auf Vollständigkeit erheben. Vielmehr soll die Übersicht dazu dienen, die hier vorliegende Arbeit in das Forschungsgebiet einzubetten und die in dieser Arbeit vorgestellten Ergebnisse abzugrenzen vom derzeitigen Stand der Forschung. Das ist wichtig, da die später vorgestellten Alleinstellungsmerkmale darin begründet sein können, weil

- verschiedene Probleme auf ähnliche Weise gelöst werden, aber im hier vorgestellten System anders weiterverarbeitet werden;
- Lösungen aus verwandten Problemstellungen an die hier vorliegende Problemstellung angepasst und modifiziert sind;
- Neue Ansätze entwickelt werden, die bei Kenntnis des aktuellen Forschungsstandes besser einzuordnen sind.

2.1 Sensoren

Die in diesem Abschnitt vorgestellten Sensoren werden für das Anwendungsfeld sichere Mensch–Maschine–Kooperation eingesetzt oder dienen einem ähnlichen Forschungs– oder Anwendungsfeld.

Einige der hier vorgestellten Systeme werden sich deshalb im nächsten Abschnitt wiederfinden. Neben den Sensoren werden auch ausgewählte Anwendungen zitiert, um gleichzeitig für die spätere Einordnung der Sensoren und Verfahren zu sensibilisieren.

Da sich die hier vorliegende Arbeit hauptsächlich auf einen optischen Sensor konzentriert, der in der Lage ist, Tiefeninformationen aus der Szene in Echtzeit zu liefern, wird zwischen optischen und nicht-optischen Systemen bewusst unterschieden.

Das menschliche Auge gilt als das Sinnesorgan, durch welches die meisten Informationen bereitgestellt werden. Es wird parallel zu den optischen Systemen daran geforscht und entwickelt, auch die anderen Sinne technisch umzusetzen und weiter zu entwickeln..

Sensoren, basierend auf dem Lichtflugzeit-Prinzip, werden mit großer Wahrscheinlichkeit in zukünftigen Zellen integriert sein und eine Schlüsselrolle für die Informationsgewinnung spielen (siehe dazu [Som 2007, Som 2008]).

2.1.1 Optische Systeme

Ein grundsätzlicher Vorteil optischer Systeme liegt darin, berührungslos und räumlich-zeitlich hochaufgelöst Informationen über die Umwelt zur Verfügung zu stellen. Die kognitiven Fähigkeiten des Menschen erlauben es bspw., ein räumliches Abbild der Szene durch binokulares Sehen zu realisieren. Dabei wird u.a. die unterschiedliche Anordnung der menschlichen Augen ausgenutzt.

In den nächsten Unterabschnitten werden einige technische Systeme vorgestellt, welche ebenfalls in der Lage sind, die Szene räumlich zu erfassen.

Monokulares Sehen

Eine der zum jetzigen Zeitpunkt am häufigsten eingesetzte Technik für die Bildverarbeitung und die Bildfolgenauswertung ist das passive monokulare Sehen. Aufgrund der kostengünstigen Bezugsmöglichkeit wird diese Technologie einer breiten Anwendermasse leicht zugänglich gemacht.

Das in [Wachter und Nagel 1999] vorgestellte Verfahren beschäftigt sich mit der Auswertung monokularer Videobildfolgen, um eine Person zu verfolgen. Im Wesentlichen kommt es zu einer Verfolgung von projizierten Modellkanten einer Hüllkörperbeschreibung des menschlichen Körpers durch Einsatz eines iterierten Kalman-Filters. Der in dieser Arbeit vorgestellte Ansatz ist für den später vorgestellten und für diese Arbeit verwendeten Tiefensensor nicht geeignet, da beim heutigen Stand der Technik an Bildkanten schwerwiegende Störungen auftreten, die bekannt sind unter der Bezeichnung *fliegende Pixel*¹.

Es sei an dieser Stelle darauf hingewiesen, dass eine Tiefenschätzung auch dann gelingen kann, wenn nur mit einer Kamera gearbeitet wird und diese sich bewegt, bzw. eine Bildfolge prozessiert wird, die von einer bewegten Kamera stammt. Eine vielversprechende aktuelle Umsetzung findet man in [Valgerats u.a. 2008]. Diese Methoden sind aber Gegenstand der aktuellen Grundlagenforschung und werden derzeit lediglich auf speziellen Testsequenzen evaluiert. Trotz positiver experimenteller Ergebnisse ist nicht abzusehen, wann derartige Methoden in eine komplexe Systemarchitektur integriert werden und dort robuste Ergebnisse unter realen Bedingungen liefern können.

Binokulares Sehen und Mehrkamerasysteme

Die meisten Systeme, welche auf Triangulationstechniken basieren, sind sogenannte Stereosysteme. Sie werden z.B. dadurch realisiert, dass zwei Kameras in unterschiedlicher Lage montiert werden. Durch eine nachgeschaltete Bildauswertung können aufgrund von Merkmalskorrespondenzen die gewünschten Tiefeninformationen ermittelt werden. Diese Art der Szenerekonstruk-

¹engl. flying pixels

tion ist theoretisch sehr gut untersucht und verstanden (siehe [Hartley und Zisserman 2004], [Schreer 2005], [Ma u.a. 2006]) und bildet den Kernbereich des Computersehens.

Einige aktuelle Verfahren konzentrieren sich auf die Verwendung mehrerer Kameras und verwenden, aufbauend auf [Potmesil 1986], modifizierte Varianten des Silhouetten–Schnitt–Verfahrens. Aktuelle Umsetzungen und Erweiterungen zeigen eine prinzipielle Tauglichkeit der Rekonstruktion menschlicher Bewegung in einem 3D Modell in Echtzeit, d.h. mehr als 20 Bilder pro Sekunde. Ein Vergleich ausgewählter bestehender Verfahren findet man in [Michoud u.a. 2007].

Eine andere Möglichkeit besteht darin, aus mehreren Ansichten Merkmale der Kamera, wie bspw. Farbe, mit zu integrieren. Ein Beispiel hierfür wird in [Nickel 2008] vorgestellt.

2.1.2 Kompakte optische Systeme zur Messung von Distanzen

Tiefen– bzw. Distanzmessungen, basierend auf optischen Systemen, haben eine wichtige Bedeutung für das Computersehen. Im Wesentlichen lassen sich diese optischen Systeme in vier Kategorien einteilen:

1. Stereoskopie
2. Aktive Triangulation
3. Lichtflugzeit basierte Systeme
4. Interferometrie

Die Stereoskopie wurde bereits in Abschnitt 2.1.1 eingeführt.

Angelehnt an die menschliche Physiologie werden mindestens zwei Kameras in allgemeiner Lage montiert, so dass diese einen möglichst großen, gemeinsamen Bereich der Szene abbilden. Um räumliches Sehen zu ermöglichen, werden deshalb sogenannte konvergente Stereosysteme aufgebaut. Dabei schneiden sich die beiden Sehstrahlen bei der Projektion auf eine Ebene.

Aus den intrinsischen und extrinsischen Parametern der Kameras und den Gesetzmäßigkeiten der Epipolargeometrie lässt sich die Szene rekonstruieren.

Die Abbildung 2.1 stellt ein achsparalleles Stereosystem dar. Die beiden Kameras sind horizontal verschoben. Im Gegensatz zu einem konvergenten Stereosystem sind die Koordinatensysteme der beiden Kameras identisch orientiert. Die beiden Bildebenen I_1 und I_2 sind parallel verschoben. Die optischen Zentren C_1 und C_2 sind horizontal verschoben. Die Basislänge B definiert den Abstand der beiden optischen Zentren. Die Brennweiten f_1 und f_2 definiert den Abstand der Brennpunkte zur Bildebene. Es wird im Folgenden von einer identischen Brennweite f ausgegangen. Ein Raumpunkt M wird durch die optischen Zentren der Kameras auf die jeweilige Bildebene projiziert. Die dadurch entstehenden Projektionen von M sind m_1 und m_2 . Die Disparität wird normalerweise in Pixelkoordinaten angegeben und ist bestimmt durch

$$\delta = u_1 - u_2 \quad . \quad (2.1)$$

Definiert man die Breite eines CCD-Elements durch d_u , so erhält man nach Anwendung des Strahlensatzes die Rauntiefe

$$\rho = \frac{B \cdot f}{\delta \cdot d_u} \quad . \quad (2.2)$$

Die Disparität ist deshalb eine Größe, die sich reziprok proportional zur Entfernung verhält. Im Falle eines konvergenten Stereoaufbaus befinden sich die beiden Kameras in allgemeiner Lage.

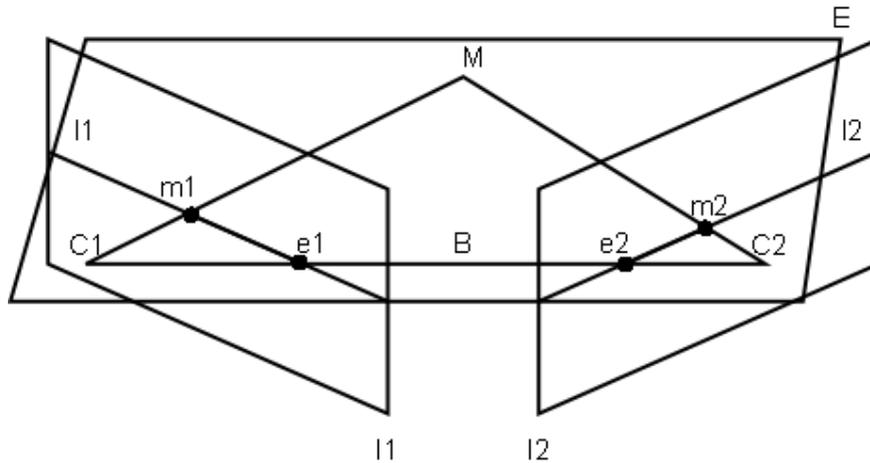


Abbildung 2.2: Darstellung zur Epipolargeometrie.

welche die Epipolarbedingung erfüllen, korrespondierende Bildpunkte sind, gilt im Allgemeinen nicht. Das liegt daran, dass ein Raumpunkt, projiziert entlang des optischen Strahls der einen Kamera, identisch abgebildet wird, in der anderen Kamera zu unterschiedlichen Abbildungen führen kann.

Die bisher beschriebene Transformation zwischen zwei Ansichten berücksichtigt lediglich die extrinsische Beziehung der beiden verwendeten Kameras. Werden zusätzlich die intrinsischen Parameter berücksichtigt, gelangt man zur sogenannten Fundamental-Matrix

$$F = A_2^{-T} E A_1^{-1} \quad . \quad (2.6)$$

Die intrinsische Matrix ist folgendermaßen festgelegt:

$$A = \begin{pmatrix} f \cdot k_u & 0 & u_0 \\ 0 & f \cdot k_v & v_0 \\ 0 & 0 & 1 \end{pmatrix} \quad . \quad (2.7)$$

Die in der instrinsische Matrix A auftauchenden Skalierungsgrößen k_u und k_v stellen die Beziehung zwischen Sensor- und Bildkoordinaten her. Die beiden Skalare u_0 und v_0 sind Verschiebungsgrößen zum Ursprung des Bildkoordinatensystems. Bei der Kamerakalibrierung müssen u.a. diese Parameter geschätzt werden.

Es ergibt sich damit die folgende, vereinfachte Vorgehensweise:

1. Festlegung der Kameraanordnung
2. Kalibrierung der Kameras
3. Detektion von Merkmalen
4. Evtl. Rektifikation der Ansichten
5. Korrespondenzanalyse
6. Tiefenbestimmung

7. Datenaufbereitung durch Interpolation

Die Vorteile bei dieser Vorgehensweise zur Tiefenbestimmung liegen darin, dass der Preis für heutige Kamerasysteme sehr günstig ist. Außerdem sind in der Vergangenheit die theoretischen Grundlagen für die Mehransichtengeometrie und allgemeiner für das Tiefensehen herausgearbeitet worden (siehe bspw. [Horn 1986], [Faugeras 1993], [Faugeras und Luong 2001], [Forsyth und Ponce 2003] und [Ma u.a. 2006]). Trotzdem existieren bekannte Probleme, die den heutigen Stand der Forschung im Bereich des Computersehens beschäftigen. Diese Probleme konzentrieren sich insbesondere auf die Merkmaldetektion und die Korrespondenzanalyse (siehe dazu bspw. [Tuytelaars und Mikolajczyk 2007]).

Man möchte u.A. die folgenden Punkte optimal umsetzen:

- Geometrieinvariante Merkmale.
- Subpixelgenaue Merkmale.
- Finden einer Korrespondenzabbildung zwischen Merkmalen unterschiedlicher Ansichten.
- Nachbearbeitung der Tiefendaten (Behandlung von Verdeckungen, Ausreißern und Löchern; Umgang mit Homogenitäten und Beleuchtungsstörungen).

Es gibt einige weitere technische Systeme, die das räumliche Sehen ermöglichen:

- Interferometrie,
- Gestalt aus Schattierung,
- Photogrammetrisches Stereo und
- Tomographie.

Der interessierte Leser findet eine gute Einführung und weitere vertiefende Literaturhinweise in [Jähne 2006]. Aufgrund ihres speziellen technischen Aufbaus, disqualifizieren sich diese Techniken zum gegenwärtigen Zeitpunkt für den Einsatz einer sicheren Arbeitsraumüberwachung.

2.1.3 Nicht optische Systeme für die Mensch–Roboter–Kooperation

In der Forschung werden auch nicht optische Systeme für die Mensch–Roboter–Kooperation eingesetzt. Dazu zählen:

- **Akustische Systeme (z.B. Ultraschall)**

Beim Ultraschall werden akustische Wellen mit 20 kHz (entspricht der oberen Hörschwelle des Menschen!) bis 1 GHz ausgesendet. Gemessen wird bei diesem Verfahren die Echolaufzeit. Damit lassen sich Informationen über die Tiefe gewinnen. Allerdings sind Ultraschallwellen abhängig von der Dichte, dem Druck und der Temperatur des Ausbreitungsmediums und damit im Umfeld der Robotik vergleichsweise weniger universell einsetzbar als optische Systeme.

Eine weitere Besonderheit liegt in der Abstrahlcharakteristik, die einer Keule gleicht. Die Abstrahlcharakteristik stellt auch ein Problem für die Anwendung in der Arbeitsraumüberwachung dar. Denn die Abstrahlcharakteristik ist stark exzentrisch geformt (Keulencharakteristik), wodurch die Objekte, welche sich in geringem Abstand zur Hauptachse der Keule befinden, nicht detektiert werden können.

- **Taktile Sensoren**

Taktile Sensoren werden in der Robotik im Allgemeinen, aber auch im Speziellen für die sichere Mensch–Roboter–Interaktion eingesetzt. Durch Auswertung von Druckprofilen können bspw. Formen unbekannter Objekte erkannt werden. Bei der Mensch–Roboter–Interaktion werden die Druckprofile im Kontaktfall ausgewertet und je nach Bedürfnis Ausweichbewegungen oder eine Reduktion der Arbeitsgeschwindigkeit eingeleitet.

- **Induktive Näherungssensoren**

Das Funktionsprinzip dieses Sensors basiert auf dem Einsatz einer Spule und eines Ferritkerns. Die von Wechselstrom durchflossene Spule erzeugt ein elektromagnetisches Wechselfeld. Ein elektrisch leitender Körper, der in das Feld eindringt, veranlasst das Magnetfeld, eine Spannung zu induzieren. Die Spannungsänderung beeinflusst eine Änderung der Energie des Feldes, welche gemessen werden kann und ausgenutzt wird, um als Detektor eingesetzt zu werden.

- **Kapazitive Näherungssensoren**

Kapazitive Näherungssensoren basieren auf einem Kondensator. Der Kondensator ist ein elektrisches Bauelement und besteht aus zwei Elektroden. Legt man eine Spannung an, laden sich die Elektroden elektrisch auf. Die Ladung auf den Elektroden erzeugen ein elektrisches Feld, welches sich zur angelegten Spannung proportional verhält.

Die Kapazität des Sensors beschreibt diese Proportionalität. Den Betrag der Ladung, welcher von den Elektroden aufgenommen werden kann, wird auch von den Materialeigenschaften des Objekts im Feld des Kondensators bestimmt.

Die Dielektrizitätskonstante ist eine skalare Größe, welche die Materialeigenschaften ausdrückt. Die Veränderung der Dielektrizitätszahl kann verwendet werden, um Objekte im Feld des Kondensators zu detektieren.

In Anwendungen mit industriellen Robotern werden besondere Sensoren verwendet, bei denen die Elektroden nicht in einer Komponente untergebracht werden. Stattdessen wird eine Elektrode auf dem Roboter angebracht, die Andere ist mit dem Grundpotential der Erde verbunden. Die Meßelektrode muss isoliert werden, da der Roboter geerdet ist und überwiegend aus Metall besteht. Zur Erhöhung der Meßempfindlichkeit werden relative Änderungen, d.h. gegen einen Referenzwert, ausgewertet. Zusätzlich werden flächendeckende Elektroden verwendet (siehe z.B. [Heiligensetzer 2003]) um einen Volumengewinn bzgl. des Meßfeldes zu erzielen.

Kontaktbasierte Sensoren

Kraft–Momenten–Sensoren messen Kräfte und Momente. Durch Auswertung der Sensorsignale ist es durch diese Sensoren bspw. möglich, den Roboter durch den Menschen gezielt zu führen. Diese Art von Sensoren werden ebenfalls für die Mensch–Roboter–Interaktion eingesetzt (siehe bspw. [Yigit 2005]). In vielen Anwendungen im Bereich der Robotik - sei es Industrie-, Service- oder Humanoide–Robotik - kann die Schätzung der Trägheitsparameter eines am Endeffektor montierterterter Kraft-Momenten-Sensors Vorteile bringen.

Als Trägheitsparameter werden die Masse, die Koordinaten des Massenschwerpunktes, die Trägheits- und die Deviationsmomente erkannt. Häufig werden dazu Kraft–Momenten–Sensoren eingesetzt, die an der Handwurzel des Roboters montiert werden. Zur Schätzung der

Trägheitsparameter ist es notwendig, die auf den Roboter einwirkenden Kräfte und Drehmomente zu messen. Zudem sind Beschleunigungen und Winkelgeschwindigkeiten von Interesse. Die Beschleunigungen können entweder von speziellen Sensoren gemessen oder aus den Gelenkwinkelmessungen des Roboters berechnet werden. Um in sicherheitskritischen Bereichen, wie bspw. der Mensch–Roboter–Kooperation eingesetzt zu werden, muss die Schätzung der Trägheitsparameter zur Laufzeit erfolgen. Zudem müssen problematische Eigenschaften von Kraft–Momenten–Sensoren (hoher Preis, umfangreiche Umbaumaßnahmen am Steuerschrank, etc.) berücksichtigt werden [Winkler 2006a].

2.1.4 Das Lichtflugzeit–Prinzip und eine technologische Realisierung

In den vorhergehenden Abschnitten wurde auf verschiedene Technologien hingewiesen, die zum Stand der Forschung zählen, insbesondere Sensoren für die Bereitstellung von Distanzwerten. Die wichtigsten Kategorien sind die passive Triangulation, die aktive Triangulation und die Interferometrie.

In diesem Abschnitt soll eine Technik beschrieben werden, welche für die hier vorliegende Arbeit die Grundlage bildet. Die im Folgenden dargestellten Resultate sind eine Zusammenfassung der wichtigen Grundlagen aus den Dissertationen von [Lange 2000], [Justen 2001], [Heinol 2001] und [Schneider 2003]. Durch diese Zusammenfassung soll es möglich sein, eine Abgrenzung gegenüber bisherigen Technologien für das Tiefensehen deutlich zu machen und um damit verständlich zu machen, weshalb ein System für die sichere Mensch–Roboter–Kooperation von der Photonen–Misch–Technologie profitiert.

Das Lichtflugzeit–Prinzip wird realisiert, indem ein moduliertes Lichtsignal durch eine aktive Beleuchtung abgestrahlt wird und durch eine Empfängeroptik wieder empfangen wird. Durch Kreuzkorrelation des gesendeten Signals mit dem empfangenen Signal und der genauen Kenntniss der Lichtgeschwindigkeit ist es möglich, die Lichtflugzeit des modulierten Signals zu bestimmen und damit schließlich die Entfernung zu Szenehindernissen zu bestimmen (siehe Abbildung 2.3). Die Realisierung hochpräziser Zeitmessungen, basierend auf dem Lichtflugzeit–Prinzip, sind deshalb besonders wichtig. Auf der Grundlage dieses Gedankenganges basiert das in dieser Arbeit verwendete Photonen–Misch–Gerät (PMD³), eingesetzt in die O3D-Kamera [O3D-PMD]. Die O3D-Kamera wird in dieser Arbeit eingesetzt. Sie verfügt derzeit über eine Auflösung von bis zu max. 204x204 Bildpunkten und einer Bildwiederholrate von bis zu max. 40 Bildern pro Sekunde (Amplituden– und Tiefenbilder).

Durch die rasante Miniaturisierung in der Mikrosystemtechnik gelingt es, die aktive Beleuchtung und den Detektor sehr nahe beieinander zu platzieren. Dadurch verhindert man bspw. Abschattungseffekte, wie sie bspw. bei der passiven Triangulation auftreten. Die kritischen Abschattungseffekte stellen ein großes Problem für die bisherigen Triangulationstechniken dar. Beim passiven Stereo kann man diesem Problem entgegenwirken, indem bspw. mehrere Kameras eingesetzt werden. Damit erhöht sich der Hardwareaufwand, insbesondere aber der Aufwand für Berechnung und Auswertung. Ein Vorteil für die kompakte Lösung der PMD–Technologie.

Die optischen Systeme, basierend auf dem hier skizzierten Lichtflugzeit–Prinzip, können sich ihrerseits durch folgende Punkte unterscheiden:

- Die verwendeten Modulationssignale,

³PMD steht für *photonic mixing device*. Diese Bezeichnung hat sich durchgesetzt und wird auch vom Hersteller im Namen des Produktes geführt.

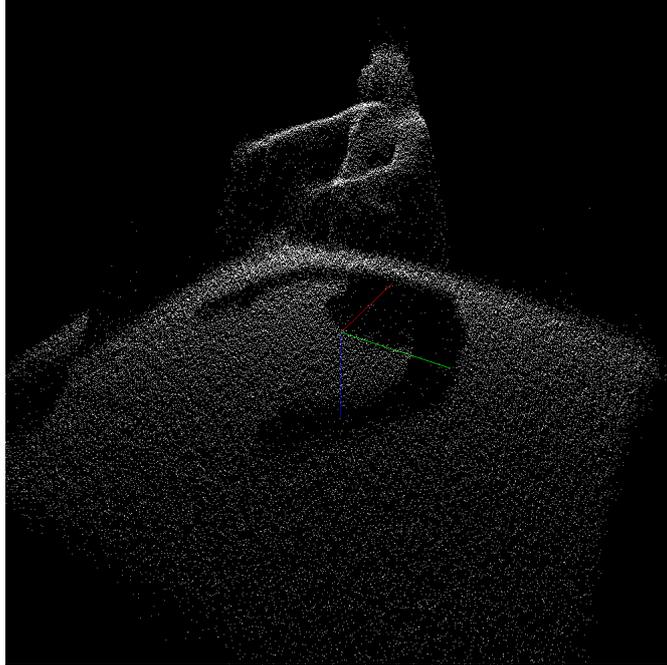


Abbildung 2.3: Darstellung eines Menschen im Raum, aufgezeichnet mit dem PMD Cam-Cube 2.0 (204 × 204-Bildpunkte Auflösung).

- die Art der modulierten Lichtquellen und
- die Empfängermodule, bzw. die verwendeten Detektoren.

Eine in der Praxis übliche Vorgehensweise basiert auf der Demodulation. Dabei wird ein emittiertes, modulierte Lichtsignal gegen ein synchron gestartetes Referenzsignal korreliert. Die dabei entstehende Phasenverschiebung ermöglicht die Distanzberechnung. Die präzise Zeitmessung wird deshalb durch die Phasenmessung ersetzt. Auch die Realisierung des Lichtimpulses kann auf unterschiedliche Weise geschehen. Üblicherweise wird heute eine kontinuierliche, modulierte Welle verwendet. Aber auch andere Signalformen, wie bspw. gepulste Modulation oder binär bzw. ternär kodierte Signale, sind denkbar.

Da für ein 3D optisches System für eine Vielzahl von Bildpunkten die Distanz ermittelt werden muss, werden die dazu benötigten 2D-(De-)Modulatoren durch Pocketszellen oder Kerrzellen realisiert.

Um neben der Phase auch die Modulationamplitude zu bestimmen, wird ein sogenanntes 4 Tap-Lock-In CCD [Spirig u.a. 1995, Spirig und Seitz 1996, Spirig u.a. 1997] verwendet. Es besteht auch die Möglichkeit, komplexere Signale zu demodulieren, da mehrere Abtastwerte gespeichert werden können.

Im weiteren Verlauf wird aber der Fokus auf die sinusmodulierten Signale gelegt. Eine Verbesserung des Lock-In CCD Sensors wird in [Lange u.a. 1999] vorgestellt.

Die Bestrahlungsstärke, welche durch den PMD-Sensor wahrgenommen wird, lässt sich durch folgende lineare Gleichung beschreiben:

$$E(t) = G'_0 + A' \cdot M(t) \quad , \text{ mit } M(t) = M(t + T) \quad . \quad (2.8)$$

Dabei ist $M(t)$ eine periodische Funktion, mit der der Periodendauer T , die durch den PMD-Sensor erzeugt und zur Ausleuchtung der Szene verwendet wird. Der Wert G_0 ist eine konstante Größe und integriert die Hintergrundbeleuchtung sowie den Gleichstromanteil (DC) der aktiven Beleuchtung. Durch A' wird die Amplitude, zugehörig zur normalisierten Modulation $M(t)$ beschrieben. Mit der Annahme, dass sich der photoempfindliche Halbleiter linear verhält, gilt für das induzierte elektrooptische Signal:

$$S(t) = G_0 + A \cdot M(t) \quad . \quad (2.9)$$

Dabei sind G_0 und A linear skalierte Versionen von G'_0 und A' . Diese Annahme gilt nur für den Idealfall, da in der Realität beide Größen durch verschiedene optische und elektrooptische Eigenschaften des Systems beeinflusst werden (siehe dazu [Lange 2000], [Justen 2001] und [Schneider 2003]). Die Remission des ausgestrahlten, modulierten Lichts, erfährt eine Phasenverzögerung aufgrund der Lichtgeschwindigkeit. Diese Phasendifferenz soll durch ϕ ausgedrückt werden:

$$M(t) = M'(t - \frac{\phi}{\omega}) \quad . \quad (2.10)$$

Idealerweise demoduliert der Sensor das Signal $S(t)$ durch Korrelation mit einer durch die Heavyside-Funktion H abgebildeten Funktion R . Die folgende Kreuzkorrelation beschreibt diesen Vorgang:

$$\begin{aligned} C(\theta) &= \int_0^{mT} S(t) \cdot R(t + \frac{\theta}{\omega}) \\ &\stackrel{\text{Def.}}{=} \int_0^{mT} (G_0 + A \sin(\omega t - \phi)) \cdot H(\sin(\omega t + \theta)) \\ &= \int_0^{mT} (G_0 + A \sin(\omega t - \phi - \theta)) \cdot H(\sin(\omega t)) \\ &= m \int_0^{\frac{T}{2} + \frac{\phi + \theta}{\omega}} (G_0 + A \sin(\omega t - \phi - \theta)) \\ &= m \int_{\frac{\phi + \theta}{\omega}}^{\frac{T}{2}} (G_0 + A \sin(\omega t - \phi - \theta)) \\ &= mT \left(\frac{A}{\pi} \cos(\phi + \theta) + \frac{G_0}{2} \right) \quad . \quad (2.11) \end{aligned}$$

θ ist die Kreuzkorrelationsvariable. Die Funktion $C(\theta)$ wird durch den PMD-Sensor umgesetzt. Man ist daran interessiert, ϕ , A und G_0 zu bestimmen um damit u.A. die Distanzmessung möglich zu machen. Außerdem gilt, dass $\phi = \frac{2\omega r}{c}$ ist.

Da man drei Unbekannte hat, aber nur über eine Gleichung verfügt, kann man durch Mehrfachabtastungen des Signals das dadurch entstehende lineare Gleichungssystem lösen. Die Lösung gelingt, indem zu unterschiedlichen Zeitpunkten (Phasen) die Funktion $C(\theta)$ abgetastet wird. Die Abtastzeitpunkte werden so gewählt, dass $\theta \in [0, 2\pi]$ gilt.

Die so gewonnenen Gleichungen sind durch das Messrauschen gestört. In der Praxis werden deshalb mehrere Abtastpunkte verwendet, um schließlich ein überbestimmtes Gleichungssystem zu erhalten. Dieses Gleichungssystem lässt sich dann z.B. im Sinne der kleinsten Fehlerquadrate lösen. Damit erhält man eine robustere Schätzung der Parameter. Die Abbildung 2.4 visualisiert den Abtastvorgang.

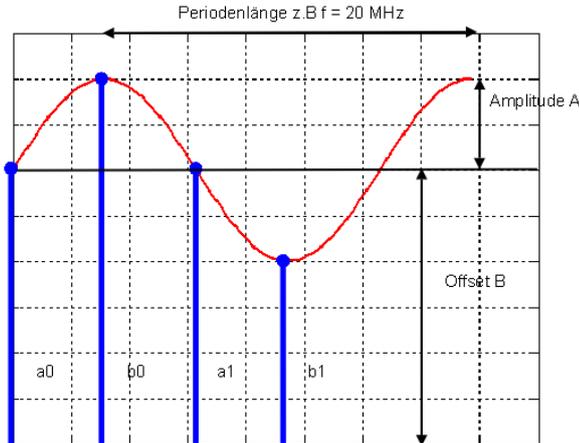


Abbildung 2.4: Vierphasige Abtastung eines sinusmodulierten Signals.

Die PMD-Technologie ermöglicht es in einem Zeitpunkt, zwei Abtastungen gleichzeitig durchzuführen. Die Resultate sind bezüglich der Phase um 180° gedreht. Es wird deshalb lediglich ein zweiter Abtastzeitpunkt gewählt, wobei die Phase hier 90° ist und damit die zweite Messung in einer Phasenlage von 270° aufgenommen wird.

Die Phasenverzögerung lässt sich dann bestimmen:

$$\phi = \text{atan} \left(\frac{C(\tau_3) - C(\tau_1)}{C(\tau_0) - C(\tau_2)} \right) . \quad (2.12)$$

Die Abtastzeitpunkte werden dabei durch τ_i ausgedrückt, mit $i \in [0, 3]$, und korrespondieren mit den Phasenlagen aus Abbildung 2.4.

Für die Amplitude gilt:

$$A = \frac{\delta}{\Delta t \cdot \sin \delta} \cdot \frac{\sqrt{(A_{\tau_3} - A_{\tau_1})^2 + (A_{\tau_0} - A_{\tau_2})^2}}{2} . \quad (2.13)$$

Der Aufpunkt wird bestimmt durch die einfach gewichtete Mittelung:

$$B = \frac{1}{4 \cdot \Delta t} \sum_i A_i . \quad (2.14)$$

2.2 Ausgewählte Forschungsgruppen

Nachdem in den vorhergehenden Abschnitt einige Sensoren vorgestellt wurden, werden in diesem Abschnitt ausgewählte Forschungsgruppen und ihre Realisierungen für eine Mensch-Roboter-Interaktion bzw. Kooperation diskutiert.

Die bisher vorgestellten Techniken um eine sichere Mensch–Roboter–Kooperation zu gewährleisten, sind wenig bis gar nicht in die Entwicklung durchgedrungen. Dies liegt mit Sicherheit auch an der herausfordernden Aufgabenstellung, neben der Überwachung der Szene und des Szeneverstehens, gleichzeitig Lösungsstrategien anzubieten, welche das Risiko für den Menschen während der Interaktion minimieren müssen, d.h. die Entwicklung eines kognitiven Industrieroboters. Des Weiteren behandeln die meisten Ansätze aus dem vorhergehenden Abschnitt nicht die auftretenden Verdeckungsprobleme, die jederzeit entstehen können, wenn sich Menschen in einer Szene dynamisch verhalten und gefährden in diesem Punkt die Sicherheit für den Menschen.

Abgesehen von diesem Aspekt wird in der industriellen Robotik zusätzlich verlangt, die Szene mit zu modellieren. Dabei belasten statische Objekte die Berechnungen weniger, als die dynamischen unbekannt Objekte, die es ebenfalls zu berücksichtigen gilt.

Die wesentlichen Punkte für eine sichere Mensch–Roboter–Interaktion sind deshalb:

- Eingesetzte Sensorik
- Verwendete Roboter
- Algorithmik
- Umgebungsmodellierung
- Risikoquantifizierung
- Echtzeitfähigkeit
- Möglichkeiten der Szeneinterpretation
- Risikominimierende Maßnahmen

Die im Folgenden vorgestellten Forschungsgruppen sind chronologisch sortiert. Außerdem helfen die folgenden Gruppierungen, die Darstellung zu strukturieren:

- Industrielle Forschung und Entwicklung
- Forschungseinrichtung national
- Universitäre Forschung (national und international)

2.2.1 Industrielle Forschung und Entwicklung

Fa. Reis Robotics, Obernburg, Deutschland

Die Fa, Reis Robotics ist ein Roboterhersteller und Systemintegrator für Robotersysteme. Der Hersteller favorisiert derzeit die Verwendung eines Laser–Scanners, um eine Vordergrunddetektion im Arbeitsbereich des Roboters zu realisieren (siehe dazu [Som 2003], [Som 2005] und [Som 2007]). So kann beispielsweise detektiert werden, ob eine Person sich innerhalb eines Gefahrenbereichs befindet.

Genau genommen kann hier allerdings nicht von einer Personenerkennung gesprochen werden, da lediglich eine Fläche in paralleler Ausrichtung zum Boden vom Sensor erfasst wird. Anschließend wird ein Zylinder um den vom Sensor detektierten Vordergrund modelliert. Der

kleinste Abstand zwischen dem Zylinder und dem Roboter dient als Eingabe in eine Funktion, die ihrerseits die maximale Arbeitsgeschwindigkeit des Roboters bestimmt.

Durch die Reduktion der Arbeitsgeschwindigkeit soll Sicherheit für den Menschen erreicht werden.

Der Vorteil, einen Laser-Scanner zu verwenden, liegt darin begründet, dass diese Technologie bereits sicherheitszertifiziert ist. Deshalb befinden sich erste Prototypen im industriellen Einsatz.

Es ist jedoch fragwürdig, inwieweit eine solche Realisierung dem weiteren Fortschritt dient, denn wichtige Merkmale für die Interpretierbarkeit der Szene, und damit auch zur Einschätzung des Risikos, sind nur beschränkt möglich. Die Möglichkeit eine wirkliche Kooperation zwischen Mensch und Roboter schließt sich für eine Vielzahl von Szenarien aus, da wichtige Extremitäten wie Arme und Hände überhaupt nicht erkannt werden. Außerdem ist die Approximation des menschlichen Körpers durch einen Zylinder als unzureichend einzustufen. Ein positiver Aspekt dieses Systems liegt allerdings darin, dass es dem Hersteller gelungen ist, sein System in die Robotersteuerung zu integrieren.

Zu berücksichtigen bleibt, dass in den zitierten Vorträgen und nach Rücksprache mit den Entwicklern keine experimentellen Ergebnisse vorliegen, die verschiedenartige Szenarien experimentell untersuchen und diskutieren.

Fa. Pilz, Sindelfingen, Deutschland

Die Fa. Pilz [Pilz 2007] stellte 2007 ein Kamerasystem zur sicheren Arbeitsraumüberwachung vor. Der Arbeitsraum wird dabei von der Decke durch drei eng beieinander liegenden Kameras ausgeleuchtet. Die Bildauswertung geschieht weitestgehend auf der Basis von Stereoalgorithmen.

Statt eine physikalische Trennvorrichtung zu verwenden, können bis zu maximal 50 statische, virtuelle Schutzzäune errichtet werden. Das Kamerasystem ist nicht in der Lage, die Vorgänge in der Szene auszuwerten. Es findet lediglich eine Vordergrunddetektion statt.

Die Erkennung geschieht innerhalb von 350 ms und erreicht damit bei weitem keine Echtzeitanforderung im Sinne der Bildfolgenauswertung. Trotzdem darf diese Technologie in die Produktion integriert werden.

Fa. Castell, Erlanger, USA

Die Fa. Castell bietet seit 2007 ein kompaktes System, bestehend aus vier Kameras, für die Arbeitsraumüberwachung an. Da aber bis zum heutigen Zeitpunkt keine genaueren technischen Daten dem Produkt zu entnehmen sind, soll an dieser Stelle lediglich auf das System aufmerksam gemacht werden. In der Praxis wird das Kamerasystem an der Decke befestigt, um dann virtuell Bereiche zu definieren, welche die Sicherheit bei der Mensch-Maschine-Interaktion zu gewährleisten.

2.2.2 Forschungseinrichtungen und universitäre Forschung

Fraunhofer Institut für Produktionstechnik und Automatisierung (IPA), Stuttgart, Deutschland

Die Arbeitsgruppe Robotersysteme des Fraunhofer IPA beschäftigt sich intensiv mit der Realisierung einer sicheren Mensch-Roboter-Kooperation. In der Arbeit von [Winkler 2008] wird

ein System vorgestellt, das ein Kamerasystem, basierend auf dem Lichtflugzeit-Prinzip einsetzt, um den Arbeitsraum zu überwachen. Kollisionen werden in diesem System im kartesischen Raum betrachtet, da die Szenedaten fusioniert werden können. Dies ist vorteilhaft, da der zugrundegelegte Sicherheitsmechanismus darin besteht dynamische, virtuelle Schutz-zonen zu errichten. Diese Zonen lassen sich im kartesischen Raum visualisieren, wodurch die Erfassbarkeit verbessert wird.

Es wird ein Konzept vorgeschlagen, welches die Überwachung in drei Bereichstypen einteilt:

1. Bereiche, in denen zwingend Messwerte geliefert werden müssen, um z.B. eine Abschätzung durch den Roboter zu detektieren.
2. Sicherheitskritische Bereiche, in denen sich weder Objekte noch Menschen aufhalten dürfen.
3. Bereiche, die von der Kollisionsbetrachtung ausgeschlossen sind.

Um Gefahrensituationen zu verhindern, wird auch bei diesem System die Arbeitsgeschwindigkeit geregelt. In einem Umweltmodell werden die Roboter abgebildet. Die Roboter werden systemintern durch quaderförmige Hüllkörpermodelle repräsentiert. Die Vorteile des hier vorgestellten Systems liegen in der Abstandsberechnung. Die dynamischen Sicherheitsbereiche sind vorteilhaft für den Anwender, insbesondere wegen der Berücksichtigung des Umweltmodells. Es lässt sich intuitiv planen und kritische Situationen können vorab am Rechner diskutiert werden.

Zu kritisieren bleibt lediglich, dass die kinematische Struktur eines sich in der Zelle bewegenden Menschen nicht umgesetzt wird. Für komplexe Kooperationsszenarien ist es von großem Vorteil, wenn zumindest die Erkennung der Arme realisiert ist.

Ein besonders schwerwiegender Nachteil ist die Laufzeit der derzeitigen Realisierung. Da mehr als 15 Bilder pro Sekunde für die Auswertung benötigt werden, ist die Echtzeitfähigkeit nicht gewährleistet.

Universität Stuttgart, Deutschland

Die Dissertation von [Thiemermann 2005] befasst sich mit der direkten Mensch-Roboter-Kooperation in der Kleinteilmontage unter Verwendung eines SCARA-Roboters. Insbesondere wird in dieser Arbeit eine ganzheitliche Betrachtung der Problemstellung eingenommen, um damit ein effizientes System zu entwickeln, welches optimal ergonomisch ist.

Gemeint ist damit die Optimierung hinsichtlich Beeinträchtigungen, Ermüdungserscheinungen, Sicherheitsaspekten und anderen, sich nachteilig auswirkende Faktoren für den Menschen. Basierend auf der Auswertung von Farbbildfolgen werden Merkmale wie bspw. der kürzester Abstand zur Werkzeugspitze des Roboters, die Relativgeschwindigkeiten und die Beschleunigungen des Roboters gemessen.

Anschließend wird mit Hilfe eines neuronalen Netzes eine (einsträngige) unscharfe Logik eingelernt. Die eingelernte, unscharfe Logik wird in einer Pilotanlage eingesetzt und passt die Arbeitsgeschwindigkeit des Roboters an.

Die Schätzung des Risikos wird durch ein eingelerntes Fuzzy-System realisiert. Es handelt sich hierbei um ein einfaches System, welches in komplexen Szenarien schnell zu einer unüberschaubaren Anzahl von Regeln kommen kann, was an der einsträngigen Struktur des Systems liegt. Trotzdem geht dieses System in diesem Punkt weiter, als die bisherigen vorgestellten Ansätze.

Als nachteilig muss die auf Farbmessungen basierte Erkennung der Hände gewertet werden. In der Ausarbeitung von [Thiemermann 2005] gibt es weiterhin keine Hinweise über Probleme mit der Farbauswertung und der Schwierigkeiten bei der Auswertung unter sich verändernden Lichtverhältnissen.

Technische Universität Berlin, Deutschland

An der Technischen Universität wurde in Zusammenarbeit mit dem Fraunhofer Institut für Produktionstechnik (IPK) eine Veröffentlichung für die flexible Mensch–Roboter–Kooperation publiziert [Krüger u.a. 2005]. Aus einem trinokularen Kamerasystem basierend auf handelsüblichen Farbbildkameras (CCD, RGB), werden die Kopfposition⁴ und die Hände erkannt. Generell werden drei Merkmale aus den Farbbildern extrahiert:

- Farbwerte
- Textur
- Bewegungsschätzungen

Die Erkennung der Hände basiert auf einem physikalischen Modell, einem sog. dichromatischen Reflektionsmodell. Dabei werden zwei Hauptmerkmale aus den Farbbildern der Kameras extrahiert (Melanin und Haemoglobin). Die rot–grün–Verteilung wird für eine Klassifikation bei der Auswertung der Bildpunkte verwendet. Aus dem Abstand der Hände wird die Arbeitsgeschwindigkeit des in Kooperation stehenden Roboters angepasst. Die Kollisionsvermeidungsstrategie wird durch den Stillstand der Maschine realisiert.

Bei dem hier vorgestellten System ist der Einsatz einer dritten Kamera vorteilhaft für die Generierung von Tiefenkarten. Ein derartiges trifokales System ist robuster als ein bifokales System [Schreer 2005]. Von Vorteil ist auch die Möglichkeit, echtzeitfähige Berechnungen durchzuführen.

Allerdings basiert die Erkennungsleistung in Abhängigkeit von Farbwerten. Es fehlt die genaue Darstellung, wie die Bewegungen erkannt werden. Da die Merkmale nicht weiterverarbeitet werden, wird hier auch keine Interpretation der Szene realisiert.

Universität British Columbia, Kanada

Einen Überblick für die Forschungsaktivitäten einer sichere Mensch–Roboter–Interaktion, wie sie an der Universität British Columbia vollzogen wird, liefert die Dissertation von [Kulić 2005]. Erarbeitet werden in dieser Dissertation Sicherheitsstrategien für Menschen, die mit einem Puma Roboter vom Typ 560 (Kleinbauroboter) in Interaktion treten.

Das Interessante, im Hinblick auf eine sichere Mensch–Roboter–Kooperation im industriellen Umfeld, liegt in der Formulierung einer Methodik zur Bestimmung des Gefahrenmaßes⁵. Dieses Maß wird ermittelt, indem die folgenden Merkmale:

- Roboterkonfiguration
- relative Geschwindigkeit

⁴Die Autoren sprechen allgemein von Personenverfolgung! Die Posenerkennung einer Person scheint sich aber lediglich auf die Erkennung von Kopfpositionen und der Hände zu beschränken.

⁵engl. measure of danger

- kürzester Abstand

verknüpft werden.

Falls erforderlich, wird eine Umplanung der ursprünglich vorgesehenen Bahnplanung ausgeführt. Die Autorin stellt fest, dass dafür der Einsatz diverser Sensoren eine Notwendigkeit darstellt und löst das Wahrnehmungsproblem durch Einsatz eines Stereofarbbild-Kamerasystems und weiterer Sensoren für die Aufzeichnung von physiologischen Daten.

Durch Auswertung der Farbbilder wird die in Interaktion tretende Person lokalisiert, das Gesicht erfasst und die Orientierung des Kopfes geschätzt.

Die physiologische Signalverarbeitung erfasst den Herzschlag, die Absonderung von Schweiß und unterschiedliche Beanspruchung der Muskulatur. Dazu werden verschiedene Sensoren eingesetzt, wie bspw.:

- ein Elektrokardiograph
- ein Sensor zur Messung der Leitfähigkeit der Hautoberfläche (SCFlex-Pro)
- ein Elektromyograph

Die Mustergewinnung, Merkmalerzeugung und Merkmalsauswertung ist bzgl. der Berechnungszeiten unkritisch. Allerdings muss beachtet werden, dass im industriellen Umfeld durchaus körperlich anstrengende Arbeiten verrichtet werden, so dass Merkmale, wie bspw. die elektrische Leitfähigkeit der Haut, als weniger geeignet angesehen werden müssen. Außerdem ist das Anbringen der Sensorik mit nicht zu unterschätzendem Aufwand verbunden und kann die körperliche Bewegungsfreiheit des Werkers stark einschränken. Ähnliche Kriterien gelten auch für die Schätzung mentaler Zustände, durch z.B. den Durchmesser der Pupillen, die Herzrate, usw..

Nachdem die Gefahr mit Hilfe der Merkmalsauswertung quantifiziert wird, besteht die Möglichkeit, die ursprünglich geplante Bahn des Roboters umzuplanen, um damit eine potentielle Gefahr zu minimieren. Um eine tatsächliche Interaktion zuzulassen wird vorgeschlagen, nicht den kürzesten Abstand zum Roboter zu betrachten, sondern stattdessen den kürzesten Abstand zum Masseschwerpunkt des Roboters. Der Vorteil bei dieser Vorgehensweise ist, dass im Falle einer Kooperation, bei der ein direkter Kontakt mit dem Menschen erwünscht ist, die Planung nicht zum Stillstand kommen muss. Allerdings setzten die Autoren voraus, dass lediglich die Hände mit dem Roboter in Kontakt treten. Die Berücksichtigung anderer Körperteile findet nicht statt. Diese Tatsache stellt eine potentielle Gefahr für den Menschen dar.

Auch unter Berücksichtigung der Gesichtserkennung und Abstandsberechnung muss davon ausgegangen werden, dass durch Verdeckungen oder anderen Störungen (z.B. sich verändernde Lichtverhältnisse), eine zuverlässige Szeneinterpretation nicht gewährleistet werden kann. Die Autoren berichten nicht über diese bekannten Probleme!

Das Planungsverfahren für die Minimierung des Risikos basiert auf der Minimierung eines Potentialfelds (siehe bspw. [Latombe 1991], insbesondere Kap. 7).

Grundlage dieser Verfahren ist die Formulierung eines Funktionals. Dieses Energiefunktional soll durch das Planungsverfahren minimiert werden. Dabei werden die Güte der zuvor aufgezählten Merkmale in dem Formalismus berücksichtigt.

Insbesondere die Integration einer Bahnplanung und damit die Möglichkeit, zukünftig flexibel auf sich ergebende Anforderungen in der Roboterzelle reagieren zu können, muss als Vorteil gewertet werden.

Wenig vorteilhaft ist die Repräsentation einzelner Körperteile durch einzelne Kugeln. Diese Vorgehensweise ist zu ungenau. Damit ergeben sich auch starke Einschränkungen bzgl. des Anwendungsfeldes. Der Einsatz komplizierter Apparaturen, z.B. EEG, ist fragwürdig im Rahmen der Industrierobotik, da er zu aufwendig ist.

Offen gebliebene Fragen beziehen sich auf die Echtzeitfähigkeit des Systems.

Universität Bayreuth, Deutschland

Die Forschungsgruppe um Prof. Henrich hat, einen wesentlichen Beitrag für den Forschungsbereich sichere Mensch–Roboter–Kooperation und Interaktion im industriellen Umfeld geleistet. Um einen umfassenden Überblick zu geben, werden in diesem Abschnitt die wichtigsten Arbeiten der Forschungsgruppe chronologisch vorgestellt. Die erste ausgewählte Arbeit [Ebert und Henrich 2001] stellt eine Analyse und einen Entwurf für ein Systemkonzept für die sichere Überwachung eines gemeinsamen Arbeitsbereichs von Mensch und Roboter vor. Der technische Aufbau des Prototypen besteht aus einem sechs-achsigen Industrieroboter und vier Grauwertkameras. Herzstück des Systems ist die Verwendung von zwei a priori berechneten Tabellen (T1 und T2), welche implizit die Kamerakalibrierung und die Geometrie der Roboter kodieren. Die Bilder der Kameras werden fusioniert, nachdem so genannte Merkmalsbilder erzeugt wurden. In einem vorverarbeitendem Schritt werden alle Konfigurationen des Roboters nacheinander angefahren. Für jede Konfiguration jeder Kamera wird dann eine Zeile in der Tabelle T1 reserviert und alle Bildpunkte, welche belegt sind und zum Roboter gehören, gespeichert.

Diese Vorgehensweise hat den Vorteil, äußerst effizient Belegungen zu prüfen, allerdings ist dazu eine zeitaufwendige Vorverarbeitung nötig, die eine festgelegte Trajektorie des Roboters voraussetzt. Des Weiteren wird eine große Menge an Speicherplatz verbraucht, was die Integration in die Robotersteuerung derzeit nicht möglich macht.

Aus der Tabelle T1 wird anschließend die Tabelle T2 erzeugt. In der Tabelle T2 werden alle auffälligen Belegungen gespeichert. Diese Belegungen gehen aus den Differenzbildern jeder Kamera hervor und werden generiert, falls eine Blockierung vorliegt. Die Autoren stellen selbst fest, dass für ihre vorgeschlagene Vorgehensweise keine Robustheitsuntersuchungen bzw. experimentellen Ergebnisse vorliegen und deshalb die Grenzen dieses Konzepts nicht bekannt sind.

Ein weiteres Problem sind Fantomhindernisse, also Hindernisse, die detektiert werden, obwohl sie nicht vorhanden sind, was eine Hindernisvergrößerung zur Folge hat.

In [Ebert und Henrich 2002a] wird ein Vorgehen für eine rein bildbasierte Kollisionserkennung vorgestellt. Es werden wieder vier Kameras verwendet, um das System zu realisieren. Die Auswertung der Bilder basiert auf einem einfachen Differenzenverfahren und Binärbildern. Bildpunkte werden dabei aktiv, sobald sich ein Hindernis oder der Roboter selbst im Bild befindet.

In einem zweiten Schritt wird auf der Grundlage von sog. Roboter–Testbildern der Roboter aus den Bildern herausgerechnet, um schließlich Kollisionsbilder zu interpretieren und um festzustellen, ob eine Kollision vorliegt oder nicht. Wichtig bei dieser Vorgehensweise sind die Zustandsbilder, bei denen jeder Bildpunkt genau einem von vier möglichen Zuständen zugeordnet wird.

Die sich ergebenden Zustände sind:

- Hintergrund

- Hindernis
- Unbekannt
- Fehler

In einem ersten Schritt werden die sogenannten Hindernisbilder erzeugt. Ähnlich verhält es sich bei der Erzeugung der Kreuzungsbilder. Hier können die Bildpunkte ebenfalls genau einen von vier Zuständen annehmen (Hintergrund, Unbekannt, Fehler oder Hindernis).

Die Autoren berichten von Diskretisierungsfehlern, welche die Güte der Schätzungen beeinflussen, aber auch von einer, zu diesem Zeitpunkt fehlenden, optimalen Anordnung der Kameras, wodurch einige Konfigurationen nicht ausgewertet werden können.

Die Publikationen [Ebert und Henrich 2002b] und [Ebert und Henrich 2003a] sind Übersichtsartikel zum SIMERO-System⁶. Im Wesentlichen werden die bisher genannten Publikationen um die Fähigkeit einer statischen Bahnplanung ergänzt, d.h. dass zu einem bestimmten Zeitpunkt zwei Raumpunkte definiert werden und dann eine Bahn in Abhängigkeit der Roboterkonfiguration geplant wird.

Ist diese Bahn bestimmt, kann diese bis zum Erreichen der Zielkoordinate nicht geändert werden. Sie ist deshalb statisch.

In [Ebert und Henrich 2003a] wird explizit darauf hingewiesen, dass eine probabilistische Bahnplanung [Kavraki und Latombe 1998] verwendet wird, die auf Punkt-zu-Punkt Bewegungen basiert. Der Vorteil dieser Methode besteht in einer flexiblen, effizienten Realisierung von geplanten Bahnen, allerdings nur für statische Szenen. Des Weiteren werden i.A. keine glatten Bahnverläufe generiert, da zwischen zwei Punkten lediglich ein Geradensegment abgefahren wird. Deshalb muss bei Verwendung dieser Methode, z.B. durch Überschleifen [Siegert 1999], nachträglich geglättet werden. Dabei ist zu berücksichtigen, dass durch das Überschleifen zusätzliche Punkte im Konfigurationsraum erzeugt werden, die streng genommen wieder einer Risikoauswertung unterzogen werden müssen. Der Einsatz von Überschleifen zur Glättung der Robotertrajektorie kann auch deshalb rechenintensiv werden.

Eine detaillierte Beschreibung der Bahnplanung findet man außerdem in der Dissertation von [Ebert 2003b].

Im Wesentlichen werden auch hier nochmals die vier wichtigen Komponenten ausführlich diskutiert:

- Bildverarbeitung
- Robotermodellierung
- Kollisionserkennung
- Bahnplanung

Des Weiteren wird in dieser Arbeit die Notwendigkeit dargestellt auf eine dynamische Umwelt reagieren zu müssen. Im Rahmen der Dissertation werden dazu mehrere Standardverfahren (Wellenausbreitungsverfahren, Tiefensuche, HD-I und Probabilistischer Bahnplaner) untersucht. Außerdem werden sechs Regeln von dem Autor vorgeschlagen die zwingend vom Bahnplaner berücksichtigt werden sollen. Diese Regeln werden im Rahmen der hier vorliegenden Dissertation neu diskutiert (siehe bspw. 7.2).

⁶SIMERO: Sicherheitsstrategien für die Mensch-Roboter-Kooperation.

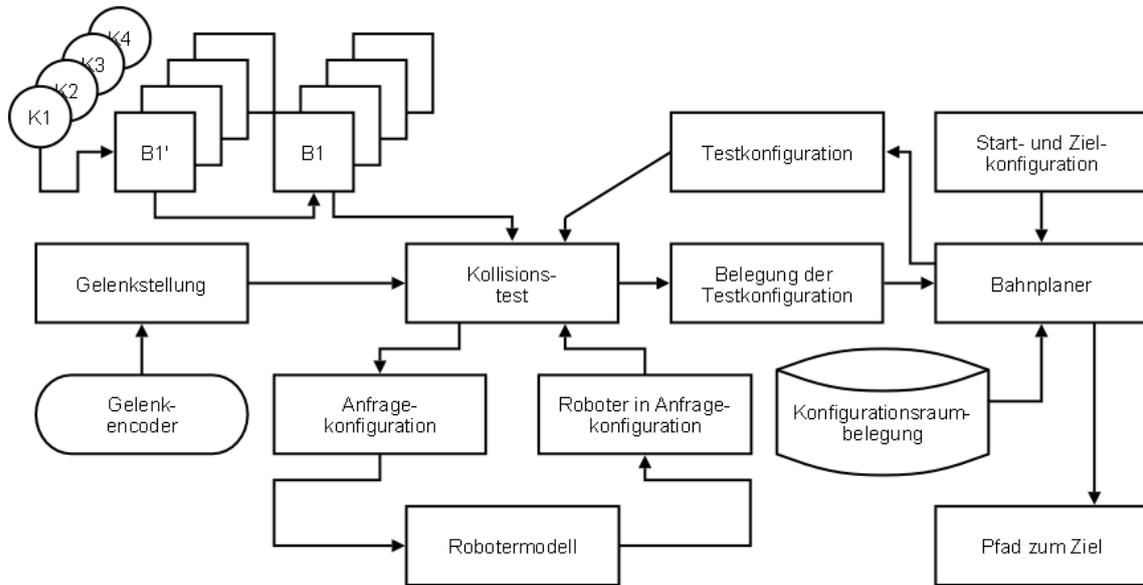


Abbildung 2.5: Modifizierte Systemarchitektur nach [Ebert 2003b].

Eine Modifikation der bisher vorgestellten Publikationen wird in [Gecks und Henrich 2004] erläutert. Von den Autoren wird der bisherige Ansatz von [Ebert 2003b] reproduziert und um ein weiteres Szenario erweitert. Die durchgeführten Experimente zeigen, dass das System mit 7 Bildern pro Sekunde die Echtzeitfähigkeit der Kameras stark unterschreitet. Eine tatsächliche, dynamische, reaktive Bahnplanung wird auch in dieser Arbeit nicht vorgestellt. Für das Experiment wird ein Überwachungsbereich von $1.5 \times 2 \times 2 \text{ m}^3$ ausgeleuchtet. Dazu werden 8192 Konfigurationen im Vorfeld aufgezeichnet. Die Auflösung der zu prozessierenden Differenzbilder beträgt 64^2 -Bildpunkte.

In [Henrich u.a. 2004] wird eine Erweiterung des bildbasierten Kollisionstest für ein probabilistische Bahnplanungsverfahren vorgestellt und verglichen. Für die Auswertung des Szenebereichs wird hier auf das Verfahren von [Ebert und Henrich 2002a] zurückgegriffen. Hinzu kommt ein Kollisionstest. Ausgehend von den linearen Segmentstücken (Punkt-zu-Punkt) wird der vom Roboter überstrichene Weg durch umrahmende Quader approximiert. Für das approximierte Volumen wird ein Kollisionstest durchgeführt.

Das resultierende Verfahren wird anschließend experimentell untersucht, indem zwei Arbeitsraumkonfigurationen mit bzw. ohne Hindernisse, basierend auf dem Einzelschritt-Kollisionstest und dem neuartigen Volumen-Kollisionstest unterschieden wird. Es gibt keinen Hinweis, dass das Planungsverfahren in Echtzeit läuft.

Auch in [Gecks und Henrich 2005] gibt es keinen Hinweis über eine echtzeitfähige Planung einer Bahn. Der bisherige Ansatz [Henrich u.a. 2004] wird dahingehend erweitert, dass Montageaufgaben eines Roboters in die Kollisionserkennung integriert werden. Die Autoren unterscheiden dazu vier Fälle:

- freie Bewegungen,
- überwachte Bewegungen,
- kleine Bahnen,

- große Bahnen.

In der Arbeit von [Kuhn u.a. 2006] wird ein System vorgestellt, welches die maximale Arbeitsgeschwindigkeit eines Industrieroboters regelt. Die Regelung geschieht durch Auswertung von Bildfolgen, basierend auf dem in [Ebert 2003b] vorgestellten Ansatz. Es werden fusionierte Daten, basierend auf Bildern und Kräften, hervorgehoben.

Des Weiteren wird der in [Gecks und Henrich 2005] diskutierte Ansatz erweitert, indem je nach Phase der Kooperation, das Hüllkörpermodell des Roboters angepasst wird. Damit wird prinzipiell zum ersten Mal auf die Möglichkeiten einer realen Kooperation Bezug genommen. Allerdings wird nicht zwischen Menschen und Objekten unterschieden, wodurch diese Vorgehensweise diskutierbare Konsequenzen hat, falls im Kooperationsmodus nicht der Mensch, sondern ein Hindernis Kontakt mit dem Roboter hat.

Diese Vorgehensweise wird auch in [Henrich und Kuhn 2006] vorgeschlagen. Eine Erweiterung dieser Arbeit bezieht sich auf die Integration eines Zustandsgraphen, welcher Übergänge zwischen den in [Kuhn u.a. 2006] vorgestellten vier Fällen realisiert.

Das in [Gecks und Henrich 2007] vorgestellte Verfahren verbessert die bisher vorgestellten Bahnplanungsalgorithmen und diskutiert erstmalig anhand experimenteller Ergebnisse die für die Planung benötigte Zeit. Es stellt sich heraus, dass das bisherige Planungsverfahren mit ca. 10 Bildern pro Sekunde arbeitet. Der Flaschenhals ist die Sensorauswertung, welche von den Autoren mit 70 ms beziffert wird.

Eine schnelle Variante für die Bestimmung von kürzesten Abständen zwischen bekannten und unbekanntem Objekten wird in [Kuhn u.a. 2007] vorgestellt. Die Autoren berichten von ca. 30 Bildern pro Sekunde für die Abstandsberechnung. Ausgehend von Einzelbildern der Szene werden Bereiche des Vordergrundes und des Hintergrundes klassifiziert [Ebert 2003b]. Anschließend werden, eine kalibrierte Kamera vorausgesetzt, Bildpunkte in den Szenebereich abgebildet. Die in dieser Arbeit vorgestellte Vorgehensweise ist flexibel, da sie rein bildgetrieben arbeitet. Systematische Fehler, verursacht durch den Aufnahmeprozess, können nicht ausgeschlossen werden. Ein weiterer Nachteil bzgl. der Genauigkeit liegt darin, dass die Verwendung eines Mehrkamerasystems die Genauigkeit erhöhen kann. Derzeit wird deshalb eine konservative Schätzung auf der Basis von sich ausbreitenden Kreisen realisiert. Die Schätzung findet somit rein lateral statt.

In den bisherigen Publikationen werden von der Forschergruppe um Herrn Prof. Henrich lediglich handelsübliche CCD-Kameras verwendet. In [Henrich und Gecks 2008a] wird erstmals eine Vorgehensweise zur Realisierung einer sicheren Mensch-Roboter-Koexistenz vorgestellt, welche Tiefenbild-Kameras, basierend auf dem Lichtflugzeit-Prinzip, verwenden.

Es wird eine Architektur favorisiert, welche beide Technologien in die Auswertung integriert und explizit darauf hingewiesen, dass die vorgeschlagene Vorgehensweise flexibel auf andere Anwendungsfelder der Arbeitsraumüberwachung ausgedehnt werden kann. Die Verwendung von Tiefenbildkameras ermöglicht die Modifikation der bisherigen Auswertearchitektur, insbesondere die Hintergrundmodellierung. Um eine konservative Schätzung zu erhalten, werden die Informationen der einzelnen Kameras (Polyeder) des heterogenen Kameranetzes geschnitten. Dadurch erhält man eine grobe Abschätzung aller in der Szene befindlichen Objekte.

Ein kritischer Punkt dieser Architektur ist die dort dokumentierte Bahnplanung zur Laufzeit, welche wie die zuvor beschriebene Planung [Gecks und Henrich 2007] auf einzelne Roboterkonfigurationen angewiesen ist und mit 5 ms je Konfiguration die Echtzeitfähigkeit nur knapp unterschreitet, da die Menge der zu prüfenden Konfigurationen je nach Abtastung hunderte Konfigurationen umfassen kann. Die Autoren verwenden deshalb eine auf Abtastung des

Suchraums basierte Vorausberechnung. Damit wird ein statisches Netz erzeugt, auf welchem zur Laufzeit geplant werden kann. Die propagierte Vorgehensweise kommt einer lokalen Suche gleich. Wie dieses Verfahren sich im Falle von Verklemmungen oder Zeitüberschreitungen verhält, ist aus der Arbeit nicht ersichtlich.

Eine Zusammenführung bereits publizierter Ergebnisse zu einem modifizierten Gesamtsystem bildet die Grundlage für die Arbeit von [Henrich u.a. 2008b]. Lediglich die Integration von Farbe als Merkmal für die Segmentierung wird in dieser Arbeit berücksichtigt. Des Weiteren wird ein silhouettenbasierter Rekonstruktionsansatz um Epipolarlinien erweitert. Diese Vorgehensweise unterstützt die Erzeugung des Hindernisbildes, welches zur Unterscheidung einzelner Objekte des Vordergrundes dient. Für diese modifizierte Vorgehensweise wurden keine Experimente publiziert. Es ist deshalb nicht abzuschätzen, wie robust diese Vorgehensweise sich unter industriellen Bedingungen verhält, vor allem das Verhalten gegenüber bekannten Störungen, wie bspw. sich verändernde Lichtverhältnisse.

Es wird von den Autoren kritisiert, dass durch die Verwendung von neuartigen 2,5D-Kameras lediglich einzelne Körperteile erkannt werden können. Die hier vorliegende Ausarbeitung wird zeigen, dass ein universeller Einsatz dieser Kameratypen bessere Lösungen bieten kann, als die bisherigen Systeme zur sicheren Mensch-Roboter-Kooperation dies tun.

Ein Vorteil bei der Entwicklung des durch [Henrich u.a. 2008b] vorgeschlagenen Systems liegt in der Verwendung eines einfachen Ansatzes zur Rekonstruktion der Szene. Dieser Ansatz wird durch Integration von Epipolarlinien weiter robustifiziert. Des Weiteren können durch diese Vorgehensweise Objekte an der Werkzeugspitze erkannt werden.

Nachteilig ist hier, dass es keine Bestrebung gibt, den Menschen oder zumindest einzelne Glieder seiner kinematischen Kette zu erkennen, obwohl mehrmals von Mensch-Roboter-Kooperation die Rede ist. Stattdessen wird auch hier auf Farbinformationen von handelsüblichen Kameras gesetzt und die sich damit ergebenden Probleme nicht diskutiert. In den zitierten Arbeiten werden wenige experimentelle Untersuchungen durchgeführt, welche die Tauglichkeit der Verfahren diskutieren lassen.

Karlsruher Institut für Technologie (KIT), Deutschland

Die Universität Karlsruhe (TH) ist mit mehreren Forschungsgruppen an der Weiterentwicklung von Mensch-Maschine-Systemen beteiligt.

Die Forschungsgruppe um Herrn Prof. Wörn zielt auf Verfahren für die industrielle und humanoide Robotik.

[Heiligensetzer 2003]

In der Dissertation von [Heiligensetzer 2003] wird ein System für eine sichere Mensch-Roboter Kooperation vorgestellt, welches auf der Auswertung von drei Überwachungssensoren basiert:

- kapazitive Näherungssensoren,
- taktile Flächensensoren,
- Sensoren für die Überwachung von Motorströmen.

Es wird hervorgehoben, dass die Verwendung mehrerer Sensoren die Sicherheit erhöht, da für den Fall eines Sensorausfalls zwei weitere Sensoren zur Verfügung stehen, welche die Sicherheit gewährleisten (Redundanzprinzip). Ein Kraft-Momenten Sensor wird zusätzlich eingesetzt, um eine direkte, auf physischem Kontakt basierende Kooperation zwischen Mensch und

Roboter zu ermöglichen. Dadurch sollen die besonderen Fähigkeiten des Menschen und des Roboters genutzt werden, um zukünftig neue Anwendungsfelder zu erschließen oder bestehende Anwendungen zu vereinfachen. Die Vereinfachungen haben Einfluß auf die Konzeption und Realisierung von gemeinsamen Arbeitsräumen.

Des Weiteren hat die Dissertation weitere Schwerpunkte in der Umweltmodellierung und der Kommandierbarkeit des Roboters durch den Menschen.

[Yigit 2005] In dieser Dissertation werden Methoden einer taktil gekoppelten Kooperation zwischen einem Menschen und einem anthropomorphen Roboterarm erarbeitet. Es wird darauf hingewiesen, dass die Verwendung von Sensoren Kosten bzgl. des Raum- und Konstruktionsaufwandes einsparen können. Die Einsparung von Raum geht z.B. damit einher, dass entsprechende Schutzvorrichtungen (derzeitig trennende Schutzzäune) eingespart werden können. Deshalb wird auch in dieser Arbeit auf Sicherheitsfragen eingegangen.

Gefordert wird ein System, welches in der Lage ist, die Intentionen des Benutzers vorausschauend zu erkennen, proaktiv zu handeln, um so effiziente Sicherheitsmechanismen zu gewährleisten und im Sinne der Kooperation entsprechendes Hintergrundwissen zur Verfügung zu stellen. Das in [Yigit 2005] vorgeschlagene System konzentriert sich bzgl. dieser Aspekte auf Methoden für einen taktil gekoppelten Roboterarm und stützt sich dabei auf eine in [Stelling 1999] formulierte Definition der Kooperation bei Mensch-Maschine Systemen:

Zeitlich eng begrenzte Zusammenkunft zwecks Bewältigung eines gemeinsamen Ziels.

Durch Verwendung eines sogenannten Handlungsautomaten zur Modellierung von kooperativen Aufgaben durch Verkettung von Aktionsprimitiven, wird ein Beitrag zum Szeneverstehen geleistet. Durch die Aktionsprimitive besteht die Möglichkeit, Sicherheitsaspekte in die Kooperation zu integrieren. Es werden folgende wichtige Aspekte in dieser Arbeit geleistet:

- Eine neuartige reaktive Regelung auf der Grundlage von Oberflächenkräften,
- Modellierung von Reflexen für nicht vorhergesehene Kontakte,
- Eine Kollisionserkennung.

Eine Forschungsgruppe um Herrn Prof. Stiefelhagen integriert ihre Arbeit in den Serviceroboter Armar-III.

[Nickel 2008] Diese Arbeit leistet einen Beitrag zur visuellen Benutzermodellierung mit Personenverfolgung und Zeigegestenerkennung. Die Arbeit wird auch für den humanoiden Roboter ARMAR-III [Asfour u.a. 2006] verwendet.

Die in der Arbeit ausgearbeiteten Verfahren werden in Mehrkamerasystemen eingesetzt. Die Verfahren realisieren eine vollständige Perzeptionskette zur Lokalisierung von Personen, der Verfolgung von Händen und Orientierungsänderungen des Kopfes, sowie der Erkennung menschlicher Zeigegesten.

Die Personenlokalisierung basiert auf einem Partikelfilter [Isard und Blake 1998]. Die Robustheit des Algorithmus wird u.a. durch eine dynamische Fusion verschiedener, aus Bildfolgen extrahierter Merkmale erreicht. Die Merkmale sind im Wesentlichen Farbwerte und durch Bewegungen erzeugte Pixelverschiebungen.

Durch die Verwendung mehrerer Merkmale verhält sich das System robust, auch wenn einzelne Merkmale nicht aus den Mustern extrahiert werden können, z.B. wegen Verdeckungen oder aber bei Ausfall einzelner Kameras.

Das System wird überwiegend auf einem Stereokamerakopf eingesetzt. Die Erkennung von Zeigegesten geschieht auf der Grundlage von Hidden-Markov-Modellen. Um die a priori benötigten Wahrscheinlichkeiten für die Zustandsübergangsmatrizen zu bestimmen, wurden eine Vielzahl von Trainingstrajektorien ausgewertet. Die Integration der Blickrichtung für die Schätzung der Zeigegeste konnte die Erkennungsrate zusätzlich steigern.

Auch in dieser Arbeit wird versucht, durch Farbwerte die Schätzergebnisse zu verbessern, ohne jedoch deutlich auf die wohlbekanntesten Nachteile einzugehen.

Offene Fragen beziehen sich auf das Verhalten des Systems bei Verdeckungen und die Echtzeitfähigkeit.

Institut für Anthropomatik Das Institut für Anthropomatik, geleitet von Herrn Prof. Dillmann, hat in der Vergangenheit in großem Maße zur Weiterentwicklung der Service-Robotik und dadurch auch im Speziellen zur Mensch-Roboter-Interaktion beigetragen. Im Folgenden werden deshalb ausgewählte aktuelle Forschungsarbeiten aus der Forschungsgruppe um Herrn Prof. Dillmann vorgestellt.

Eine grundlegende Arbeit für die sichere Mensch-Roboter-Interaktion ist die Schätzung der menschlichen Pose aus Sensordaten. Ein flexibles System zur Integration multi-modaler Eingabedaten ist das VooDoo-System [Knoop u.a. 2006].

Das VooDoo-System erlaubt die Verfolgung einer menschlichen Pose mit mindestens 10 Bildern pro Sekunde. Dabei wird versucht, ein intern repräsentiertes 3D Mensch-Modell an die Sensordaten anzupassen. Dies gelingt u.a. durch Einsatz des ICP-Algorithmus [Horn 1987].

Verarbeitet werden Rohdaten von unterschiedlichen Sensoren (2D und 3D). Insbesondere auf Merkmalen basierte 2D Verarbeitung und 3D Sensordaten spielen eine signifikante Rolle. Auch die Integration von einem auf dem Lichtflugzeit-Prinzip basierenden Kamerasystem wird berücksichtigt.

Das 3D Menschmodell wird grafisch durch Kegelstümpfe dargestellt. Die interne Strukturierung findet dabei in einer baumartigen Struktur statt. Es werden bis zu 10 Kegelstümpfe verwendet, um die menschliche Pose zu modellieren.

Eine Erweiterung des VooDoo-Systems wurde in [Löschet u.a. 2009] publiziert. Dabei geht es um die automatische Initialisierung eines Mensch-Modells am Anfang der Verfolgung. Vorausgesetzt wird lediglich, dass der Benutzer frontal zum System steht. Die eigentliche Initialisierung gelingt durch eine in zwei Phasen aufgeteilte Vorgehensweise. Dabei kommt ein Silhouetten-basiertes Verfahren in Kombination mit einer kaskadierten Metrik zum Einsatz. Die Initialisierung verhält sich robust gegenüber Beleuchtungsvariationen in der Szene und ist dadurch vermutlich robuster als die bisher vorgestellten Verfahren. Es wird von den Autoren darauf hingewiesen, dass dies deshalb gelingt, da aus Farbbildern erzeugte Merkmale nicht berücksichtigt werden. Allerdings verwendet das Verfolgungssystem VooDoo anschließend Farbmerkmale bei der Verfolgung von Personen, insbesondere des Kopfes und der Hände.

Die erste Phase bildet Hypothesen über das Vorhandensein einer Person aus den Sensordaten. Die zweite Phase verifiziert die Hypothesen und stimmt diese ab.

Die gesamte Initialisierung benötigt maximal 20 Sekunden.

Ein weiteres entwickeltes System zur markerlosen Verfolgung, basierend auf Stereobildern, wird in [Azad u.a. 2008] vorgestellt. Das System basiert auf einem erweiterten Partikel-Filter mit dem Ziel, den Oberkörper des Menschen in Echtzeit zu verfolgen. Wichtige Merkmale stammen dabei von 3D (Farb-)Informationen über die Position der Hände und des Kopfes. Außerdem werden zusätzlich Tiefendaten verarbeitet die vom Stereokamerasystem bereitgestellt

werden. Der Partikelfilter wird u.a. dazu verwendet, die inverse Kinematik der Arme zu schätzen. Die Vorgehensweise sieht auch eine priorisierte Fusionsmethode, sowie eine adaptive Schulter-Positions-Schätzung vor. Die gesamte Oberkörperverfolgung gelingt mit 15 Bildern pro Sekunde, inklusive einer vollautomatischen Initialisierung.

Das zugrundeliegende Mensch-Modell umfasst 14 Freiheitsgrade. Die Freiheitsgrade des Nackens und die Rotationsparameter der Armgelenke um die eigene Achse werden nicht berücksichtigt. Die gesamten Parameter werden durch einen einzelnen Partikelfilter geschätzt. Die Schätzung verhält sich robust, falls die Person frontal zum Stereokamerasystem steht. Die Bewegungsschätzungen des Menschen können bspw. auf den humanoiden Roboter Armar-III übertragen werden [Do u.a. 2008].

2.3 Normen und Vorschriften

Die im Rahmen dieser Arbeit eingesetzten und entwickelten Methoden und das anschließend zusammengesetzte Gesamtsystem erhebt nicht den Anspruch, eine Sicherheitszertifizierung zu erlangen.

Das hier vorgestellte System kann dies nicht leisten, da bis heute keine speziell für diese Thematik erforderlichen Normen und Vorschriften veröffentlicht wurden. D.h. konkret, es gibt bis heute keine Norm, die direkt Bezug nimmt zur sicheren Mensch-Roboter-Kooperation. Insbesondere müsste diese Norm eine Position zum physischen Kontakt beziehen, welche bei der Kooperation eintreten kann. Wie mit dieser Tatsache umgegangen wird, soll in den folgenden Abschnitten geklärt werden. Dazu werden im Folgenden relevante Normen und Richtlinien aufgezählt und anschließend erläutert:

- EG-Maschinenrichtlinie [EG2006] (vormals [EG98])
- Grundbegriffe; Sicherheit von Maschinen [DIN12100] (vormals [DIN292])
- Sicherheit bei Industrieroboter [DIN10218] (vormals [DIN775])
- Trennende Schutzeinrichtungen [DIN953]
- Sicherheitsbezogene Teile von Steuerungen [DIN954-1]
- Leitsätze zur Risikobeurteilung [DIN1050]
- Nutzungssicherheit kraftbetätigter Tore [DIN12453]

Die [EG2006] definiert grundlegende Sicherheits- und Gesundheitsanforderungen. Die Erfüllung der Sicherheitsanforderungen wird als Konformitätsnachweis bezeichnet und muss von jedem Hersteller eigenverantwortlich erbracht werden. Dies gilt insbesondere für Maschinen, die in der Europäischen Union in Verkehr gebracht werden. Der Konformitätsnachweis wird vom Hersteller auf dem Produkt durch das CE-Zeichen ausgedrückt. Im Gegensatz zur alten [EG98] ist in der neuen Richtlinie festgelegt, dass auch unvollständige Maschinen (Teilmaschinen) die Richtlinie erfüllen können.

Unvollständige Maschinen können keine bestimmte Tätigkeit ausführen. Sie müssen zuerst eingebaut werden in eine Anlage oder Gesamtmaschine inkl. aller Schutzeinrichtungen. Danach gilt die Maschine als vollständig. Dazu gehören z.B. die Industrieroboter (Bearbeitungsmaschinen).

Die Aufnahme von Teilmaschinen wie Industrieroboter in die Maschinenrichtlinie wird dadurch gerechtfertigt, dass diese über richtlinienkonforme Ausrüstungen verfügen, z.B. Umhausungen und Not-Halt-Schalter. Für Teilmaschinen muss deshalb zukünftig eine Einbauerklärung und eine zusätzliche Risikobewertung angefertigt werden. Die Aushändigung der Risikobewertung muss durch den Kunden vertraglich geregelt werden.

Um das CE-Zeichen verwenden zu dürfen, muss das Konformitätsbewertungsverfahren durchlaufen werden.

In den Anwendungsbereich der Maschinenrichtlinie fallen derzeit auch die sogenannten Sicherheitsbauteile. Dazu zählen auch Schutzeinrichtungen zur Detektion von Personen. Da diese jedoch auch anderen Richtlinien unterliegen, müssen zusätzliche Qualitätssicherungsmaßnahmen erfüllt werden.

Ein weiterer Hinweis in den Richtlinien bezieht sich auch auf automatische Maschinen, zu denen ebenfalls die Industrieroboter gezählt werden.

Darin wird gefordert, dass bestimmte automatische Abläufe beobachtet und quittiert werden müssen, z.B. durch Zustimmungsschalter, die während des Beobachtungszeitraums gedrückt werden müssen. Vom Anwendungsbereich der Sicherheitsanforderungen ausgenommen sind insbesondere für *Forschungszwecke entwickelte und hergestellte Maschinen*. Das umfasst insbesondere die in dieser Arbeit vorliegenden Entwicklungen und Experimente.

In der [DIN12100] werden Normen für die Sicherheit von Maschinen geregelt. Es handelt sich um eine sogenannte Typ A Norm (Sicherheitsgrundnorm), die Grundbegriffe, Gestaltungsleit-sätze und allgemeine Aspekte für Maschinen, Geräte und Anlagen klärt. In dieser Norm werden auch die Kategorien des durch die Maschine verursachten Gefährdungspotentials bestimmt.

Spezielle Fragen zur Sicherheit von Industrierobotern sind in der [DIN10218] geregelt. Wobei der Anwendungsbereich der [DIN10218] die Konstruktion (Teil 1) und der Bau des Roboters (Teil 2) ist und damit maximal die technischen Voraussetzungen für den assistierenden Betrieb festlegt. Die [DIN10218] ist auch für die Integration der Roboteranlage verantwortlich. Im Falle eines gemeinsamen Arbeitsraumes, unter Berücksichtigung eines autonomen, automatischen Betriebs, muss die Arbeitsgeschwindigkeit entsprechend des in der [DIN13855] geforderten Mindestabstandes einhalten und bei Unterschreitung des Mindestabstandes gewährleistet werden, dass ein sicherer Stopp stattfindet.

Eine Alternative besteht in der Leistungs- und Kraftbegrenzung durch Verwendung einer sicheren Steuerung nach [DIN13849] (Kategorie 3) (siehe Abbildung 2.6). Die grundlegenden Details findet man in [DIN10218] (Abschnitt 5.9).

Die bisher verwendeten trennenden Schutzeinrichtungen sind durch [DIN953] geregelt. Sie sind Teil der Sicherheit von Industrierobotern. Das Dokument unterscheidet zwei Hauptgruppen. Zum einen die feststehenden, trennende Schutzeinrichtungen und die beweglichen, trennenden Schutzeinrichtungen. Darüberhinaus werden Hinweise über die Gestaltungsanforderungen festgelegt. Aus einem im Dokument angegebenen Diagramm können angemessene Schutz-einrichtungen ausgewählt werden.

In der Norm [DIN954-1] werden Anforderungen an sicherheitsbezogene Teile von Steuerungen beschrieben. Dazu wird ein iterativer Prozess formuliert, welcher aus den folgenden fünf Schritten besteht:

1. Gefährdungs- und Risikoanalyse
2. Maßnahmen zur Verringerung des Risikos durch Methoden der Steuerung
3. Festlegung von sicherheitstechnischen Anforderungen

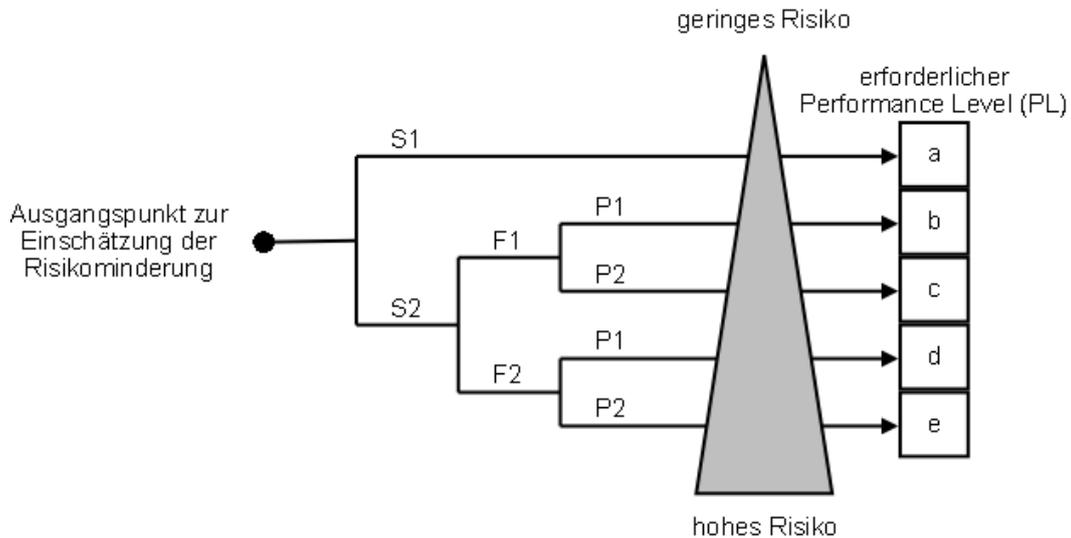


Abbildung 2.6: Einordnung in die Sicherheitskategorien nach [DIN13849].

4. Gestaltung
5. Validierung

Zusätzlich werden die sicherheitsbezogenen Teile der Steuerung in Kategorien eingeteilt:

- **Kategorie B** Das Auftreten eines Fehlers kann zu Verlust der Sicherheitsfunktion führen.
- **Kategorie 1** Mildert die Forderungen von Kategorie B ab, indem von einer geringen Auftretswahrscheinlichkeit des Einzelfehlers ausgegangen wird. Um für eine geringe Ausfallwahrscheinlichkeit zu garantieren, müssen sicherheitstechnisch bewährte Bauteile verwendet werden .
- **Kategorie 2** Die Maschinensteuerung muss in angemessenen Zeitabständen die Sicherheitsfunktion überprüfen. Auch hier kann es zu Ausfällen der Sicherheitsfunktion kommen.
- **Kategorie 3** Es wird gefordert, dass Einzelfehler in angemessener Zeit erkannt werden. Dadurch soll garantiert werden, dass die Sicherheitsfunktion nicht außer Kraft gesetzt wird.
- **Kategorie 4** Fehlerquellen, die zum Ausfall der Sicherheitsfunktion führen, müssen rechtzeitig erkannt werden. Damit soll die Grundlage geschaffen werden, dass die Sicherheitsfunktion erhalten bleibt, trotz mehrfachen Fehlereintritts.

Die Zugehörigkeit einer Schutzeinrichtung zu einer der eben genannten Kategorien hängt von den folgenden Faktoren ab:

- **S Schwere der Verletzung** Hier wird zwischen leichten (reversiblen; S1) und schweren (irreversiblen oder Tod; S2) Verletzungen unterschieden.

- **F Häufigkeit und/oder Dauer der Gefährdungsexpansion** Hier wird zwischen seltener bzw. kurzer Exposition (F1) und häufiger bzw. lange andauernder Exposition (F2) unterschieden.
- **P Möglichkeit zur Vermeidung der Gefährdung** Es kommt zu einer Klassifizierung nach möglichem (P1) und kaum möglichem Vermeiden der Gefährdung (P2).

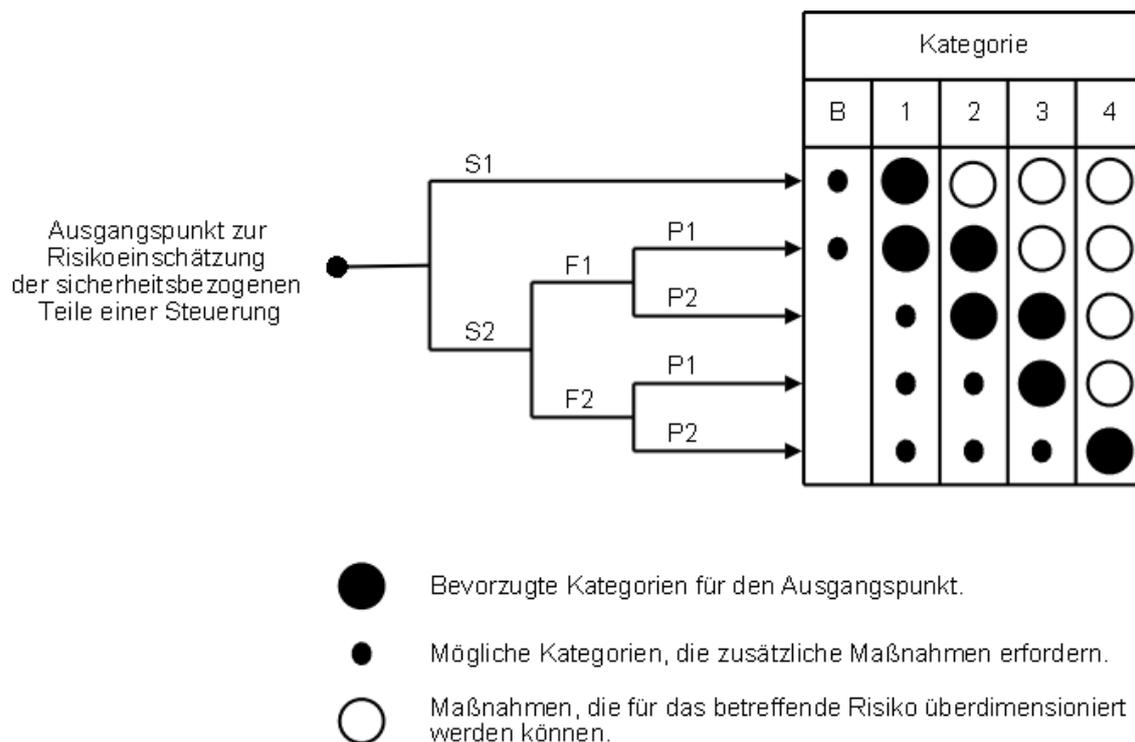


Abbildung 2.7: Einordnung in die Sicherheitskategorien nach [DIN954-1].

Bei der direkten Kooperation zwischen Mensch und Roboter existiert eine Gefahrenquelle, die u.a. durch die Dynamik des Roboters provoziert wird. Die Sicherheit für den Menschen bei der Kooperation wird durch die hohe Arbeitsgeschwindigkeit des Roboters und der damit verbundenen kinetischen Energie gefährdet. Deshalb müssen steuerungstechnische Maßnahmen, z.B. durch Reduktion der Arbeitsgeschwindigkeit oder Ausweichbewegungen, vorgenommen werden.

Während der Mensch-Roboter-Kooperation teilen sich der Mensch und der Roboter einen gemeinsamen Arbeitsraum. Deshalb ist von einer sogenannten langen Gefährdungsexpansion auszugehen. Des Weiteren muss von irreversiblen Verletzungen ausgegangen werden, falls es zu einem Unfall kommt. Das in dieser Arbeit vorgestellte System muss demnach mindestens die Sicherheitskategorie 3 der [DIN954-1] erfüllen (siehe auch Abbildung 2.7).

Die [DIN1050] ist eine Norm zur Definition der Risikoelemente. Dabei wird das Risiko, bezogen auf die betrachtete Gefährdung als eine Funktion von Ausmaß des möglichen Schadens, der durch die betrachtete Gefährdung, verursacht werden kann, und der Wahrscheinlichkeit des Eintritts dieses Schadens, betrachtet. Die Wahrscheinlichkeit des Schadeneintritts wird bestimmt durch:

- Häufigkeit und Dauer der Gefährdungsexposition.
- Eintrittswahrscheinlichkeit eines Gefährdungsereignisses.
- Möglichkeit zur Vermeidung oder Begrenzung des Schadens.

Die Norm fordert die Durchführung einer Risikoanalyse und Risikobewertung.

Die auf den ersten Blick sonderbar erscheinende Forderung, dass die quantitative Bewertung der Sicherheit für Industrieroboter durch die Norm zur Nutzungssicherheit für kraftbetätigte Tore [DIN12453] geregelt wird, ist insofern verständlich, da es bis heute keine explizite Norm für Industrieroboter und deren Dynamik gibt. In der Norm wird ein Grenzwert für die zulässige Kontaktkraft angegeben, wobei die aus der Dynamik des Roboters entstehenden Kontaktkräfte weniger als 400 N betragen dürfen. Wird dieser Wert überschritten, muss garantiert werden, dass die statische Kontaktkraft nach maximal 0,75 s auf den Grenzwert von 150 N reduziert wird.

Da es keine Normen für eine sichere Mensch–Roboter–Kooperation gibt, werden Normen verwendet, die aus möglichst naheliegenden Anwendungsbereichen kommen.

Entscheidend für diese Arbeit ist, dass die grundsätzlichen Forderungen eines als sicher geltenden Systems nicht gänzlich außer Acht gelassen werden sollen. Deshalb sollen im Rahmen dieser Arbeit verschiedene Methoden eingesetzt und entwickelt werden, welche die Sicherheit für den Menschen gemäß der hier vorgestellten Normen fördert, auch wenn eine Sicherheitszertifizierung nicht Gegenstand dieser Forschungsarbeit ist.

2.4 Abschließende Bemerkung und Diskussion

In diesem Abschnitt wurde ein Überblick zum Stand der Forschung und Entwicklung für den Bereich sichere Mensch–Roboter–Kooperation zu vermitteln. Damit können die in dieser Arbeit entwickelten Methoden von den vorgestellten Realisierungen abgegrenzt werden. Das hier erarbeitete System setzt hauptsächlich auf ein neuartiges Kamerasystem [O3D-PMD], welches auf dem Lichtflug–Prinzip basiert.

Aus diesem Grund wurde diesem Kapitel ein zusätzlicher Abschnitt über Sensoren gewidmet, um Unterschiede bzw. Vor- und Nachteile zwischen den verschiedenen Sensoren darzustellen. Der hier verwendete Sensor verfügt über Schwächen, wie bspw.:

- die verrauschten Tiefenmessungen,
- Bewegungsartefakte und
- eine vergleichsweise geringfügige Auflösung.

Andererseits gibt es Vorteile, wie bspw.:

- dichte Tiefenkarten, die mit bis zu 20 Bildern pro Sekunde zur Verfügung gestellt werden,
- eine Fremdlichtunterdrückung und
- die Möglichkeit des Nachtsehens.

Andere Sensoren haben ihrerseits Vor- und Nachteile.

Es wurde bewusst darauf geachtet, den Stand der Forschung und Entwicklung möglichst umfassend zu beschreiben. Deshalb wurde eine grobe Einteilung nach Firmen, die bereits verfügbare Produkte herstellen oder gerade mit der Entwicklung beschäftigt sind, zu diskutieren, aber insbesondere Forschungsgruppen und Forschungseinrichtungen. Die Auswahl muss im Rahmen dieser Arbeit auf eine begrenzte Anzahl von Stellvertretern beschränkt werden. Es wurde versucht, diese so auszuwählen, dass einerseits akute Forschungstrends aus speziellen Disziplinen vorgestellt werden und andererseits Systeme diskutiert werden, die das allgemeinere Ziel der sicheren Kooperation bzw. Interaktion haben.

Außerdem ist ein Abschnitt über Normen und Richtlinien integriert. Im Rahmen dieser Forschungsarbeit geht es nicht darum, eine Sicherheitszertifizierung anzustreben. Trotzdem haben die Anforderungen für die Sicherheitszertifizierung den Entwurf der Systeme⁷ beeinflusst.

In den folgenden Kapiteln wird gezeigt, wie zwei Systeme realisiert werden, welche mit einem einzelnen Sensor in der Lage sind, eine sichere Mensch-Roboter-Kooperation bzw. – Interaktion zu gewährleisten.

⁷Es werden im Rahmen dieser Arbeit zwei Systeme erarbeitet!

Kapitel 3

Systemarchitektur

In diesem Kapitel werden zwei Architekturen für die Mensch–Roboter–Kooperation vorgestellt und diskutiert. Die in den nachfolgenden Kapiteln vorgestellten Methoden sind die Grundlagen der beiden Architekturen.

Der erste Entwurf wird realisiert, um einen auf Techniken der Bildfolgenauswertung konzentrierten Demonstrator zu entwerfen.

Die zweite Architektur ist eine Weiterentwicklung des ersten Entwurfs. Die softwaretechnische Realisierung und der reale Demonstrator der beiden Architekturen lässt aber klare Unterschiede erkennen. Die Unterschiede werden in den folgenden Abschnitten dargestellt.

Statt die gesamte Auswertung auf Methoden zu stützen, welche keine Meta–Informationen aus den Bildmustern erschließen, wird bei der zweiten Architektur Wissen aus den Bildmustern extrahiert.

Schließlich werden Merkmale aus ausgewählten Bildpunkten gebildet, die letztlich eine Rekonstruktion der realen Szene ermöglichen. Die Rekonstruktion beschränkt sich dabei auf wenige wichtige Parameter, die im Wesentlichen die Kinematik der Roboter und Teile der menschlichen Kinematik umfassen.

3.1 Entwurf einer ersten Systemarchitektur

Die Auswertung und Interpretation von Bildfolgen, aufgezeichnet von Sensoren, nennt man Bildfolgenauswertung. Dabei können mehrere verschiedenartige Sensoren zum Einsatz kommen. Werden mehrere Sensoren eingesetzt, so müssen die aufgezeichneten Bilder fusioniert werden. Es muss dabei beachtet werden, dass die einzelnen Aufnahmen von unterschiedlichen Sensortechnologien kommen können (heterogene Fusion) und die Aufzeichnungspunkte synchronisiert werden müssen. Im Rahmen dieser Dissertation kommt es lediglich zum Einsatz eines homogenen Sensornetzes, bestehend aus PMD–Sensoren, bzw. eines einzelnen PMD–Sensors.

Sind alle Aufnahmen korrekt erfasst, wird mit der Segmentierung begonnen. Unter Segmentierung soll hier die Trennung des Vordergrundes vom Hintergrund verstanden werden. Dies geschieht bei Einsatz der PMD–Technologie unter Verwendung verschiedenartiger, aus unterschiedlichen Mustern (Tiefen–, Amplituden– und Grauwertinformationen) generierte, Merkmale. Im Wesentlichen sind dies:

- Struktur,

- Tiefe und
- Bewegung.

Sobald die Segmentierung abgeschlossen ist, wird mit der eigentlichen Identifikation von Bildpunkten des Vorder- und Hintergrunds begonnen. Dazu werden a priori bekannte Modelle herangezogen, welche iterativ gegen die einzelnen Segmente registriert werden.

Dieser Sachverhalt wird anhand der Rekonstruktion der menschlichen Kinematik aus Bildfolgen veranschaulicht:

Durch eine Bewegungsanalyse und unter Einsatz einer sogenannten Hintergrundsubtraktionstechnik (siehe bspw. [Stauffer und 1999]) werden bewegte Objekte in der Szene identifiziert. Die Tiefeninformation, welche die PMD-Technologie in Echtzeit liefert, ist ein weiteres Muster, um zwischen Objekten und Agenten zu unterscheiden. Um die Unterscheidung zwischen Agent und Objekt zu verbessern, wird weiteres Vorwissen in den Schätzprozess integriert. Dies wird im Rahmen der Schätzung der menschlichen Kinematik durch Verwendung eines parametrisierten Modells (Hüllkörpermodell) realisiert. Da die Anpassung des Menschmodells ein nichtlineares Schätzproblem darstellt, werden iterative Löser verwendet.

Nachdem man eine annehmbare Schätzung gefunden hat, lassen sich die identifizierten Szenobjekte in ein Umweltmodell integrieren. Das Umweltmodell umfasst hier ein Menschmodell, verschiedene Objektmodelle (z.B. Steuerungskästen der Roboter usw.) und zwei Robotermodelle. Die bekannten Parameter der Modelle erlauben es, die Geometrien abzuleiten und diese der Bahnplanung inklusive Abstandsberechnung zur Verfügung zu stellen.

Innerhalb der Bahnplanung kommt es zu einer Verarbeitung der geometrischen Informationen, so dass die Eingabedaten effizient verarbeitet werden können.

Die Ergebnisse der Bahnplanung werden anschließend der Robotersteuerung weitergereicht. Die Ausführungen der Ergebnisse der Bahnplanung auf der Robotersteuerung beeinflussen das Robotermodell. Der Zyklus schließt sich, wenn in einem neuen Zeitschritt weitere Aufnahmen der eingesetzten Sensoren gemacht werden.

Das im Rahmen dieser Arbeit verwendete Menschmodell verfügt über mehr als 40 Freiheitsgrade, wobei derzeit nur wenige, aber für die Sicherheit des Menschen wichtige Parameter geschätzt werden. Bei den geschätzten Parametern handelt es sich für die erste Architektur um die Position und Orientierung des Mensch-Modells. Bei der Realisierung der zweiten Architektur sind weitere Parameter, wie bspw. die Arme, die Beine und die Kopforientierung in den Schätzprozess integriert.

Bei der ersten Architektur wird versucht die Hülle des Menschen präzise zu erfassen und auf der Grundlage der umhüllenden Kurve den Abstand des Menschen zum Roboter zu schätzen (siehe Abbildung 3.1).

Die der Konturverfolgung zugrundeliegende Idee ist es, die Schätzungen des Optischen Flusses in die Konturverfahren zu integrieren, um damit einen geschlossenen systematischen Ansatz für die dynamische Konturverfolgung zu erhalten. Die technischen Details werden später in 4.3.2 motiviert. Die für die Modellgleichung verwendeten Teilgleichungen für die aktiven Konturen stammen aus [Xu und Prince 2000], die Gleichungen für die Berechnung der Optischen-Flussfelder sind durch [Brox u.a. 2004] motiviert und für die Bedürfnisse der hier vorliegenden Anwendung weiterentwickelt (siehe [Graf und Wörn 2008a] und [Graf und Wörn 2008c]). Neuartig ist u.a. die Kombination beider Ansätze unter Berücksichtigung effizienter numerischer Löser, wie bspw. die Methode der konjugierten Gradienten ([Graf und Wörn 2008b]). Dadurch können im Hinblick auf die Anzahl der Freiheitsgrade und die Berechnungsgeschwin-

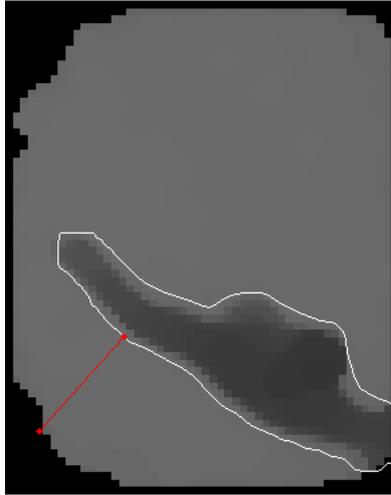


Abbildung 3.1: Abstandsberechnung zwischen Mensch und Roboter, basierend auf einem Hüllmodell.

digkeit, präzise Algorithmen in das System integriert werden. Die Echtzeitfähigkeit bleibt erhalten (>15 Bilder pro Sekunde).

Um sich vom bisherigen Stand der Forschung abzugrenzen, werden in der folgenden Liste die wichtigen Unterschiede aufgelistet:

1. Die bisherigen Sensoren, basieren auf einfachen (Farbbild-) CCD-Kameras, wohingegen in dieser Arbeit ein ressourcenschonender Aufbau der Demonstratorzelle mit nur einer einzelnen Tiefenbildkamera realisiert wird.
2. Sofern parametrische Kurven verwendet werden, wird in den entsprechenden Veröffentlichungen über kein systematisches Vorgehen berichtet. Um Bewegungen zu schätzen, wird auf Ansätze zurückgegriffen, die nicht auf Bewegungsmustern basieren, sondern stattdessen auf Muster wie bspw. Hautfarbe oder anderen heuristischen Merkmalen. Im Rahmen dieser Arbeit wird stattdessen eine neuartige Grundgleichung für die Konturerfolgung eingeführt. Diese Gleichung basiert auf einem Variationsansatz, der zwei wichtige Methoden der Bildauswertung verknüpft und durch Umschreiben der Modellgleichungen von [Brox u.a. 2004] können effiziente Löser verwendet werden, welche eine echtzeitfähige Anwendung möglich machen (siehe [Graf und Wörn 2008b]).

Die im Rahmen dieser Arbeit entwickelten Ansätze basieren, so weit nicht anders angegeben, auf der Auswertung von Tiefenbildfolgen, eines an der Decke der Demonstratorzelle montierten Sensors. Details zu den hier verwendeten Verfahren werden in den späteren Abschnitten diskutiert.

Die Abstandsmessung zwischen Mensch und Roboter geschieht hier durch Berechnung des kürzesten Abstandes von der Kontur des Menschen zur Kontur des Roboters (siehe bspw. Abbildung 3.1).

3.2 Entwurf einer zweiten Systemarchitektur

Die Mensch–Roboter–Kooperation wird in der Forschung zur Zeit derart realisiert, dass der Einsatz verschiedenartiger Sensoren, wie bspw. Farbbildkameras [Thiemermann 2005], Stereokameras bzw. Mehrkamerasysteme [Henrich und Gecks 2008a], Fusion haptischer und kapazitiver Sensoren [Heiligensetzer 2003], künstlicher Haut [Göger u.a. 2006] und photonische Mischsensoren (PMD, mit 64x50 Bildpunkten Auflösung) [Winkler 2008] zum Einsatz kommen, um mit diesen Sensoren eine möglichst sinnvolle Kooperation bzw. Interaktion zu ermöglichen.

Bisher wurde das während der Kooperation entstehende Risiko für den Menschen, durch die kürzeste Distanz zwischen beiden Akteuren bestimmt (siehe bspw. [Kuhn u.a. 2007]). Um das Risiko zu minimieren, wurden in der Vergangenheit einfache Strategien, wie bspw. die Reduktion der Arbeitsgeschwindigkeit vorgeschlagen [Kuhn u.a. 2006], oder aber, der Roboter weicht aus [Gecks und Henrich 2007].

In diesem Abschnitt wird eine Systemarchitektur vorgestellt werden, die den bisherigen Stand der Forschung dahingehend erweitert, dass auf der Grundlage eines neuartigen optischen Sensorsystems die Erkennung des Menschen, d.h. Teile seiner Kinematik, markerlos in Echtzeit auf ein virtuelles Menschmodell übertragen werden und damit schließlich eine flexiblere Kooperation zwischen Menschen und Roboter stattfinden kann. Bisherige Ansätze betrachten lediglich Bildpunktballungen (z.B. [Kuhn u.a. 2007]) oder eine Approximation durch grobe geometrische Primitive (z.B. [Kuhn u.a. 2006]), ohne genau zu verstehen, was in der Szene vor sich geht. Eine derartige Vorgehensweise hat den Nachteil, dass eine direkte Kooperation nicht sinnvoll stattfinden kann, zumindest nicht dann, wenn man erreichen möchte, die Szene zu interpretieren und um damit Auswerteverfahren eine fruchtbare Grundlage für weitere Entwicklungsmöglichkeiten zu sorgen. Damit kann schließlich die Sicherheit für den Menschen, aber auch die Kooperationsfähigkeit erhöht werden - ein wichtiger Schritt hin zur kognitiven Industrierobotik.

Eine Voraussetzung für eine real stattfindende Kooperation kann dadurch definiert sein, dass der Mensch bspw. durch seine Hände mit dem Roboter in direkten Kontakt steht. Steht der Mensch allerdings mit dem Rücken nahe zum Roboter, verändert sich das Risiko für den Menschen. Ansätze, die den Menschen und seine Pose nicht berücksichtigen, werden in dieser Arbeit als kritisch eingestuft oder aber arbeiten weniger effizient. Kennt man jedoch die kinematische Konfiguration des Menschen, kann eine differenziertere Schätzung des Risikos stattfinden.

Außerdem wird ein Algorithmus zur Behandlung von sich verdeckenden Armen vorgestellt, welcher es erlaubt, Verdeckungen, die durch den Roboter oder aber den Menschen selbst erzeugt werden, zu behandeln.

Die zuverlässige Schätzung der menschlichen Pose, die effiziente Abstandsberechnung zwischen den in der Szene befindlichen Akteuren, sowie eine Quantifizierung des Risikos durch Verwendung einer zweisträngigen unscharfen Logik, bilden neben anderen wichtigen Methoden die Grundlage für die Systemarchitektur, welche auf den Tiefen- und Amplitudenbildern des PMD-Sensors basieren. Unter Berücksichtigung der Sensorinformationen werden klassische Verfahren der Bildfolgenauswertung (z.B. Optischer Fluss), der Schätztheorie (z.B. Kalman-Filter), einer (zweisträngigen) unscharfen Logik und des Maschinellen Lernens (z.B. Unterstützungsvektormethoden) für die Systemarchitektur entwickelt und experimentell untersucht.

3.2.1 MAROCO - Ein Softwaresystem für die sichere Mensch-Roboter-Kooperation

Die Realisierung der Systemarchitektur resultiert in einem modularen Softwaresystem mit der Bezeichnung MAROCO¹, welches eine Grundlage für zukünftige Untersuchungen komplexer menschenzentrierter Systeme bietet. Der gesamte Auswertezyklus inkl. risikominimierender Strategien (Reduktion der Arbeitsgeschwindigkeit und Ausweichbewegung der Roboter), geschieht dabei in Echtzeit, d.h. für mehr als 30 Bilder (Amplituden- und Tiefenbilder) pro Sekunde. Für die Auswertung wird ein handelsüblicher Rechner verwendet.

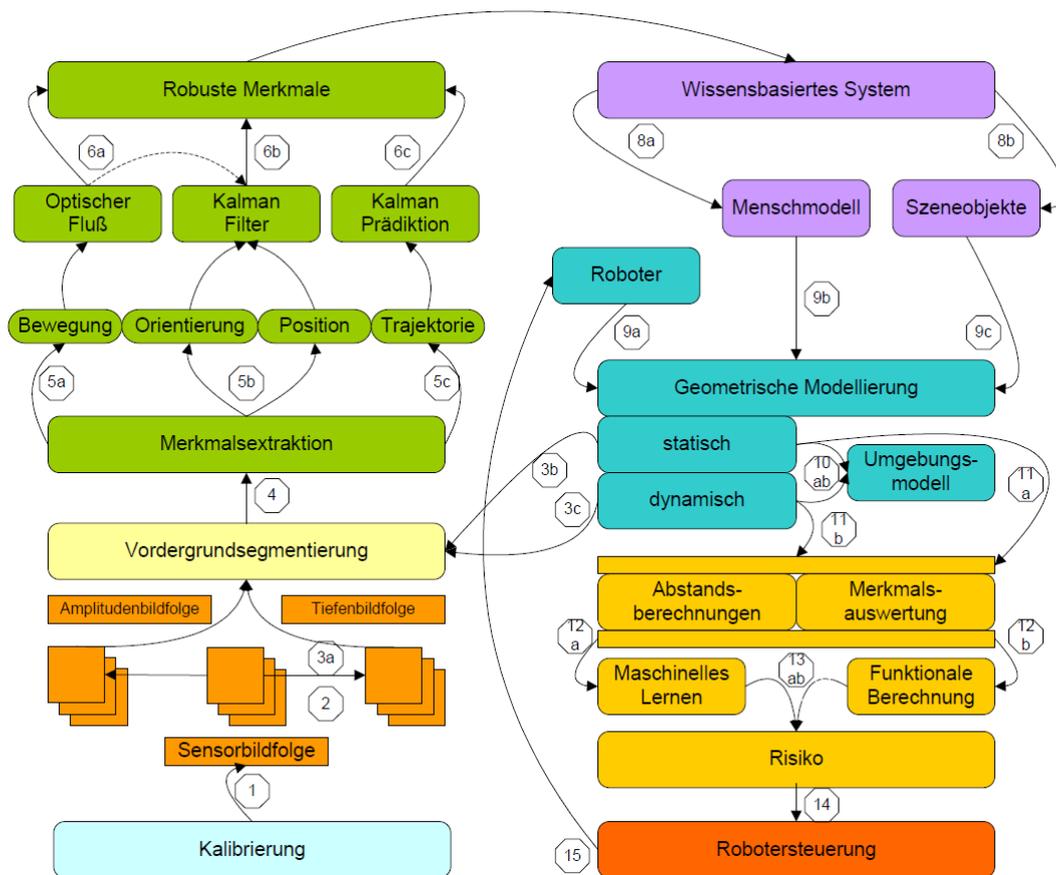


Abbildung 3.2: Eine Architektur für die sichere Mensch-Roboter-Kooperation als Auswertezyklus von Bildfolgen.

Die im Folgenden vorgestellte Architektur wurde bereits in [Graf und Wörn 2009a] detailliert beschrieben. Die in Abbildung 3.2 dargestellte modulare Architektur ist in Block-Notation angegeben und ist im Rahmen dieser Arbeit vollständig implementiert worden. MAROCO ist das daraus entstandene Software-System. Für die Module gab es keine, abgesehen von den Unterstützungsvektormethoden, zur Verfügung stehenden Bibliotheken. D.h. die Module mussten vollständig implementiert und getestet werden.

Mit MAROCO ist es möglich, die praktische Umsetzung zu visualisieren, aber auch Experi-

¹MAROCO ist das, im Rahmen der Dissertation, entwickelte Softwaresystem und setzt sich aus dem engl. Wörtern *huMAN ROBot COoperation* zusammen.

mente durchzuführen, diese aufzuzeichnen und damit reproduzierbar zu machen. Außerdem kann durch das System ein Vergleich zwischen konkurrierenden Ansätzen realisiert werden. In den folgenden Abschnitten sollen einige Details der Architektur erläutert werden.

Eine weitere Vertiefung findet anschließend in den folgenden Abschnitten statt, die sich auf Details der einzelnen Module konzentrieren.

Auf das spezielle Thema Kalibrierung wird hier nicht weiter eingegangen, da dafür bereits entsprechende Bibliotheken von Projektpartnern [Fuchs und Hirzinger 2008] oder haus eigenen Lösungen [Mamaev u.a. 2008] zur Verfügung standen.

Das Besondere am [O3D-PMD] ist, dass neben den Tiefeninformationen eine zusätzliche Informationsquelle zur Verfügung steht. Dabei handelt es sich um das sogenannte Amplitudenbild. Die Amplitudenwerte der O3D-Kamera sind ein Maß für die Remission der aktiven Beleuchtung und werden u.a. von der Remissionsfähigkeit der betrachteten Szenegegenstände beeinflusst. Die verschiedenen Informationen der Bildquellen werden an das nächste Modul weitergereicht (Abbildung 3.2(3a)).

Die Bildverarbeitung, im Gegensatz zur Bildfolgenauswertung, ist ein Vorgang, welcher Bilder in Bilder transformiert. Ein wichtiges Gebiet der Bildverarbeitung ist die Segmentierung, durch die wichtige Informationen, wie bspw. Bildpunkte korrespondierend zu Hindernissen oder Personen in der Szene, von unwichtigen Informationen, z.B. Rauschen oder Beleuchtungseinflüssen, getrennt werden.

Die in dieser Arbeit verwendete Vorgehensweise basiert auf Gaußschen-Mischverteilungen, die bildpunktweise angewandt werden. Da die Verteilungen zur Laufzeit an die sich verändernden Verhältnisse der Szene angepasst, handelt es sich um eine Methode des maschinellen Lernens. Im weiteren Verlauf der Arbeit wird aus dem Zusammenhang klar, auf welche Teildisziplin des maschinellen Lernens Bezug genommen wird.

Die hier verwendete Technik ist eine Weiterentwicklung der Arbeit von [Stauffer 2002]. Zusätzlich wird die Arbeit um einige Ideen von [Lee 2005] erweitert und schließlich, im Rahmen dieser Arbeit, adaptiert an die Bedürfnisse der Mensch-Roboter-Interaktion und unter Berücksichtigung der Besonderheiten einer auf dem Lichtflugzeit-Prinzip basierenden Kamera. Der Ansatz wird robuster, indem die Amplitudenbilder in den Schätzprozess integriert werden. Eine Anpassung der bisherigen Modelle ist deshalb nötig.

Mit dem durch die Hintergrundmodellierung eingelernten Modell ist es schließlich möglich, durch einfache Subtraktion, eine erste Segmentierung des Vordergrundes vorzunehmen. Die Trennung des Vordergrundes vom Hintergrund ist ein erster Schritt, Wichtiges von Unwichtigem zu trennen.

Später angewandte Methoden der Bildfolgenauswertung, wie bspw. differentielle Techniken, reagieren empfindliche auf lokale Störungen von Bildpunkten. Deshalb wird zusätzlich zur Vordergrund-Hintergrund-Segmentierung ein Verfahren integriert, welches die Bildinformationen angemessen regularisiert. Insbesondere muss die Regularisierung in der hier vorliegenden Arbeit adaptiert an die Bildstrukturen stattfinden. Deshalb sind klassische Vorgehensweisen wie bspw. die Filterung mit einer isotropen Gaußfunktion wenig empfehlenswert. Stattdessen wurden lokal-adaptierende Ansätze entwickelt und experimentell untersucht.

Die bis dahin vorverarbeiteten Bilder werden einem sich anschließenden Modul zur Verfügung gestellt, dessen Aufgabe darin besteht, Merkmale zu extrahieren (Abbildung 3.2(4)). Diese Merkmale können Ballungen von Bildpunkten, oder aber auch einzelne Bildpunkte sein. Die extrahierten Merkmale dienen später der Schätzung ausgewählter kinematischer Parameter eines Menschmodells, oder dienen der noch später folgenden Risikoquantifizierung. Im Wesentlichen werden Skelettierungen, Einzelpunktextraktionen und andere wichtige Merkmale,

z.B. Orientierungsmerkmale, gebildet.

Ein Beispiel dieser Vorgehensweise ist in Abbildung 3.3 zu sehen. Hier wurde die Person frontal abgebildet. Die verwendeten Algorithmen werden im weiteren Verlauf der Arbeit vertieft dargestellt und diskutiert.

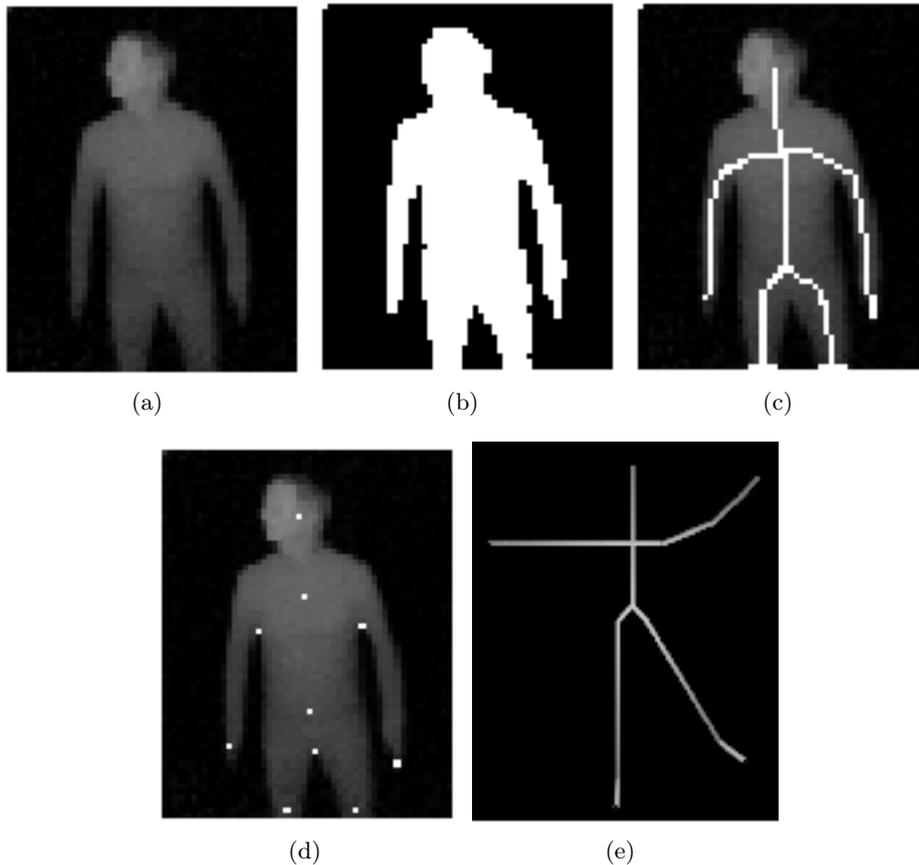


Abbildung 3.3: Illustration zur Berechnung ausgewählter kinematischer Parameter, basierend auf Tiefendaten zur markerlosen Rekonstruktion einer vereinfachten menschlichen Kinematik.

Die Abbildung 3.3(a) zeigt ein vorverarbeitetes Tiefenbild mit den zuvor beschriebenen Methoden. In 3.3(b) wird das Resultat der Segmentierung und Binarisierung dargestellt. In 3.3(c) wird auf der Grundlage, von später erläuterten Algorithmen, eine Skelettierung der Tiefendaten vorgenommen. In 3.3(d) werden aus signifikanten Merkmalen, z.B. Endpunkten, Kreuzungspunkten und geometrisch ermittelten Punkten, Merkmale extrahiert. In 3.3(e) werden diese Merkmale dazu verwendet ein stark vereinfachtes kinematisches Menschmodell zu rekonstruieren, welches sich auf wichtige kinematische Parameter für die Mensch-Roboter-Kooperation konzentriert.

Schließlich erhält man die folgenden Merkmalsgruppen:

- Positionsschätzungen.
- Bewegungsschätzungen.

- Orientierungsschätzungen.
- Trajektorien.

Da die so gewonnenen Merkmale teilweise aus den Bildfolgen gewonnen werden, handelt es sich hier um Bildfolgeauswertetechniken (siehe Abbildung 3.2(6a, 6b, 6c)). Die Merkmale und Muster werden weiter verarbeitet durch die Berechnungen des Optischen Flusses (Abbildung 3.2(6a)), Anwendung eines adaptierten Kalman-Filters (Abbildung 3.2(6b)) und der Kalman-Prädiktion (Abbildung 3.2(6c)).

Da die eingesetzten Module auf Ausreißer angemessen reagieren können, erhält man robuste Merkmale. Insbesondere im Hinblick auf eine sichere Mensch-Roboter-Kooperation stellt diese Vorgehensweise eine neue Qualität bei der Realisierung des Systems dar.

Die robusten Merkmale werden dann an ein Modul mit der Bezeichnung „Wissensbasiertes System“ (Abbildung 3.2(7)) weitergereicht. Dahinter verbirgt sich eine Entscheidungslogik, welche auf der Grundlage der robusten Merkmale bestimmt, ob das Merkmal dazu beiträgt, ein Menschmodell aufzubauen (Abbildung 3.2(8a)), d.h. Informationen zur menschlichen Kinematik liefert, oder es sich um Szeneobjekte (Abbildung 3.2(8b)) handelt. Die Position des Roboters steht nach dem Kalibrierprozess ebenfalls zur Verfügung. Falls der Roboter in den Szenebereich der Kamera fährt, können die dadurch belegten Bildpunkte dem Roboter zugeordnet werden. Da die durch den Roboter erzeugten Bildinformationen für die Erkennung des Menschen störend sind, können diese deshalb verworfen werden. Diese Vorgehensweise hat darüberhinaus den Vorteil, dass man unterscheiden kann, ob eine Verdeckung des Menschen durch den Roboter stattfindet. Unter gewissen Annahmen und in speziellen Fällen, kann man das Menschmodell rekonstruieren. Eine wichtige Eigenschaft, insbesondere für die Aufgaben, welche einen direkten Kontakt zwischen Mensch und Roboter erfordern.

Das Software-System MAROCO realisiert eine Systemarchitektur, welche ein flexibles und komplexes kinematisches Modell für einen menschlichen Körper bereitstellt (Abbildung 3.2(8a)). Da im Rahmen der Mensch-Roboter-Kooperation ein einzelnes 3D-Kamerasystem eingesetzt wird, welches an der Decke des Demonstrators montiert ist, können aufgrund von Abschattungen lediglich eine bestimmte Auswahl an Parametern (>12) geschätzt werden (siehe Abbildung 3.4).

Wichtige kinematische Parameter, die berücksichtigt werden sind:

- Die Kopforientierung (Blickrichtung).
- Die Orientierung der Schulter.
- Die Konfiguration der Ober- und Unterarme.
- Die Körpergröße.
- Die Position.

Die Bildpunktballungen, die keiner Quelle zugeordnet werden können, werden durch umhüllende geometrische Körper (z.B. Zylinder) erfasst und in Abhängigkeit der initial geschätzten Körpergröße des Aktors angepasst.

Damit stehen dem Softwaresystem MAROCO die geometrischen Beschreibungen des Roboters, des Menschen und der Szeneobjekte zur Verfügung. Es handelt sich also um einen einfachen Klassifikator, welcher zwischen den drei Kategorien unterscheidet.

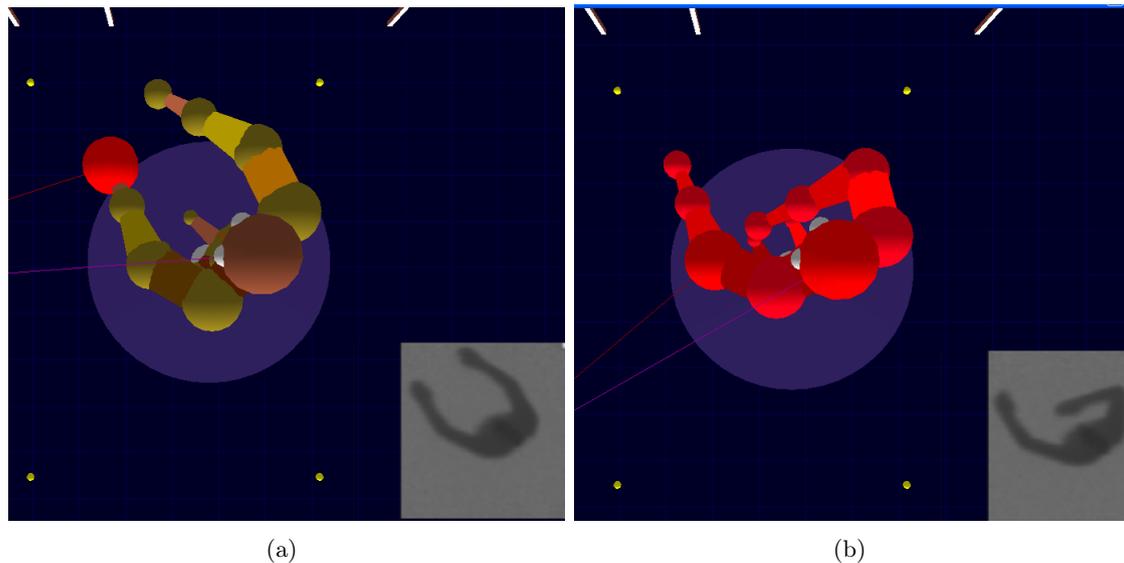


Abbildung 3.4: Zwei unterschiedliche Aufnahmen einer Person mit der Tiefenbildkamera und ihre Rekonstruktion in das virtuelle Umgebungsmodell von MAROCO.

Im nächsten Schritt werden alle Objekte in ein geometrisches Szenemodell abgebildet. Das Szenemodell hat diverse Vorteile. Man kann sich während der Laufzeit die Geschehnisse der realen Szene, virtuell abgebildet in MAROCO, betrachten. Es besteht die Möglichkeit verschiedenartige Szenarien einzuspielen, welche die Sicherheitsgewährleistung beeinflussen, was eine intuitive Diskussion fördert. Außerdem wird zusätzlich zum Szenemodell eine Schnittstelle implementiert, welche es ermöglicht, die Steuerung der Roboter unabhängig vom physikalischen Modell zu verwenden. Damit kann man gezielt kritische Situationen durchspielen, um intuitiv sicherheitskritische Parameter zu bestimmen. Dies ist ein wichtiger Vorteil für den Einsatz von Methoden des überwachten Lernens. An dieser Stelle sei auch darauf hingewiesen, dass das finale System in Echtzeit, d.h. gemessen am Takt des Kamerasystems läuft. Ein System, basierend auf einer reinen Bildfolgenverarbeitung, wird zum gegenwärtigen Zeitpunkt kaum Chancen haben, dieser Anforderung gerecht zu werden, insbesondere wenn man berücksichtigt, dass sich die Szeneanalyse bei einer derartigen Vorgehensweise schwierig gestaltet.

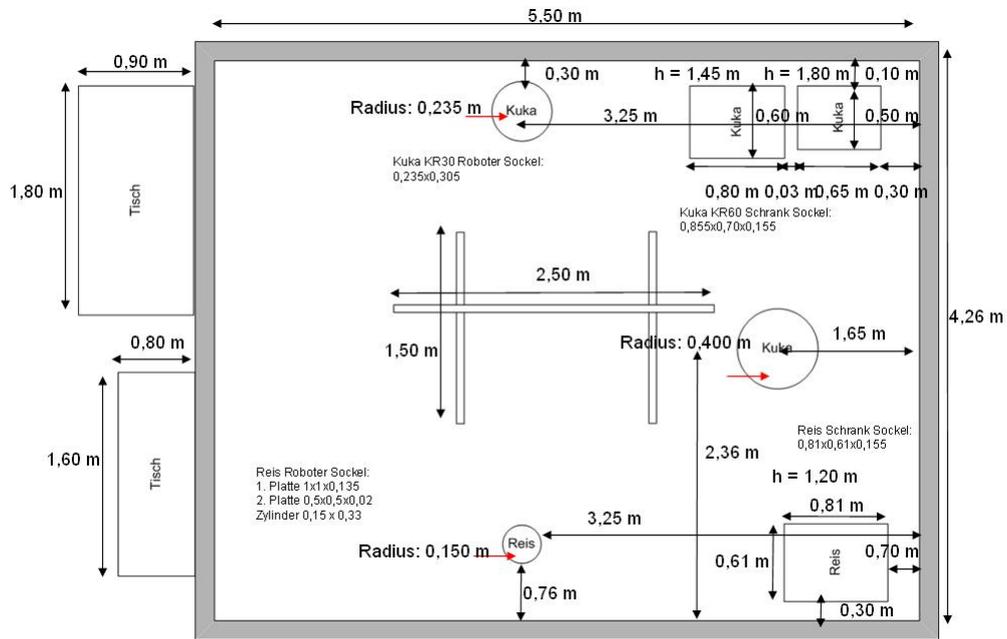
Die robusten Merkmale und die effiziente Abstandsberechnungen zwischen den in der Szene befindlichen Entitäten, bilden die Grundlage für die Risikoquantifizierung. In MAROCO werden mehrere Möglichkeiten für die Risikoquantifizierung untersucht. Einerseits werden Methoden des Maschinellen Lernens eingesetzt (Abbildung 3.2(12a)), andererseits stehen Implementierungen zur Verfügung, welche als Modul zusammengefasst die Bezeichnung „Funktionelle Bewertung des Risikos“ tragen (Abbildung 3.2(12b)).

Auf der Grundlage der Berechnung des (geschätzten) Risikos können dann Entscheidungen getroffen werden (Abbildung 3.2(14)), die das Risiko für den Menschen minimieren. Eine einfache Möglichkeit ist, die Arbeitsgeschwindigkeit des Roboters zu reduzieren. Eine aufwendigere Variante berücksichtigt die Neuplanung einer Bahn (Abbildung 3.2(15)). Auch für diesen Fall erweist sich die virtuelle Szenemodellierung von MAROCO als Vorteil. Durch MAROCO ist es möglich, die Bahn des Roboters theoretische in jedem Zeitschritt neu zu planen. Beide Varianten, die Reduktion der maximalen Arbeitsgeschwindigkeit und die Umplanung einer

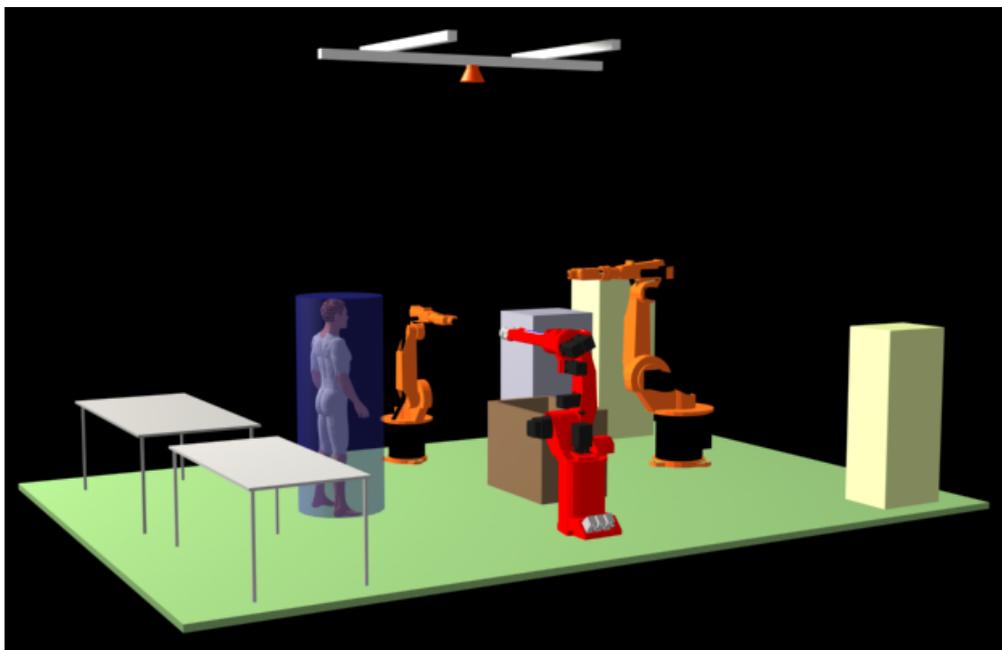
kritischen Bahn, beeinflussen die Kinematik des Roboters (Abbildung 3.2(15)).

Bei der Realisierung der zweiten Architektur wurde zusätzlich die Demonstratorzelle erweitert. Hinzugekommen ist ein KUKA KR60HA Industrieroboter. Zwei unterschiedliche Sichten auf den Aufbau der Zelle sind in der Abbildungen 3.5 zu finden.

Schließlich ist es möglich, in Echtzeit, d.h. im Takt der Kamera, ein virtuelles Umgebungsmodell aufzubauen (Abbildung 3.6).



(a)



(b)

Abbildung 3.5: (a) Vermessung der erweiterten Demonstratorzelle. (b) Visualisierung des zweiten Demonstratoraufbaus.

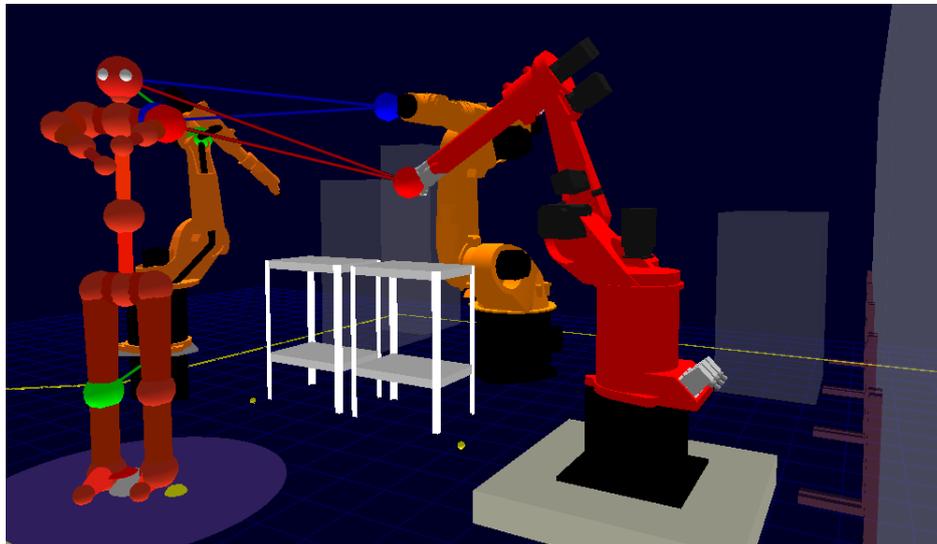


Abbildung 3.6: Darstellung einer komplexen rekonstruierten Szene durch MAROCO.

Kapitel 4

Bildfolgenauswertung

Dieses Kapitel beginnt mit der Beschreibung der Vorverarbeitung von Tiefenbildern. Da es sich um verrauschte Tiefenbilder handelt, müssen die Aufnahmen angemessen regularisiert werden. Dazu wird im ersten Abschnitt ein neuartiges Verfahren vorgestellt, welches effizient implementiert wurde und zu hervorragenden Ergebnissen, im Vergleich zu verschiedenen klassischen Ansätzen, geführt hat.

Die Bildpunkte der regularisierten Bilddaten müssen nun klassifiziert werden nach bedeutungsvollen Bildpunkten zugehörig zum Vorder- oder Hintergrund. Ein beliebter Ansatz für CCD-Kameras sind die in [Stauffer 2002, Lee 2005] vorgestellten Verfahren. Da im Rahmen dieser Arbeit mit einem tiefauflösenden Sensor gearbeitet wird, ist eine Adaption der Verfahren nötig um bessere Segmentierungsergebnisse zu erzielen. Dazu wird ein neuer Bewertungsformalismus eingeführt, welcher neben den Tiefeninformationen auch die Amplitudeninformationen des Sensors mitberücksichtigt.

Anschließend wird in Abschnitt 4.3.2 das theoretische Modell zur Erzeugung der Merkmale für die erste Architektur vorgestellt und diskutiert. Dazu wird ein hochgenaues Optisches Fluss Schätzverfahren in das Rahmenwerk eines Variationsansatzes zur Bestimmung von Randkurven eingeführt. Zusätzlich wird eine vollautomatische Initialisierung diskutiert. Die Bestimmung der Optischen Flussfelder wird um einen erheblichen Faktor gesteigert, indem die komplizierten punktweisen Löser in ein kompaktes Gleichungssystem umformuliert werden, wodurch effizientere numerische Methoden zur Lösung großer, schwach besetzter Gleichungssysteme angewandt werden können. Die Optischen Flussfelder werden auf die Bilder nach der Vorverarbeitung und nach der Hintergrundmodellierung angewandt.

Für die zweite Systemarchitektur wird ein neues Optisches Fluss-Modell eingeführt, welches als Verallgemeinerung der kombiniert lokalen-globalen Schätzverfahren (siehe bspw. [Bruhn u.a. 2005a]) angesehen werden kann. Die Idee liegt darin, dass der aus der lokalen Konstanzannahme resultierende Datenterm durch die eingeführte Fensterfunktion (normalerweise eine Gaußsche-Funktion) einen isotropen Charakter hat. Dies führt zur systematischen Unterschätzung von Flussfeldern. Im Rahmen dieser Arbeit wurde ein adaptives Verfahren entwickelt um anisotrope Datenterme zu generieren und um damit eine genauere Schätzung der Betragsnorm von Optischen Fluss-Feldern zu ermöglichen. Das verallgemeinerte Optische Fluss-Modell wird anschließend durch einen Mehrgitteransatz gelöst, wobei die Gleichungen auf Grafikkartenprozessoren gelöst werden, was eine vielfache Verkürzung der Berechnungszeit bringt. Abschließend wird ein 3D Optisches Flussfeld erzeugt, indem die Geometrie der Kamera miteinbezogen wird. Die Optischen Flussfelder sind wichtige Merkmale für die Bewegungs-

schätzung von Personen in der Roboterzelle und werden später bei der Kinematikschtzung benötigt.

Alle vorgestellten Methoden werden experimentell verglichen und evaluiert.

4.1 Vorverarbeitung

Bevor mit der eigentlichen Auswertung von Bildfolgen begonnen werden kann, müssen durch vorverarbeitende Schritte, die in den Bildaufzeichnungen enthaltenen Störungen minimiert werden, um die Bildsignale für nachfolgende Verarbeitungsschritte aufzubereiten.

Die in den folgenden Abschnitten dargestellten Methoden sind für diesen Zweck entwickelt. Da diese Werkzeuge universell für verschiedene Typen von Bildern einsetzbar sind, werden einige der experimentellen Ergebnisse anhand von medizinischen Bilddaten diskutiert, da diese aufgrund ihrer vielseitigen Strukturierung eine kompakte Diskussion ermöglichen.

Die klassische Vorgehensweise zur Glättung von Bilddaten geschieht durch Faltung mit einer gaußähnlichen Funktion. Gerechtfertigt wird diese Vorgehensweise durch die lineare Diffusionsgleichung

$$\frac{\partial u}{\partial t} = \Delta u \quad . \quad (4.1)$$

Dabei wird ein Bild durch die Funktion u beschrieben. Die Variable t steht für die Zeit. Mit $u(0, x) = u_0(x)$ und $u_0 \in L_2(\mathbb{R}^n)$ existiert die Lösung

$$u(t, x) = \begin{cases} u_0(x), & t = 0, \\ \left(u_0 * \frac{1}{(4\pi t)^{\frac{n}{2}}} \exp\left(-\frac{x^2}{4t}\right) \right) (x), & t > 0. \end{cases} \quad (4.2)$$

Die Herleitung gelingt im Wesentlichen durch die Anwendung der Fouriertransformation und anschließender partieller Integration. Damit reduziert sich die partielle Differentialgleichung auf eine gewöhnliche Differentialgleichung. Durch Rücktransformation der Lösung der Gleichung erhält man schließlich den sogenannten Gauß-Kern.

Die Faltung des Bildes mit dem Gauß-Kern führt zu einem Diffusionsvorgang. Die Diffusion wird durch den Konzentrationsgradienten ∇u in dem Medium u verursacht. Es entsteht durch den Gradienten eine Strömung $\sigma = -D\nabla u$, welche dem Gradienten entgegengerichtet ist. D wird als Diffusionstensor bezeichnet. Der Tensor reduziert sich auf eine skalare Größe, falls $\nabla u \parallel \sigma$ gilt. In diesem Fall spricht man von isotroper Diffusion.

Eine in der Praxis übliche Vorgehensweise modifiziert die Lösung der isotropen Diffusion, indem mit einer endlichen Matrix, abgeleitet aus der Abtastung einer Gaußfunktion, gefaltet wird. Diese Vorgehensweise ist heuristisch motiviert und kann in den meisten Standardlehrbüchern zur digitalen Bildverarbeitung nachgelesen werden (z.B. [Jähne 2006] oder [Gonzalez und Woods 2002]).

Die bisher vorgestellte Methode zur Glättung von Bilddaten ist isotrop. Das bedeutet, dass unabhängig von der Richtung gleichermaßen geglättet wird. Durch Anwendung solcher Filter verschmieren signifikante Bildinhalte, wie bspw. Kanten. Die differentiellen Methoden, z.B. die Bewegungsschtzung durch den Optischen Fluss, aber auch Techniken zur Segmentierung oder Klassifikation, benötigen oftmals diese Informationen, weshalb Filter angewendet werden müssen, die diese speziellen Strukturen erhalten.

In der Vergangenheit wurden eine Vielzahl von Arbeiten zur strukturerhaltenden Glättung publiziert. Einen wichtigen Anstoß haben diese Methoden durch die Arbeiten von [Rosenfeld und Thurston 1971] und [Wiktin 1983] erhalten.

Eine aus heutiger Sicht als Klassiker zu bezeichnende Arbeit ist [Perona und Malik 1990]. Ein Vorteil dieser Methode liegt darin, dass der Diffusionsfilter in Abhängigkeit lokaler Muster adaptiert wird (siehe Abbildungen 4.1 und 4.2). Es ist möglich, den Kontrast von Kanten in Bildern zu verstärken, bei gleichzeitiger Reduktion der Bildstörungen. Außerdem erfüllt dieser Filter die Lokalitätsforderung, d.h. Kanten bleiben bei Anwendung der Diffusion an ihrer ursprünglichen Position. In den nun folgenden Abschnitten sollen drei ausgewählte Techniken vorgestellt werden, welche durch ihre Wirkungsweise als anisotrope Filter zu bezeichnen sind.

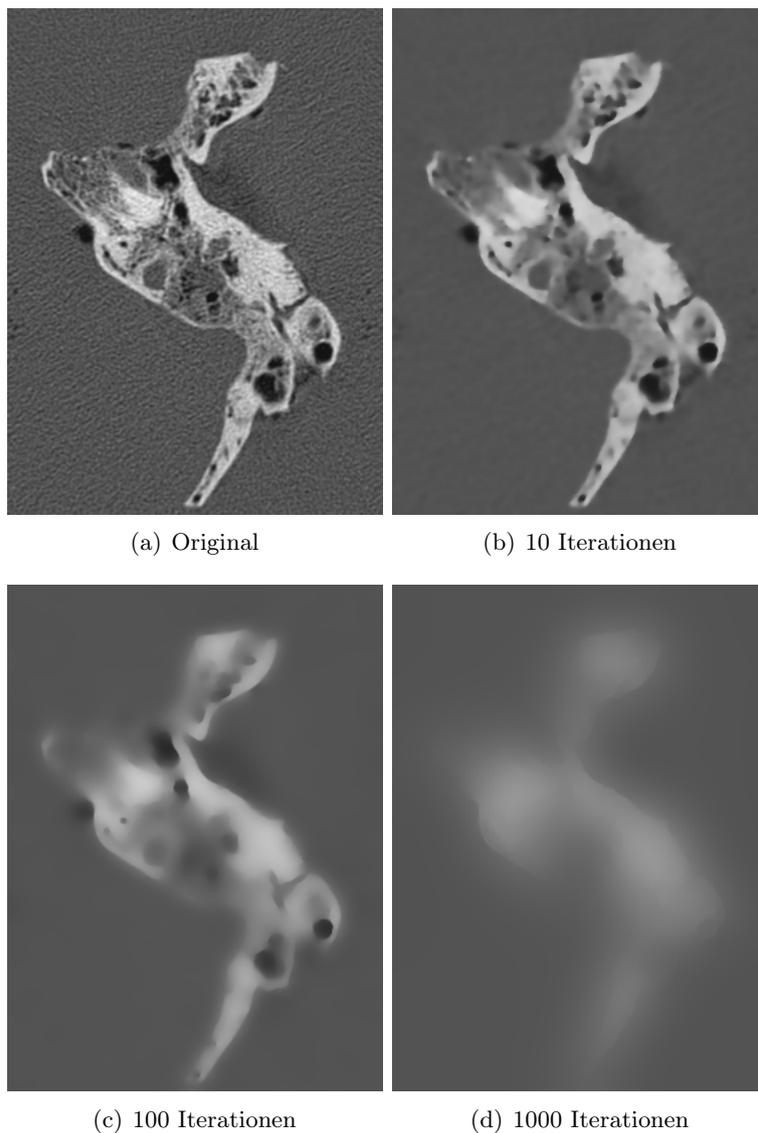


Abbildung 4.1: Abbildungen für das lineare Modell nach [Perona und Malik 1990].

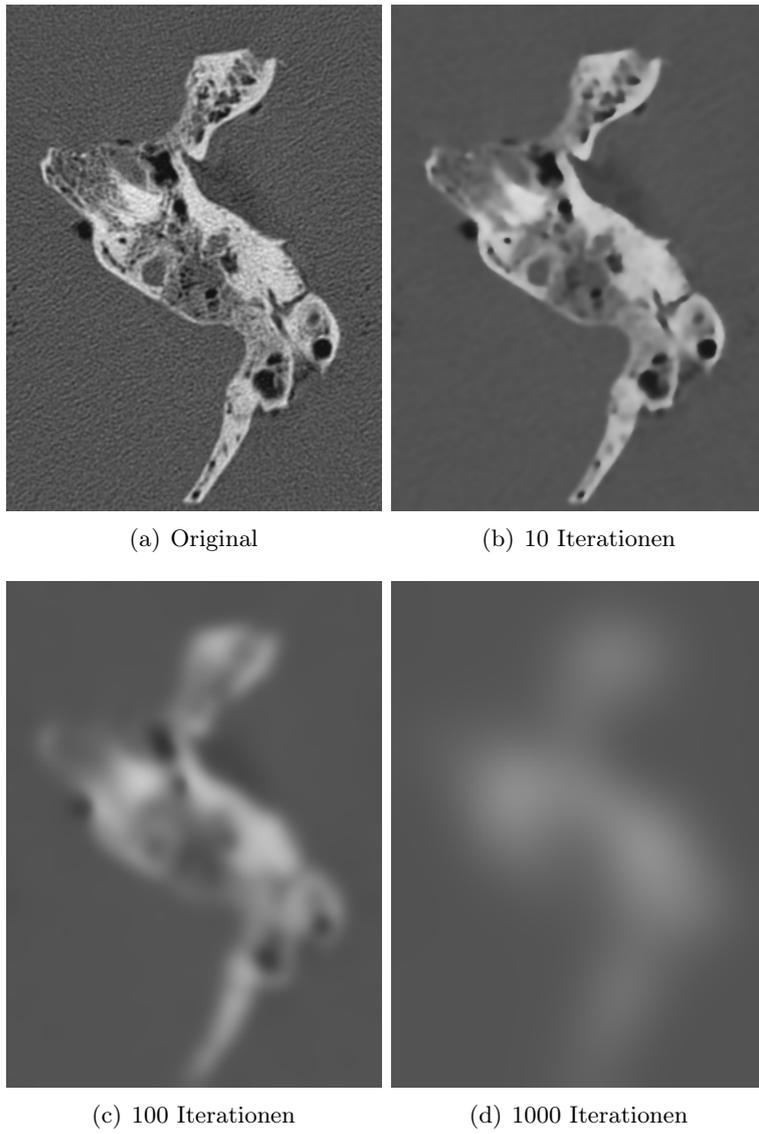


Abbildung 4.2: Abbildungen für das anisotrope Modell nach [Perona und Malik 1990].

4.1.1 Anisotrope Diffusion

In diesem Abschnitt werden die Grundlagen für die lineare und nichtlineare anisotrope Diffusion vorgestellt. Die Diffusion kann man sich veranschaulichen, indem die Grauwerte der Bilder als Konzentration betrachtet werden. Es lässt sich dazu die parabolische partielle Differentialgleichung formulieren:

$$\partial_t c(x) = -\operatorname{div} F. \quad (4.3)$$

Durch Einführung des Masseerhaltungsgesetzes in das Ficksche Gesetz erhält man:

$$F = -D\nabla c(x). \quad (4.4)$$

Damit erhält man die gewünschte Diffusionsgleichung:

$$\partial_t c(x) = \operatorname{div}(D\nabla c(x)). \quad (4.5)$$

Die Konzentration wird hier durch die Funktion c beschrieben, D kann ein Skalar oder eine Matrix sein. Ist D eine Matrix, bezeichnet man diese als Diffusionstensor. Durch ihn lassen sich verschiedenartige Diffusionsfilter realisieren. Im Wesentlichen stellt er einen Ausgleich zwischen der Strömung bzw. dem Fluss F und dem Gradienten der Konzentration her. Er ist also ein besonderer Freiheitsgrad für den Filter. Eine tiefergehende Untersuchung und Diskussion findet man in [Weickert 1998].

Die Abhängigkeit der gegenwärtigen Diffusion von dem Resultat der zeitlich vorhergehenden Diffusion führen auf eine Nichtlinearität. Deshalb muss in jedem Zeitschritt eine nichtlineare partielle Differentialgleichung gelöst werden. Der entscheidende Unterschied zur Filterung mit einem Gauß-Kern liegt in der Anisotropie, welche durch die Bilddaten selbst geregelt wird, indem entlang von Kanten, oder über homogenen Flächen hinweg die Diffusion verstärkt und orthogonal zu Kanten die Diffusion minimiert.

In [Weickert 1999] wird eine Vorgehensweise vorgestellt, welche durch Auswertung des Strukturtenors, Einfluss auf den Diffusionsvorgang nimmt. Die Gleichung (4.5) hat einen Freiheitsgrad, der durch Einführung eines Diffusionstensors festgelegt werden kann. In [Weickert 1999] wird dazu vorgeschlagen, den Strukturtenor

$$J_\rho = K_\rho * (\nabla c(x)\nabla c(x)^T) \quad (4.6)$$

in das Diffusionssystem zu integrieren. Der Parameter ρ stellt dabei ein Größe zur Beeinflussung der Varianz dar. Dadurch kann der Wirkungsbereich des Tensors gesteuert werden. Der entscheidende Punkt liegt darin, dass die Eigenwerte des so entstandenen Diffusionstensors neu bestimmt werden, durch z.B. folgende Berechnung

$$\begin{aligned} \lambda_1^{\text{neu}} &= \max \left[\epsilon, 1 - \exp \left(-\frac{(\lambda_1 - \lambda_2)^2}{\kappa^2} \right) \right] , \\ \lambda_2^{\text{neu}} &= \epsilon . \end{aligned} \quad (4.7)$$

Einen intuitiven Zugang zu dieser Vorgehensweise gewinnt man, indem man sich klar macht, dass der Strukturtenor eine 2×2 -Matrix ist, welche stets symmetrisch und positiv-semidefinit ist. Das bedeutet, dass man eine Normalform definieren kann, indem durch Anwendung einer orthogonalen Transformation die Eigenwerte des Strukturtenors durch eine Diagonalmatrix repräsentiert werden. Formal gilt:

$$J' = U^T J U = U^T \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} U , \quad (4.8)$$

wobei U eine orthogonale Matrix, z.B. $U \in S(2)$, ist. Des Weiteren kann man erzwingen, dass o.B.d.A stets $\lambda_1 \geq \lambda_2$ gilt. Damit lässt sich eine Normalform für den Strukturtensor definieren. Die Spaltenvektoren der orthogonalen Matrix U sind die zu den Eigenwerten des Strukturensors zugehörigen Eigenvektoren. In Gleichung 4.6 wird der Strukturtensor definiert durch Faltung einer Gaußfunktion mit dem äußeren Produkt der Gradienten der Bildintensitäten. Der Strukturtensor kann daher als Maß für die lokale Verteilung der Grauwertvariationen herangezogen werden. Die Eigenwerte verhalten sich proportional zur Grauwert- bzw. Intensitätsvariation innerhalb der lokalen Umgebung um einen Aufpunkt. Die Lokalität lässt sich durch Variation der Kovarianzmatrix, hier ausgedrückt durch ρ , einstellen.

Durch den Strukturtensoransatz erhält man eine analytische Aussage über die lokale Verteilung der Intensitäten bzw. die Konzentration. Der Eigenvektor, korrespondierend zum größeren Eigenwert λ_1 , zeigt damit in Richtung der stärksten Intensitätsänderung. Der Eigenvektor, korrespondierend zum kleineren Eigenwert λ_2 , zeigt in Richtung der geringsten Konzentrationsvariation.

Durch systematisches Zusammenfassen der wichtigsten Kombinationen erhält man Tabelle 4.1, welche sich auf die Grenzfälle konzentriert.

λ_1	λ_2	Strukturierung
0	0	homogen
max	0	geradlinig
> 0	> 0	eckig
max	max	große Variationen

Tabelle 4.1: Kategorisierung von Intensitätsstrukturen, modifiziert nach [Jähne 1993].

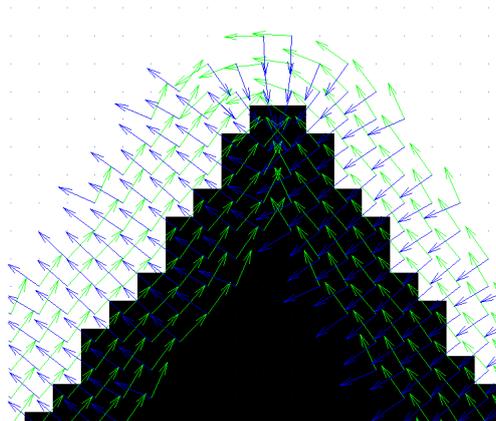


Abbildung 4.3: Visualisierung einer Intensitätsstruktur mit eingezeichneten Eigenvektoren, resultierend aus einer Strukturtensor-Analyse.

Die in Abbildung (4.4) eingezeichneten blauen Vektoren repräsentieren die Eigenvektoren, korrespondierend zu den größten Eigenwerten. Die grünen Vektoren repräsentieren dementsprechend die Eigenvektoren zugehörig zu den kleineren Eigenwerten. Die anisotrope Diffusion soll also entlang der blau eingefärbten Vektoren durchgeführt werden, wohingegen die Diffusion entlang der grün markierten Vektoren gestoppt werden soll. Als mögliches Ergebnis erhält man, das in Abbildung (4.5) gezeigte Resultat.

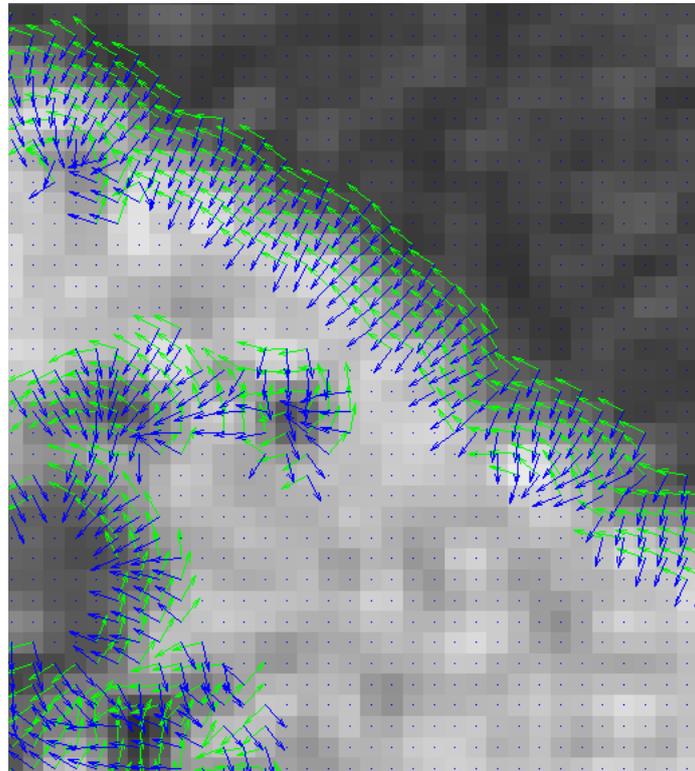


Abbildung 4.4: Visualisierung eines Eigenvektorfeldes mit unterschiedlichen Grauwertvariationen.

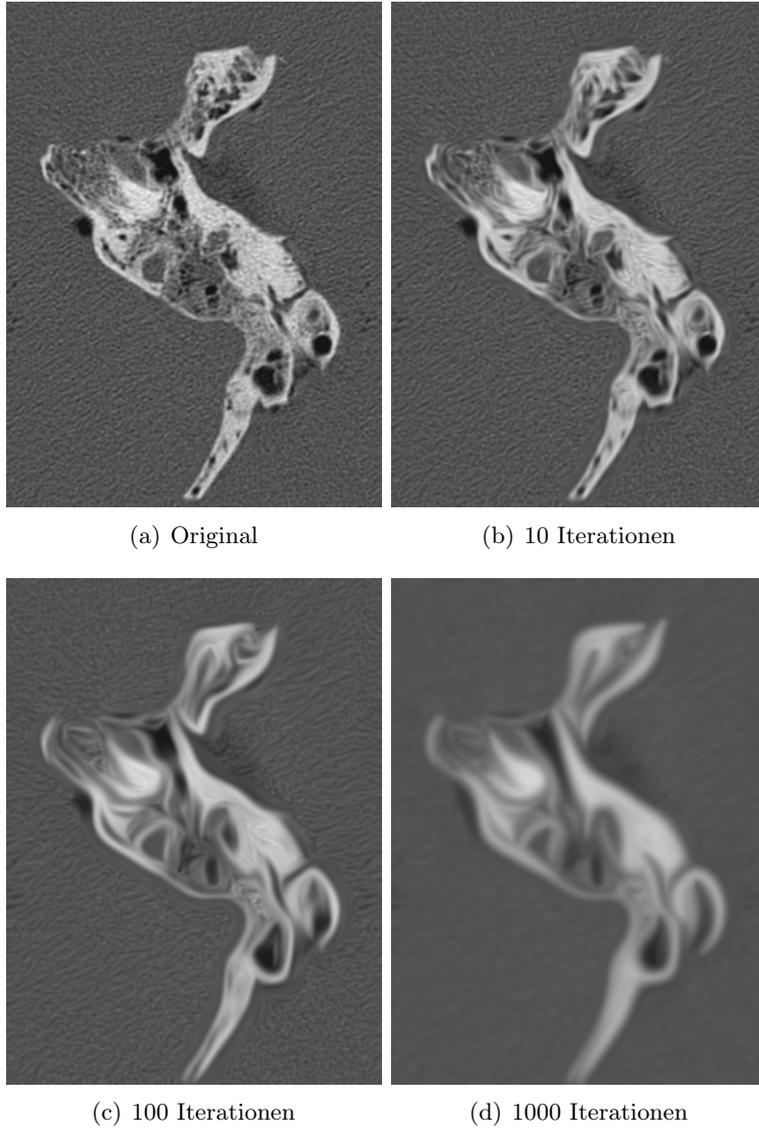


Abbildung 4.5: Verschiedene Iterationsstufen für ein medizinisches Bild mit starken Grauwertvariationen.

Bei genauerer Betrachtung der Bilder ergeben sich neben den Vorteilen, offensichtliche Nachteile. Die anisotrope Diffusion weitet Buchten oder Ecken aus, das Hintergrundrauschen wird ebenfalls nicht hinreichend gut diffundiert und verschiedene Inseln verbinden sich durch den Diffusionsvorgang im Laufe der Zeit und werden so zu nicht mehr trennbaren Bildpunktballungen. Die hier vorgestellten Methoden werden untersucht im Hinblick auf die Tauglichkeit für die Mensch–Roboter–Kooperation. Man stellt fest, dass die hier vorgestellte nichtlineare anisotrope Diffusion, unabhängig von der Güte der Ergebnisse, wegen der Berechnungszeit von mehreren Minuten ungeeignet ist für diese Arbeit.

4.1.2 Effiziente bilaterale Filterung

Das Problem der anisotropen Diffusionsfilterung liegt in der Vielzahl der iterativen Schritte, welche eine echtzeitfähige Anwendung strapazieren.

Die bilaterale Filterung verzichtet auf Iterationen und kann in einem einzigen Schritt angewandt werden. Des Weiteren wurde diese Methode im Rahmen der Dissertation effizient implementiert, so dass eine Anwendung unter Ausnutzung leistungsfähiger Grafikkarten möglich ist. Die im Folgenden vorgestellte Methode kommt damit weit unter die Echtzeitanforderung und überzeugt in diesem wichtigen Punkt. Für die PMD-Kamera konnten Messungen mit unter 1 ms pro Bild erzielt werden. Die in [Durand und Dorsey 2002] vorgeschlagene schnelle Berechnung der bilateralen Filterung, basierend auf einer Linearisierung und Anwendung einer schneller Fouriertransformation, muss deshalb nicht berücksichtigt werden, da die Beschleunigung durch Ungenauigkeiten, entstehend durch Linearisierungsschritte, erzielt wird.

Der bilaterale Filter wird erstmalig in [Tomasi und Manduchi 1998] beschrieben. Er stellt eine nichtlineare Filtermethode dar. Nichtlineare Filter werden für die Bildverarbeitung entwickelt, um signifikante Bildstrukturen bei der Filterung zu erhalten. Nichtlineare Filter gehören deshalb zur Klasse der adaptiven Filter. Der Unterschied zwischen einem linearen und nichtlinearen Filter liegt darin, dass ein nichtlinearer Filter nur von der Bildposition, nicht aber von den Bildwerten abhängt.

Die Familie der bilateralen Filter beruhen auf einer einfachen und anschaulich einleuchtenden Idee. Die Filterung an einem Aufpunkt wird durch eine gewichtete Mittelung erzielt. Die Besonderheit der gewichteten Mittelung liegt darin, dass zwei Einflussfaktoren das Verhalten des Filters bestimmen. Einerseits spielt die geometrische Distanz eine wichtige Rolle, andererseits die photometrische Ähnlichkeit, d.h. der Intensitätsunterschied eines benachbarten Bildpunktes zum Aufpunkt. Damit haben nahegelegene Bildpunkte aus der Nachbarschaft des Aufpunktes, die eine ähnliche Intensität vorweisen, größeren Einfluß auf die Filterung, als Punkte mit größerer Distanz und höherer Differenz bezüglich der Intensität.

In [Tomasi und Manduchi 1998] wird dieser Sachverhalt erstmals formalisiert und stellt damit eine wichtige Alternative zur kostenintensiven, anisotropen Filterung dar. Die theoretischen Grundlagen wurden erst später erarbeitet (siehe [Elad 2002] und [Mrázek et al. 2006]). Die bilateralen Filter können in die Theorie der robusten Schätzer eingebettet werden. Die Herleitung gelingt durch Minimierung eines Energiefunktional.

Es werden beim bilateralen Filter zwei Filtertechniken kombiniert:

- Ein ortsabhängiger Filter und
- ein Bereichsfilter¹.

¹Die Autoren von [Tomasi und Manduchi 1998] sprechen von domain filter und range filter.

Die beiden Typen von Filtern werden in den nächsten Unterabschnitten vorgestellt. Die Notation wurde modifiziert, ist aber gleichbedeutend zu der in [Tomasi und Manduchi 1998] eingeführten Darstellung.

Ortsfilterung

Die ortsabhängige Filterung lässt sich durch

$$f_o(\mathbf{x}) = \frac{1}{\eta_o} \int_{\Omega} c(\mathbf{x}, \mathbf{x}') \cdot f(\mathbf{x}') d\mathbf{x}' \quad (4.9)$$

formalisieren. Mit einer Normalisierungskonstante η_o , für die gilt:

$$\eta_o = \int_{\Omega} c(\mathbf{x}, \mathbf{x}') d\mathbf{x}' \quad . \quad (4.10)$$

Durch Ω wird der gesamte Bildbereich erfasst.

Bei der Berechnung des Bildpunktes $f_o(\mathbf{x})$ werden die benachbarten Bildpunkte durch $c(\mathbf{x}, \mathbf{x}')$ gewichtet und zusammengefasst. Üblicherweise, und auch im Rahmen der hier vorliegenden Arbeit, wird für $c(\mathbf{x}, \mathbf{x}')$ die Gauß-Funktion

$$c(\mathbf{x}, \mathbf{x}') = \frac{1}{2\pi|\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\mathbf{x} - \mu)^T \Sigma^{-1}(\mathbf{x} - \mu)\right) \quad (4.11)$$

eingesetzt.

Hierbei ist zu beachten, dass die bisherige Filterung stetig definiert ist. Da bei der späteren Realisierung die bisherigen Gleichungen diskretisiert werden, müssen Besonderheiten bei der technischen Realisierung beachtet werden, auf die nicht weiter eingegangen werden soll. Insbesondere die Behandlung des über den Randbereich der Bilddimension hinausgehende Gauß-Filter muss angepasst werden. In der Regel erlaubt man sich bei der Implementierung eine Ungenauigkeit, in dem die Maskengröße in Abhängigkeit der Varianz formuliert wird. So wird bspw. das 3-fache der Standardabweichung gewählt, um die Anzahl der Abtastpunkte der kontinuierlichen Gaußfunktion zu bestimmen. Damit die Eigenschaften für eine Mittelung erfüllt bleiben, muss anschließend re-normalisiert werden, indem über die Funktionswerte der Abtastpunkte summiert und anschließend geteilt wird.

Bereichsfilterung

Die Anwendung der Ortsfilterung ist in dem hier vorliegenden Spezialfall eine Faltung mit einer Gauß-Funktion. Dadurch finden an Übergängen wie bspw. Kanten zunächst wieder die unerwünschten Artefakte der isotropen Filterung statt. Durch Integration der Bereichsfilterung kann dieses Problem gelöst werden.

Deshalb wird der Bereichsfilter definiert zu

$$f_b(\mathbf{x}) = \frac{1}{\eta_b} \int_{\Omega} s(f(\mathbf{x}), f(\mathbf{x}')) \cdot f(\mathbf{x}') d\mathbf{x}' \quad . \quad (4.12)$$

Der Unterschied zwischen dem Ortsbereichsfilter (4.11) und dem Bereichsfilter (4.12) liegt in den Gewichtungsfunktionen $c(\cdot, \cdot)$ und $s(\cdot, \cdot)$.

Die Gewichtung des Bereichsfilters basiert auf einer Ähnlichkeitsfunktion. Die Ähnlichkeitsfunktion ihrerseits wird durch Funktionswerte, d.h. die Intensitätswerte der Bildpunkte definiert. Auch hier verhält es sich so, dass der Filter für große Unterschiede der Argumente entsprechend reagieren muss, in diesem Fall einen geringen Wert nahe Null zurück liefert. Deshalb kann auch für die Ähnlichkeitsfunktion s eine Gaußfunktion eingesetzt werden.

Diskretisierung und Diskussion

Die bisher dargestellten Gleichungen beschreiben einen kontinuierlichen bilateralen Filter. Es ist daher zwingend notwendig, eine Diskretisierung der Gleichungen vorzunehmen:

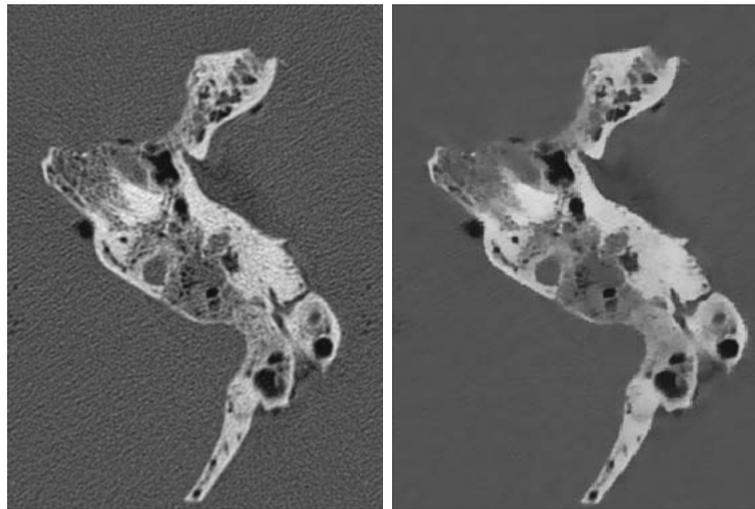
$$\begin{aligned}
f_{i,j} &= \frac{\sum_{(k,l) \in N(i,j)} c((k,l), (i,j)) \cdot s(u_{k,l}, u_{i,j}) \cdot u_{i,j}}{\sum_{(k,l) \in N(i,j)} c((k,l), (i,j)) \cdot s(u_{k,l}, u_{i,j})} \\
&= \frac{\sum_{k \in N_x} \sum_{l \in N_y} c((k,l), (i,j)) \cdot s(u_{k,l}, u_{i,j}) \cdot u_{i,j}}{\sum_{(k,l) \in N(i,j)} c((k,l), (i,j)) \cdot s(u_{k,l}, u_{i,j})} \tag{4.13} \\
&= \frac{\sum_{k=-G_x}^{G_x} \sum_{l=-G_y}^{G_y} c((i-k, j-l), (i,j)) \cdot s(u_{k,l}, u_{i-k, j-l}) \cdot u_{i,j}}{\sum_{(k,l) \in N(i,j)} c((k,l), (i,j)) \cdot s(u_{k,l}, u_{i,j})} \\
&= \frac{\sum_{k=-G_x}^{G_x} \sum_{l=-G_y}^{G_y} \exp\left(-\frac{k^2+l^2}{2\sigma_d^2}\right) \cdot \exp\left(-\frac{(u_{i-k, j-l})^2}{2\sigma_r^2}\right)}{\sum_{(k,l) \in N(i,j)} \exp\left(-\frac{k^2+l^2}{2\sigma_d^2}\right) \cdot \exp\left(-\frac{(u_{i-k, j-l})^2}{2\sigma_r^2}\right)}.
\end{aligned}$$

Die in Gleichung (4.13) durchgeführte Diskretisierung kann für eine abschließende Diskussion zur Funktionsweise des bilateralen Filters herangezogen werden. Ohne Berücksichtigung des photometrischen Terms, hier bezeichnet durch $s(\cdot, \cdot)$, wird der bilaterale Filter ein einfacher, linearer Gaußscher Filter. Die Anisotropie wird deshalb vollständig durch den photometrischen Teil beeinflusst.

Dazu eine Diskussion zweier idealer Beispiele:

1. Auf einer homogenen Bildregion wird das Argument des photometrischen Terms konstant zu null. Dadurch wird $s(\cdot, \cdot)$ konstant zu eins. Betrachtet man $s(\cdot, \cdot)$ als Gewicht für $c(\cdot, \cdot)$, wird sich das Gewicht nicht auf die Ortsfilterung auswirken. Es kommt zu einer einfachen linearen Filterung durch den Ortsbereichfilter.
2. Anders verhält es sich an einer Kante. Auf der Kante kommt es zu einem großen Intensitätsunterschied bzgl. der Bereichswerte. Dadurch wird das Argument der Exponentialfunktion in $s(\cdot, \cdot)$ aufgrund der großen Differenz einen hohen Wert erzeugen. Die Funktion selbst verhält sich reziprok zur Differenz der Argumente und erzeugt deshalb ein geringes Gewicht $s(\cdot, \cdot)$. Die Folge ist eine geringere Gewichtung von $c(\cdot, \cdot)$.

In Abbildung 4.6 ist ein Ergebnis der bilateralen Filterung dargestellt.



(a) Original.

(b) Nach Filterung.

Abbildung 4.6: Resultate der bilateralen Filterung.

4.1.3 Orientierte Medianfilterung

Die bisher vorgestellten Ansätze zur Filterung von Bildern sind Vertreter der nichtlinearen Filter. Die meisten Erweiterungen konzentrieren sich auf eine effiziente Realisierung oder für eine Erweiterung des Filters hin zur mehrkanaligen Bildverarbeitung.

Eine klassische nichtlineare Filtermethode, welche zudem positive Eigenschaften bzgl. der Kantenhaltung besitzt, ist der sogenannte Medianfilter. Der Medianfilter besitzt ebenfalls günstige Eigenschaften, die eine effiziente Implementierung ermöglichen (siehe dazu [Huang 1981] bzw. [Weiss 2006]).

Allerdings ergeben sich bei Anwendung des klassischen Medianfilters unerwünschte Effekte. Die im Rahmen dieser Arbeit entwickelte Idee basiert auf der Überlegung, die Information einer vorgeschalteten Strukturanalyse, basierend auf dem Strukturtensor zu verwenden, um damit in unterschiedlich strukturierten Umgebungen spezielle, auf dem Medianfilter basierende Faltungsmasken anzuwenden.

Analog zur anisotropen Filterung, basierend auf der Diffusionsgleichung, soll hier der Eigenvektor zum kleinsten Eigenwert die Hauptrichtung der Medianfilterung bestimmen.

Anders verhält sich der Filter in homogenen Regionen, in denen aufgrund der Analyse entschieden werden kann, den klassischen Medianfilter zu verwenden. Die hier skizzierte Idee wurde bereits in [Graf und Wörn 2007] vorgestellt und diskutiert. Der orientierte Medianfilter kann damit zur Klasse der adaptierten Filter gezählt werden und verhält sich an signifikanten Bildstrukturen, wie bspw. Kanten, anisotrop.

Die Kombination von Strukturanalyse und Medianfilterung, um dadurch einen Filter zu erhalten, der angemessen an die zugrundeliegende Bildstruktur adaptiert, geschieht in zwei Schritten:

1. Zuerst muss der Strukturtensor für jeden Bildpunkt ermittelt werden. Danach wird der Eigenvektor zum Größeren der beiden Eigenwerte bestimmt. Der sich qualifizierende Eigenwert muss allerdings eine empirisch ermittelte, d.h. an die Bilddaten angepasste Schwelle überschreiten.

- In Abhängigkeit des Eigenwertes, wird auf eine eindimensionale Maskengröße für den Medianfilter zugegriffen. Diese Maske wird dann an einem Aufpunkt entlang des Eigenvektors, zugehörig zum kleinsten Eigenwert, orientiert. Falls der Größere der beiden Eigenwerte unter einer zuvor ermittelten Schwelle liegt, wird der klassische 2D Medianfilter mit ungerader Maskengröße angewandt.

4.1.4 Experimente und Diskussion

In Abbildung 4.7 wird das Ergebnis der Medianfilterung den Resultaten der bereits vorgestellten Filter gegenübergestellt.

Der Medianfilter basiert auf einer Sortierung der Intensitäten, welche aus der Nachbarschaft eines Aufpunktes entnommen werden. Der Wert an der mittleren Position wird anschließend an die Position des Aufpunktes zurückgeschrieben. Durch die Sortierung ist die eine Hälfte der Werte größer als der Median und die anderen Hälfte der Werte kleiner als der Median. Diese Vorgehensweise stattet den Medianfilter mit wichtigen Vorteilen aus. Trotzdem kommt es bei der Anwendung größerer Masken zu Auslöschungseffekten (z.B. der rote eingekreiste Bereich in Abbildung 4.8).

Die folgende Tabelle fasst, die in Abbildung 4.8 ausgewählten Resultate, qualitativ zusammen.

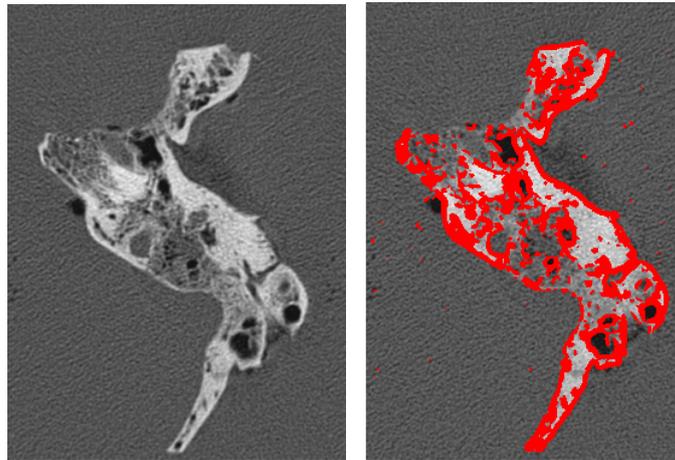
Filtertyp	SV1 rot	SV2 türkis	KA grün	HG blau	EK gelb	ZA
Isotroper Gaußscher Filter	0	-	-	+	-	++
Anisotrope Filterung (PM)	-	-	+	0	0	-/+
Anisotrope Diffusion	-	-	+	0	-	-/0
Bilaterale Filterung	++	++	++	0	++	++
Medianfilter	++	0	++	++	+	++
Orientierte Medianfilterung	+	++	++	++	++	++

Tabelle 4.2: Zusammenfassender, qualitativer Vergleich ausgewählter Filtermethoden.

Die in der Tabelle 4.2 genannten Bezeichnungen für die bewerteten Merkmale stehen für

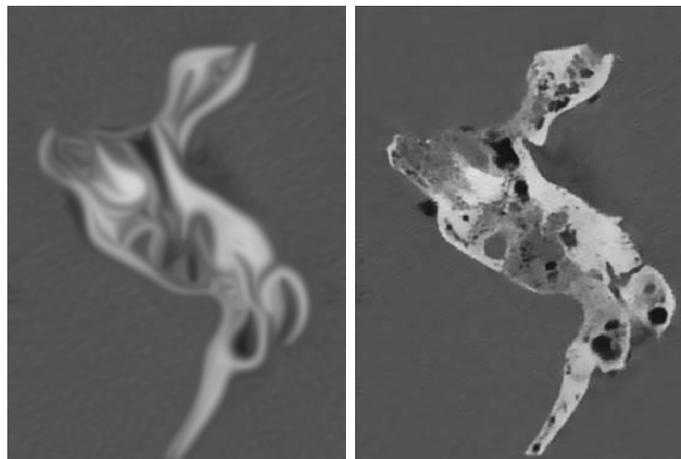
- Strukturiertem Vordergrund (SV1 und SV2).
- Kante (KA).
- Hintergrund (HG).
- Ecke (EK).
- Zeitaufwand für die Durchführung der Filterung (ZA).

Die isotrope Gaußfilterung ist dafür bekannt, dass sie Ecken und Kanten stark verschmiert. Die isotrope Gaußfilterung ist motiviert durch die Diffusionsgleichung. Der Versuch, eine passende Adaption der Parameter vorzunehmen, scheitert in der Praxis an den beiden wichtigen Fähigkeiten, signifikante Bildstrukturen zu erhalten, bei gleichzeitiger Minimierung des Rauschens. Beide Anforderung sollten möglichst optimal erfüllt werden, damit die später durchgeführte Rekonstruktion des kinematischen Menschmodells präzise durchgeführt werden kann. Die Anwendung der isotropen Gaußfilterung gewährleistet das nicht. Es ist zwar möglich, den Filter



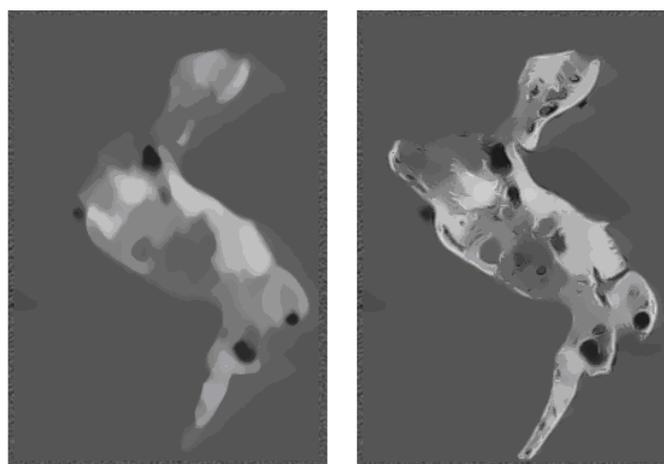
(a) Original.

(b) Strukturanalyse (rot).



(c) Anisotrope Diffusion.

(d) Bilaterale Filterung.



(e) 2D Median.

(f) Orientierter Median.

Abbildung 4.7: Überblick zu den untersuchten nicht-linearen Filtern.

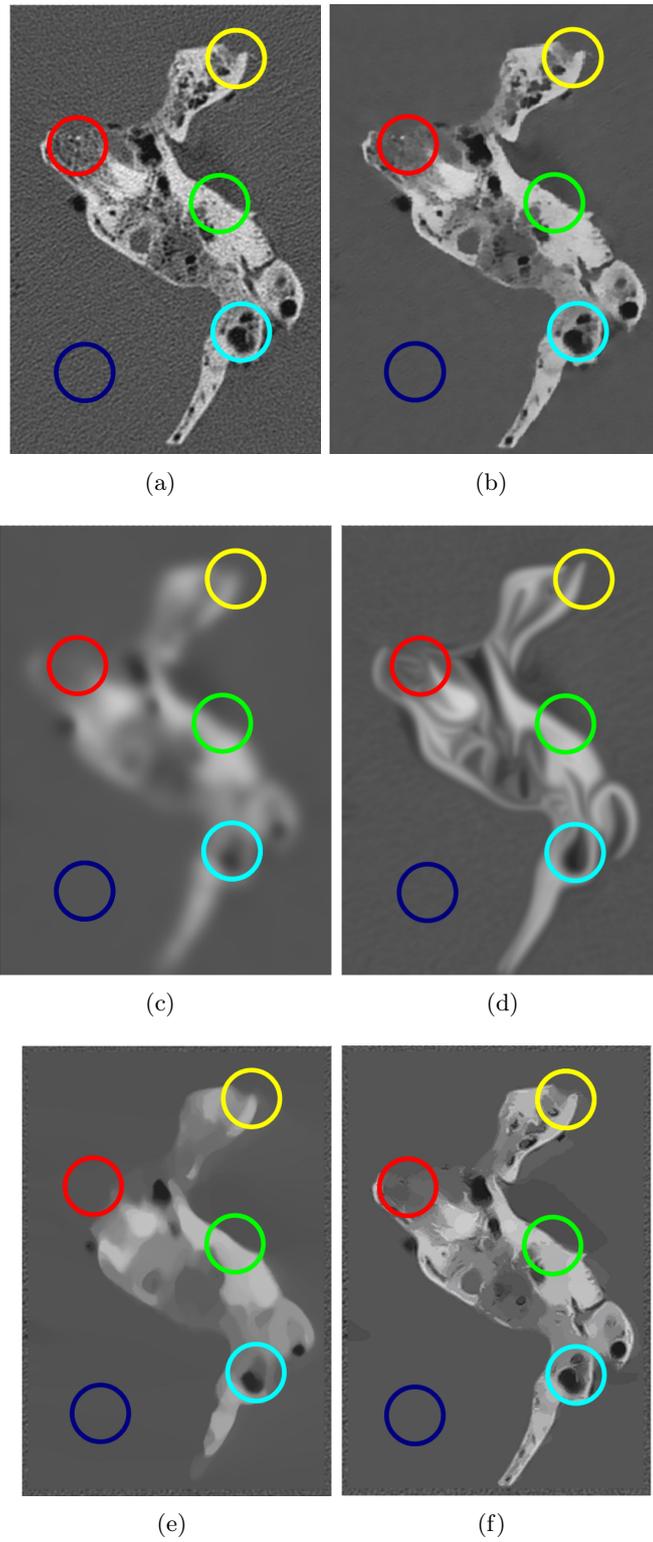


Abbildung 4.8: Darstellung der Resultate nach der Anwendung ausgewählter nicht-linearer Filter.

effizient zu implementieren, insbesondere auf neuartigen Grafikkarten, so dass quasi kaum mehr Berechnungszeit für die Realisierung benötigt wird. Die gerade diskutierten schlechten Eigenschaften disqualifizieren jedoch den Filter.

Die nichtlineare anisotrope Filterung nach [Perona und Malik 1990] erzeugt ausreichende Ergebnisse bei Ecken und Kanten, hat aber große Nachteile in den beiden Teilausschnitten des Vordergrundes. Teilweise verschwinden Informationen des Vordergrundes (SV1) oder es verschmelzen getrennte Bildstrukturen durch die Filterung (SV2). Die Kantenschärfe ist ausreichend und könnte eine Anwendung möglich machen. Die Auswertzeit wird allerdings negativ beurteilt, da der Filter nicht effizient implementiert ist. Es ist aber durchaus möglich, die Implementierung an die Möglichkeiten neuartiger Grafikkarten anzupassen.

Die nichtlineare anisotrope Diffusion nach [Weickert 1998, Weickert 1999] ist ebenfalls nicht effizient implementiert. Es wird stattdessen eine von Herrn Prof. Weickert mitunterstützte Implementierung verwendet [AnisotropeDiffusion]. Die Berechnungen von mehr als 20 min disqualifizieren diesen Filter jedoch von vornherein. Selbst wenn dieser Filter effizient implementiert wird, treten negative Eigenschaften auf (siehe SV1 und SV2). Die Kantenerhaltung ist zwar gut, jedoch bleiben die Strukturen im Hintergrund zu stark erhalten.

Die bilaterale Filterung hat in fast allen Punkten hervorragende Eigenschaften. Der gefilterte Hintergrund weist zu viele Strukturen auf. Allerdings bleiben Kanten, Ecken und insbesondere die kritischen Stellen des strukturierten Vordergrundes (SV1 und SV2) gut erhalten. Damit qualifiziert sich dieser Filter für die Anwendung in MAROCO. Eine Implementierung auf heutzutage erhältlichen Grafikprozessoren ermöglicht eine äußerst effiziente Anwendung des Filters.

Der Medianfilter ist ein klassischer Filter. Es ist bekannt, dass dieser Filter gute Eigenschaften bei der Kantenerhaltung besitzt [Winkler 2006b]. Dieser einfach zu verstehende und einfach zu implementierende Filter überzeugt ebenfalls in fast allen Punkten. Insbesondere bei der Filterung des Hintergrundes, mit der bis jetzt alle bisher diskutierten Filter mittlere bis größere Probleme haben. Schwierigkeiten hat dieser Filter bei Ecken. Größere Schwächen treten bei strukturiertem Vordergrund (siehe SV2 in 4.8) auf. Trotzdem sind die Ergebnisse überzeugend.

Der letzte hier vorgestellte Filter wurde im Rahmen dieser Arbeit entwickelt. Inspiriert durch die positiven Eigenschaften einiger klassischer Filter ist es gelungen, diese innerhalb eines Algorithmus zu vereinen. Nach der bisherigen Diskussion zeigt sich, dass besondere Eigenschaften zum Überwinden beider Probleme, das Rauschen im Hintergrund zu unterdrücken und die Strukturen des Vordergrunds zu erhalten, erzielt werden müssen. Mit dem orientierten Medianfilter geschieht dies, indem einerseits an zugrundeliegende Bildinformationen adaptiert wird (Maskenform und Maskenorientierung) und andererseits auf den klassischen Medianfilter gesetzt wird, der von vornherein sehr gute Resultate liefert.

Eine effiziente Implementierung auf heutzutage verfügbaren Grafikkarten wurde im Rahmen dieser Arbeit ebenfalls vorgenommen. Beide Filter, der orientierte Medianfilter und der bilaterale Filter, erreichen damit Laufzeiten von unter 1 ms. Der orientierte Medianfilter zeigte in fast allen Kriterien sehr gute Ergebnisse. Lediglich die Filterantwort beim strukturierten Vordergrund ist mit gut zu beurteilen.

Der bilaterale Filter und der orientierte Medianfilter erfüllen beide die Ansprüche für eine gute Vorverarbeitung und können deshalb im Rahmen dieser Arbeit eingesetzt werden.

4.2 Hintergrundmodellierung

Hintergrundrepräsentationen, gewonnen aus Trainingsdaten, sind eine häufig verwendete Technik der Bildfolgenauswertung. Bei der Bildfolgenauswertung müssen wichtige Informationen in den Bildfolgen von Unwichtigen getrennt werden. Deshalb hat man es mit einem Segmentierungsproblem zu tun.

Bei der Mensch–Maschine–Interaktion sind bspw. die wichtigen Informationen die sich bewegenden Menschen und bewegten Roboter. Die Bildpunkte, korrespondierend zu Menschen oder Robotern, werden als Bildpunkte des Vordergrund definiert. Die weniger relevanten Bildpunkte sind z.B. die Bildpunkte, erzeugt durch statische Hintergrundobjekte, wie bspw. die Steuereinheiten der Roboter. Im Idealfall soll jeder Bildpunkt eindeutig dem Vorder– oder Hintergrund zugeordnet werden.

4.2.1 Einleitung

Auf den ersten Blick lässt sich das Problem der Vordergrundsegmentierung scheinbar intuitiv lösen, indem man die Differenz zwischen einer Hintergrundrepräsentation in Form eines Bildes B und der aktuellen Bildfolge I in Abhängigkeit eines Schwellwertes Θ berechnet:

$$|I_i - B| > \Theta. \quad (4.14)$$

Diese Vorgehensweise hat gravierende Nachteile, da in der hier repräsentierten Form keine Adaptionsmöglichkeit an die sich verändernde Szene hat. In der Realität ist der Hintergrund nicht konstant. Er verändert sich mit der Zeit und wird durch folgende Punkte beeinflusst:

- Systematische Störungen des Bildaufzeichnungsprozesses.
- Veränderte Lichtverhältnisse durch externe Lichtquellen, z.B. Tageslicht, plötzliches An- oder Ausschalten von Lichtquellen².
- Störungen, erzeugt durch den Lichtausbreitungsprozess, z.B. Reflektionen und Schatteneffekte.
- Objekte, die seit längerem keine Bewegung aufweisen, z.B. Werkstücke, Behältnisse usw.

Deshalb sollte ein für die Segmentierung eingesetztes Verfahren adaptiv und lernfähig sein, um eine effiziente Klassifikation in Vordergrund und Hintergrund durchzuführen.

4.2.2 Gaußsche Mischmodelle zur Hintergrundschätzung

In diesem Unterabschnitt werden Verfahren basierend auf gaußschen Mischverteilungen für die Hintergrundmodellierung vorgestellt. Es reicht in manchen Fällen aus, einen einzelnen Bildpunkt durch eine einzige gaußsche Funktion zu modellieren. Die gaußsche Dichtefunktion kann das Rauschen der Kamera durch geeignete Parametrierung robust verarbeiten.

In der Realität wird die Szene gebildet durch die Oberflächen der Szenegegenstände. In der Realität verhält sich die Beleuchtung der Szene nicht konstant. Da Bilder außerdem mehrkanalig sein können (bspw. Farbbildkameras, oder wie hier Tiefen– und Amplitudeninformationen erzeugen) und die Szene sich dynamisch verhält, sind mehrere Verteilungen für die Modellierung nötig.

²Trotz der scheinbar gut funktionierenden Fremdlichtunterdrückung führen Fremdlichteinflüsse zu einem verändernden Messverhalten bzgl. der Amplituden und Tiefen.

Die Vorgehensweise

Das Verfahren kann in vier Phasen eingeteilt werden:

1. Zuerst werden die gaußschen Mischmodelle (GMM) definiert.
2. Anschließend muss beschrieben werden, wie sich die Parameter mit der Zeit ändern. Bei Parameteränderungen wird eine Hintergrundschätzung heuristisch durchgeführt und man erhält dann eine Entscheidung, welche Bildpunkte zum Hintergrund bzw. Vordergrund gezählt werden müssen.
3. Danach werden Bildpunkte als Zusammenhangskomponenten zusammengefasst.
4. Abschließend werden die Daten der Bildfolgenauswertung übergeben und dort weiter verarbeitet.

Gaußsche Mischmodelle

Die Werte jedes Bildpunktes werden zur Laufzeit beobachtet. Zu einem gegebenen Zeitpunkt t sei der Verlauf eines Bildpunktes bekannt:

$$\{X_1, \dots, X_t\} = \{I(x_0, y_0, i) : 1 \leq i \leq t\} \quad , \quad (4.15)$$

wobei I die Bildfolge ist.

Der Wert bleibt konstant, solange sich der Hintergrund und die Beleuchtung nicht verändern. Falls die Messungen beeinflusst werden, kann man diese Veränderung modellieren, denn falls eine Störung, z.B. durch eine Beleuchtungsänderung in einer statischen Szene auftritt, muss diese von der gaußschen Verteilung erkannt werden.

Die vergangenen Werteverläufe jedes Bildpunktes werden durch eine gaußsche Mischverteilung modelliert. Das Mischmodell eines Bildpunktes wird definiert durch die folgende Gleichung:

$$P(X_t) = \sum_{i=1}^K \omega_{i,t} \cdot G(X_t, \mu_{i,t}, \Sigma_{i,t}) \quad , \quad (4.16)$$

wobei K die Anzahl der Verteilungen, $\omega_{i,t}$ eine Schätzung des Gewichts der i -ten gaußschen Mischverteilung zum Zeitpunkt t , $\mu_{i,t}$ die Mittelwerte repräsentieren und durch $\Sigma_{i,t}$ die Kovarianzmatrizen beschrieben sind. Die Funktion G ist eine gaußsche Dichtfunktion und deshalb folgendermaßen definiert:

$$G(X_t, \mu, \Sigma) = \frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}(X_t - \mu)^T \Sigma^{-1} (X_t - \mu)} \quad . \quad (4.17)$$

Der Parameter K wird durch den zur Verfügung stehenden Speicherplatz und Rechenaufwand beschränkt. Ein in den hier durchgeführten realen Experimenten üblicher Wert für K liegt zwischen 3 und 5.

Für mehrkanalige Bilder wird aus Gründen des Rechenaufwands angenommen, dass die Werte der verschiedenen Kanäle unkorreliert sind und identische Varianz besitzen. Damit werden aufwendige Berechnungen, erzeugt durch eine notwendige Matrixinvertierung, eingespart. Die Kovarianzmatrix Σ wird deshalb folgendermaßen definiert:

$$\Sigma_{k,t} = \sigma_k^2 I \quad . \quad (4.18)$$

Durch GMM werden die Verteilungen der Werte der zuletzt beobachteten Bildfolgen beschrieben. Die neuen Werte werden dann hauptsächlich durch ein Modell abgebildet und dienen dazu, dieses Modell zu aktualisieren.

Aktualisierung der Modelle

Es gibt verschiedene Methoden, das Modell abzubilden und zu aktualisieren. Beispielsweise kann man die ML³-Methode oder den EM⁴-Algorithmus verwenden, wobei, um es vorwegzunehmen, für diese Methoden ein zu großer Rechenaufwand benötigt wird.

Deshalb wird hier eine echtzeitfähige Approximation mit K -Mittelwerten zur Aktualisierung des Modells verwendet.

Schwellwert für die Übereinstimmung

Jeder neue Wert X_t wird mit einer einzelnen gaußschen Verteilung von K Modellen verglichen. Hierfür kann man einen Schwellwert definieren. Liegt der aktuelle Wert unterhalb des Schwellwertes, ist eine Übereinstimmung gefunden.

Ein einheitlicher Schwellwert ist zwar eine einfache Lösung, kann aber zu Problemen führen, falls die Szene unterschiedlich beleuchtet wird, was bei der hier verwendeten PMD-Kamera im Allgemeinen der Fall ist. Deswegen ist der Schwellwert als das 2,5-fache der Standardabweichung der Gaußverteilung definiert. Der Schwellwert erhöht sich in dunkleren Bereichen, in denen das Rauschen zunimmt und die Standardabweichung größer ist. Dadurch erhöht sich die Chance, dass sich die in der Szene befindenden Entitäten erfasst werden.

Falls die Werte X_t mit keiner der K Verteilungen übereinstimmt, wird eine neue Verteilung erstellt. Der Mittelwert wird durch die aktuellen Werte definiert, wobei eine höhere Varianz und eine niedrige Priorität vergeben werden.

Aktualisierung der Priorität

Die Prioritäten der Verteilungen werden mit der Zeit t nach [Stauffer und 1999] selbständig aktualisiert:

$$\omega_{k,t} = (1 - \alpha)\omega_{k,t-1} + \alpha M_{k,t} \quad , \quad (4.19)$$

wobei α die Lerngeschwindigkeit ist. Falls eine Übereinstimmung getroffen wird, wird $M_{k,t} = 1$ gesetzt, sonst wird $M_{k,t} = 0$ gesetzt. Dadurch erhöht sich der Wert der Priorität, falls ein Modell bereits in der Vergangenheit verwendet wird. Falls nicht, nimmt der Wert ab. Dadurch werden die Prioritäten dynamisch zugeordnet.

$M_{k,t}$ verhält sich in diesem Fall, wie der durch Tiefpassfilterung ermittelte Mittelwert der a posteriori Wahrscheinlichkeit, die Beobachtung der Übereinstimmung eines Modell im Laufe der Zeit von 1 bis t .

Aktualisierung des Mittelwertes und der Varianz

Für ein nicht abgebildetes Modell bleiben der Mittelwert μ und die Varianz σ unverändert. Falls eine Übereinstimmung vorliegt, werden die beiden Parameter der Gauß-Funktion wie

³engl. maximum likelihood

⁴engl. expectation and maximation

folgt aktualisiert:

$$\mu_t = (1 - \rho)\mu_{t-1} + \rho X_t \quad , \quad (4.20)$$

$$\sigma_t^2 = (1 - \rho)\sigma_{t-1}^2 + \rho(X_t - \mu_t)^T(X_t - \mu_t) \quad , \quad (4.21)$$

wobei

$$\rho = \alpha\eta(X_t|\mu_k, \sigma_k) \quad (4.22)$$

abhängig von der Lerngeschwindigkeit und der Dichtefunktion ist, um die aktuellen Verteilung anzupassen. Damit werden die beiden Parameter μ und σ , ähnlich wie die Priorität $\omega_{k,t}$, mit der Zeit adaptiert.

Ein wesentlicher Vorteil bei diesem Verfahren liegt darin, dass sich das Objekt zu einem Teil des Hintergrundes entwickelt, ohne dabei den Hintergrund sofort zu überschreiben. Der Teil des Hintergrundes bleibt immer noch im Modell erhalten, aber die Wahrscheinlichkeit wird langsam reduziert. Erst nach einer bestimmten Zeit verringert sich die Priorität und der Teil des Hintergrundes wird durch ein neues Modell ersetzt, falls ein Objekt zum Hintergrund gezählt wird und dieses sich dann plötzlich wieder bewegt. Die Verteilung des vorherigen Hintergrundes existiert noch und ist verfügbar mit dem gleichen Mittelwert und der gleichen Varianz, aber mit einer geringeren Priorität. Trotzdem lässt sich der Hintergrund bzw. Vordergrund effizient rekonstruieren.

Schätzung des Hintergrundes

Wann immer sich die Parameter des Mischmodells ändern, muss eine Entscheidung getroffen werden, welche Kombination der gaußschen Dichtefunktionen das Hintergrundmodell am besten repräsentiert. Man benötigt für diese Entscheidung aussagekräftige Faktoren. In [Stauffer 2002] wird dies erreicht, indem das Verhältnis der Priorität zur Varianz verwendet wird. Deshalb wird eine Zuordnung der Gaußverteilungen durch ω/σ modelliert. Dieses Verhältnis erhöht sich, falls eine Verteilung mehr Priorität bekommt und dadurch aussagekräftiger wird, und bzw. oder falls die Varianz der Verteilung niedriger wird. Hier werden nur die übereinstimmenden Verteilungen zugeordnet, da nur die Werte solcher Verteilungen sich verändert haben. Die Zuordnung des Modells ist durch eine korrespondierende Liste beschrieben, wobei sich die am höchsten wahrscheinliche Hintergrundverteilung an oberster Position befindet. Die niedrigste Verteilung wird nach unten verdrängt und ggf. durch eine neue Verteilung ersetzt. Die ersten B Verteilungen werden als Hintergrundmodell ausgewählt.

Dabei gilt:

$$B = \arg \min_b \left(\sum_{k=1}^b \omega_k > T \right) \quad . \quad (4.23)$$

Der Schwellwert T bestimmt, wie viele Daten mindestens zum Hintergrund beitragen sollen. Durch diesen Schwellwert wird die beste Verteilung hervorgehoben. Wenn T zu klein ist, kann das Hintergrundmodell unimodal sein. Falls T einen großen Wert hat, ist das Hintergrundmodell multimodal und wird von mehreren Verteilungen approximiert. So können die Hintergrundbewegungen, die durch mehrere Intensitäten aufgenommen sind, gut beschrieben werden.

4.2.3 Parameterbestimmung und –analyse der gaußschen Mischmodelle

In [Stauffer und 1999] wird beschrieben, wie das Mischmodell die ganze Szene modelliert. Nicht nur der Hintergrund, sondern auch der Vordergrund werden durch das Mischmodell beschrieben. Um in der Praxis einen adaptiven Hintergrund zu modellieren, sind deswegen mindestens drei Gaußmodelle notwendig, wobei zwei davon verwendet werden, um den Hintergrund zu modellieren und das verbleibende Modell eingesetzt wird, um den Vordergrund zu beschreiben.

Wenn ein Bildpunkt nicht zum Hintergrundmodell gehört, gibt es zwei Möglichkeiten:

- Der Bildpunkt stimmt mit keinem existierenden Gaußmodell überein. Der Bildpunkt wird als Vordergrund erkannt und ein neues Gaußmodell wird anhand dieses Bildpunktes generiert.
- Der Bildpunkt stimmt mit einem existierenden Gaußmodell überein. Das Modell befindet sich, wegen der geringen Priorität oder einer hohen Varianz nicht im Hintergrundmodell. Der Bildpunkt wird zu diesem Modell gezählt und das Modell wird durch diesen Bildpunkt aktualisiert.

Für K sollte keine Zahl kleiner als drei ausgewählt werden, denn dann gibt es höchstens eine Gaußfunktion, die den Hintergrund modelliert. Diese verhält sich dann analog zu den Differenzbildverfahren.

Die Lerngeschwindigkeit α

In [Stauffer und 1999] ist die Lerngeschwindigkeit α durch einen bestimmten Wert festgelegt. Das führt zu Problemen bei der Aktualisierung des Hintergrundes:

- Wenn man α zu klein wählt, werden die Parameter ω und μ , sowie σ zu langsam, d.h. nicht echtzeitfähig, von der aktuellen Bildfolge gelernt. Obwohl der Hintergrund über einen bestimmten Zeitraum hinweg stabil ist, ist die Priorität nicht groß genug, um das Hintergrundmodell zu beschreiben. Falls eine Bewegung in der Szene stattfindet, wird ein neues Gaußmodell erstellt, mit kleinerem ω . Trotzdem werden die entsprechenden Bildpunkte als Hintergrund beschrieben und nicht zum Vordergrund gezählt.
- Wenn man α zu groß wählt, wird das Hintergrundmodell zu schnell an die aktuelle Bildfolge adaptiert. Daraus folgt, dass ein Objekt in sehr kurzer Zeit zum Hintergrund gezählt wird und nicht als zum Vordergrund gehörend erkannt wird.

Um dieses Problem zu lösen, wird in [Power und Schoones 2002] die Lerngeschwindigkeit α_t in Abhängigkeit von dem Zeitstempel t definiert zu:

$$\alpha_t = 1/t \quad , \quad (4.24)$$

mit $t = 1, 2, 3, \dots$. Der Vorteil liegt darin, dass das Mischmodell am Anfang für ein grosses α_t sehr schnell adaptiert. Es dauert beispielsweise für einen statischen Hintergrund nicht mehr so lange, bis eine hohe Priorität und niedrige Varianz erreicht ist. Mit der Zeit wird die Adaptionszeit immer größer, evtl. geht diese sogar gegen unendlich und die Lerngeschwindigkeit α_t wird immer kleiner und geht im Extremfall gegen 0. Man kann deshalb eine untere Grenze $\alpha_t = \alpha_{min}$ festlegen. Falls diese untere Grenze erreicht ist, wird die Aktualisierung nicht mehr verlangsamt und der Hintergrund wird von einem zeitkonstanten Wert adaptiert.

Die Aktualisierung von Mittelwert und Varianz

Ein anderer entscheidender Parameter für die Aktualisierung des gaußschen Mischmodells ist die Geschwindigkeit der Konvergenz $\rho_{k,t}$. Durch $\rho_{k,t}$ lassen sich der Mittelwert μ und die Varianz σ jedes Gaußmodells neu berechnen:

$$\mu_t = (1 - \rho_{k,t})\mu_{t-1} + \rho_{k,t}X_t \quad , \quad (4.25)$$

$$\sigma_t^2 = (1 - \rho_{k,t})\sigma_{t-1}^2 + \rho_{k,t}(X_t - \mu_t)^T(X_t - \mu_t) \quad . \quad (4.26)$$

Die Berechnung, die in [Stauffer und 1999] vorgestellt wird, ist eine Kopplung von der Lerngeschwindigkeit α und die Schätzung der gaußschen Dichtefunktion:

$$\rho_{k,t} = \alpha\eta(X_t|\mu_k, \sigma_k) \quad (4.27)$$

$$= \alpha \cdot \omega_t \cdot \frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}(X_t - \mu_t)^T \Sigma^{-1} (X_t - \mu_t)} \quad . \quad (4.28)$$

Diese Berechnung ist für die Anwendung oftmals ungeeignet, da $\rho_{k,t}$ aufgrund der gaußschen Dichtefunktion zu gering ist. Bei der Initialisierung werden σ_k in der Regel sehr groß und $\omega_{i,t}$ sehr klein gesetzt, deshalb ist $\rho_{k,t} \approx 0$. Die σ werden dadurch groß. Dies kann zu Problemen führen. Im Extremfall kann der Hintergrund nicht modelliert werden.

Eine bessere Berechnung lässt sich durch Anwendung des Online-EM⁵-Algorithmus erzielen:

$$\mu_{k,t+1} = \mu_{k,t} + \frac{P(k|X_{T+1}, \Theta)}{\sum_{t=1}^{T+1} P(k|X_t, \Theta)} (X_{T+1} - \mu_{k,t}) \quad , \quad (4.29)$$

$$\sigma_{k,t+1}^2 = \sigma_{k,t}^2 + \frac{P(k|X_{t+1}, \Theta)}{\sum_{t=1}^{T+1} P(k|X_t, \Theta)} ((X_{T+1} - \mu_{k,t})(X_{t+1} - \mu_{k,t})^T - \sigma_{k,t}^2) \quad . \quad (4.30)$$

Diese Berechnung ermöglicht eine genauere und deutlich schnellere Approximation von μ und σ . Der Hintergrund wird somit nach kürzerer Zeit gelernt und verhält sich stabiler.

Das Problem bei diesem Verfahren ist, dass die Berechnungen im Laufe der Zeit zeitaufwendiger werden. Aufgrund der Rechenkomplexität muss die Berechnung nach einer gewissen Zeit t auf die letzten L Beobachtungen eingeschränkt werden. Ab diesem Zeitpunkt verhält sich dieses Verfahren ähnlich dem Verfahren von [Stauffer und 1999].

Ab dem Zeitpunkt t profitiert das System nicht mehr von der präzisen und schnellen Konvergenz. Danach reagiert das System langsamer auf Beleuchtungsänderung bzw. Variationen in den Tiefenmessungen oder auf neu eintretende Szeneobjekte.

Eine andere Berechnung wird in [Lee 2005] vorgestellt, um eine genauere und schnellere Konvergenz bei Einführung des neuen Objektes weiterhin behalten zu können. Dazu wird ein Zähler c_k für jedes Gaußmodell eingeführt:

$$c_{k+1} = c_k + q_k \quad , \quad (4.31)$$

wobei

$$q_k = 1, \quad \text{falls } k = \mathit{arg!max}_j p_j \quad (4.32)$$

$$q_k = 0, \quad \text{sonst} \quad . \quad (4.33)$$

⁵EM steht für Expectation-Maximization

Der Zähler c_k wird addiert, falls für das Gaußmodell eine positive Übereinstimmung getroffen wird. Falls das Gaußmodell nach längerer Zeit nicht mehr aktualisiert wird und durch ein neues gaußsches Modell ersetzt wird, wird c_k auf 1 zurückgesetzt.

Die Konvergenzgröße $\rho_{k,t}$ wird dann wie folgt berechnet:

$$\rho_{k,t} = q_k \cdot \left(\frac{1 - \alpha}{c_k} + \alpha \right) . \quad (4.34)$$

Kurz nach der Initialisierung des Gaußmodells ist $\rho_{k,t} \approx 1/c_k$ und das Gaußmodell wird ähnlich gut wie bei Verwendung des Online-EM-Algorithmus aktualisiert. Mit der Zeit wird c_k größer und $1/c_k$ entsprechend kleiner. Nach einiger Zeit ist $\rho_{k,t} \approx \alpha$ und das Gaußmodell verhält sich wie das Verfahren von [Stauffer und 1999]. Der wesentliche Vorteil ist, dass man bei der Beleuchtungsänderung oder bei der Einführung des Objektes die genauere und schnellere Konvergenz nicht verliert und ein stabileres System erhält.

Die Hintergrundsubtraktion führt zu einer binären Klassifikation basierend auf $P(B|x)$, wobei x der Bildpunktwert zum Zeitpunkt t und B die Klasse des Hintergrunds ist. Das Modell wird beschrieben durch

$$P(x) = \sum_{k=1}^K P(G_k) P(x|G_k) \quad (4.35)$$

$$= \sum_{k=1}^K \omega_k \cdot g(x; \mu_k, \tau_k) . \quad (4.36)$$

Die a posteriori Wahrscheinlichkeit wird durch die Komponenten $P(G_k)$ und $P(x|G_k)$, sowie eine Schätzung der Wahrscheinlichkeitsdichte $P(B|G_k)$ formuliert:

$$P(B|x) = \sum_{k=1}^K P(B|G_k) P(G_k|x) \quad (4.37)$$

$$= \frac{\sum_{k=1}^K P(x|G_k) P(G_k) P(B|G_k)}{\sum_{k=1}^K P(x|G_k) P(G_k)} . \quad (4.38)$$

Bei der Initialisierung und Aktualisierung der Modelle sind die Komponenten $P(G_k)$ und $P(x|G_k)$ bekannt. Die Bestimmung der Wahrscheinlichkeitsdichte $P(B|G_k)$ ist ein heuristisches Problem und eine Schätzung wird anhand des Vorwissens über den Hintergrund unterschiedlich dargestellt.

In [Stauffer und 1999] wird $P(B|G_k)$ so gesetzt, dass das Gaußmodell, das die höchste Zuordnung ω/σ besitzt, zu 1 gesetzt wird, wobei die anderen zu null gesetzt werden.

In dieser Arbeit wird eine sigmoide oder s-förmige Funktion, beschrieben durch ω und σ , trainiert und die Wahrscheinlichkeitsdichte $P(B|G_k)$ approximiert wird durch:

$$P(B|G_k) = f\left(\frac{\omega}{\sigma}; a, b\right) \quad (4.39)$$

$$= \frac{1}{1 + e^{-a \cdot \frac{\omega_k}{\sigma_k} + b}} . \quad (4.40)$$

Die beiden Parameter a und b bestimmen, wie sensibel sich die Funktion verhält.

Da $P(B|x)$ und $P(B|G_k)$ nun bekannt sind, kann man den Bildpunkt x als Hintergrund erkennen, falls $P(B|x) < T$ ist. T ist ein Schwellwert und bestimmt, wie viele Daten zur Hintergrundmodellierung beitragen sollen. Es gibt derzeit keine formale Grundlage dafür, welche Werte für T eingesetzt werden sollen, deshalb muss Expertenwissen integriert werden.

Amplitudeninformation

Die Informationen, welche von der PMD-Kamera geliefert werden, bestehen aus zwei Teilen:

1. Die Tiefeninformation zur Abstandsmessung.
2. Die Amplitudeninformation in Form von Intensitäten.

Jeder Bildpunkt erhält neben der Abstandsmessung zusätzlich die Amplitude des reflektierten Signals. Die Amplitude ist von der Sendestärke der Quelle, der Empfindlichkeit der Aufnehmer und einer geometrisch bedingten Intensitätsabnahme der Distanz abhängig.

Es wird für jeden Bildpunkt gemessen, wie stark die Signalamplitude ist. Ist die Amplitude an einer Stelle zu gering, empfängt die Kamera lediglich ein schwach reflektiertes Signal. Es kann sein, dass sich das Objekt weiter entfernt von der Kamera befindet oder einfach zu schwach beleuchtet wird. Es kann aber auch sein, dass die Oberfläche des Objektes ungeeignete Eigenschaften besitzt und deshalb die Remission gering ausfällt.

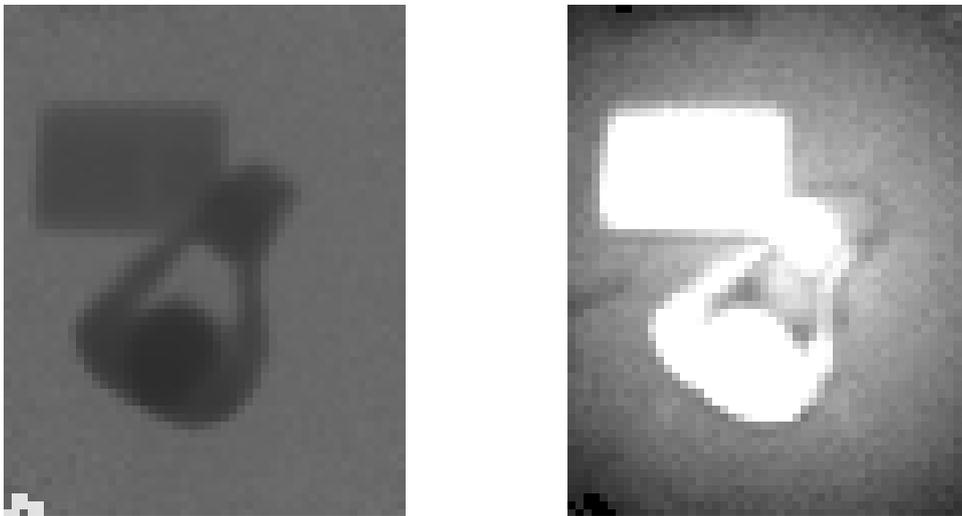


Abbildung 4.9: Darstellung der Tiefen- und Amplitudeninformation.

Die Information des Intensitätbildes ist für die Hintergrundrepräsentation nützlich. In erster Linie wird die Messunsicherheit aus dem Intensitätsbild die Aktualisierung beeinflussen. Das im Rahmen dieser Arbeit weiterentwickelte Verfahren zum Einsatz einer Hintergrundrepräsentation für Tiefenbildfolgen, basiert auf den Verfahren von [Stauffer und 1999] und [Lee 2005], wurde aber an die Bedürfnisse des neuartigen Kamerasystems angepasst.

Aktualisierung des gaußschen Modells

Die Tiefenbildfolge wurde bei den anderen Verfahren ähnlich wie die Grauwertbilder behandelt. Da die Tiefen- und Amplitudeninformation zusätzlich die Unsicherheit gegenüber der Messung darstellt, können diese mit der Lerngeschwindigkeit gekoppelt werden, um den Mittelwert und die Varianz effizienter zu aktualisieren.

Dazu wird u.A. vorgeschlagen, die Tiefeninformation selbst als zusätzliche Größe für die Gewichtung der Sicherheit der Messung zu betrachten. Man nimmt an, dass die Messunsicherheit

U proportional zum Abstand des Messobjekts ist. Bei Verwendung der Rohdaten gewinnt man den Abstand direkt aus dem Messwert X_t :

$$U_t = c \cdot X_t + u \quad . \quad (4.41)$$

Je geringer der Abstand zwischen Kamera und Messobjekt ist, desto sicherer ist die Messung. Damit kann die Varianz derart aktualisiert werden, dass sie schneller konvergiert, falls der Abstand geringer ist.

Es wird die im vorhergehenden Abschnitt vorgestellte Methode, basierend auf einem neu eingeführten Zähler für die schnelle Konvergenz, verwendet. Die Varianz wird bei der Aktualisierung mit der Messunsicherheit U_t angepasst:

$$\begin{aligned} \sigma_t^2 &= (1 - \rho'_{k,t})\sigma_{t-1}^2 + \rho'_{k,t}(X_t - \mu_t)^T(X_t - \mu_t) \quad , \\ \rho'_{k,t} &= \frac{1}{U} \cdot \rho_{k,t} \\ &= \frac{1}{c \cdot X_t + u} \cdot \rho_{k,t} \quad . \end{aligned} \quad (4.42)$$

Diese Annahme ist intuitiv, aber nicht für komplexe Anwendung geeignet. Eine Idee liegt deshalb darin, zusätzlich die Amplitudeninformation zu berücksichtigen.

Dazu definiert man die Sicherheit über Messung zum Zeitpunkt t durch:

$$S_t = c \cdot A_t + u \quad , \quad (4.43)$$

wobei A_t der Amplitudenwert zum Zeitpunkt t ist.

Ähnlich wie in (4.43) werden die Parameter folgendermaßen aktualisiert:

$$\begin{aligned} \sigma_t^2 &= (1 - \rho''_{k,t})\sigma_{t-1}^2 + \rho''_{k,t}(X_t - \mu_t)^T(X_t - \mu_t) \quad , \\ \rho''_{k,t} &= S \cdot \rho_{k,t} \\ &= (c \cdot A_t + u) \cdot \rho_{k,t} \end{aligned} \quad (4.44)$$

Mit der Zeit wird c_k größer und demnach $1/c_k$ entsprechend kleiner. Nach einiger Zeit ist $\rho_k(t) \approx \alpha$ und das Gaußmodell verhält sich ähnlich wie das Verfahren von [Stauffer und 1999]. Der Vorteil der präzisen und schnellen Konvergenz bleibt erhalten. Die Modelle, welche für längere Zeit kleinere Amplituden haben, werden entsprechend langsamer konvergieren. Das Hintergrundmodell wird schwach und unsicher erkannt und hat deshalb eine geringere Varianz.

Berechnung der Hintergrundsubtraktion

Bei der Berechnung der Hintergrundsubtraktion wird die Amplitudeninformation als ein zusätzliches Maß I_k eingeführt.

Für jedes neue Modell wird I_k durch Setzen von $I_k = 0$ initialisiert. Mit der Zeit werden diese Maße neben μ_k , τ_k und ω_k aktualisiert:

$$S_k(t) = (1 - \alpha)S_k(t - 1) + \alpha i(t), \quad \text{mit } i(t) \in [0, 1], \quad (4.45)$$

wobei $i(t)$ die aktuelle Intensität der Amplitude ist.

Wenn die I_k kleiner sind, weist dies auf eine geringere Intensität in den letzten Bildfolge hin. Das bedeutet, dass die Parameter des Modells präziser sind. Wenn das Modell ein niedriges

τ_k und ein hohes ω_k hat, kann man deshalb ausgehen, dass der Bildpunkt zum Hintergrund gezählt werden muss.

Dazu wird die folgende Gleichung vorgeschlagen:

$$P(B|G_k) = f(\omega/\sigma; a, b) \quad (4.46)$$

$$= \frac{1}{1 + e^{-a \cdot \frac{\omega_k}{\sigma_k} + b - c \cdot I_k + d}} \quad (4.47)$$

Die Parameter a und b bestimmen die Sensibilität der Funktion. Die Parameter c und d die Toleranz gegenüber der Messunsicherheit.

Durch $-c \cdot I_k + d$ bekommt ein Modell mit kleineren Werten für $I_k(t)$ weniger Gewicht und die korrespondierende s-förmige Funktion stabilisiert die Schätzung.

Im Folgenden werden die drei Verfahren experimentell verglichen.

4.2.4 Experimente und Diskussion

Die Verfahren werden auf Basis der Simulationsdaten und realen Daten durchgeführt und miteinander verglichen. Im ersten Teil der Experimente wird zuerst eine Gruppe von Simulationsdaten erstellt. Diese Daten sind eindimensional und synthetisch generiert. Prinzipiell gibt es zwei Arten von Simulationsdaten:

- Daten in denen der Hintergrund dominiert und konstant bleibt;
- Daten in denen zu unterschiedlichen Zeitpunkten Objekteinformationen hinzugefügt und entfernt werden.

Im zweiten Teil der Experimente werden reale Daten der PMD-Kamera verwendet. Die Daten werden bildpunktweise analysiert. Da die realen Daten starke Messunsicherheit aufweisen und oftmals Messfehler beinhalten, ist es wichtig zu sehen, wie sich die unterschiedlichen Verfahren verhalten.

Dazu werden folgende Szenarien aufgezeichnet und ausgewertet:

1. Ein konstanter Hintergrund, in dem kein Objekt oder Mensch eintritt (ca. 5 Minuten).
2. Keine Veränderung des Hintergrundes in den ersten Minuten. Dann durchquert ein Mensch willkürlich die Szene.
3. Ein konstanter Hintergrund (für ca. 10 Sekunden). Anschließend durchquert ein Mensch für einen längeren Zeitraum von mehreren Minuten die Szene.
4. Ein Objekt, das in einem Intervall von ca. 1 Minute auftaucht und dann wieder wegbewegt wird.
5. Ein Objekt, das in einem Intervall von ca. 30 Minuten auftaucht und dann wieder wegbewegt wird.
6. Anfänglich ein leerer Vordergrund. Dann betritt ein Mensch die Szene und zusätzliche Objekte tauchen in der Szene auf. Menschen und Objekte haben kurze Zeiten in denen sie still stehen bevor sie sich weiter bewegen.

7. Zuerst ein leerer Hintergrund. Dann betritt ein Mensch die Szene. Später erscheinen verschiedene Objekte, die sich bewegen, dann für eine kurze Zeit still stehen, bevor sie sich weiterbewegen oder verschwinden.
8. Eine leere Szene am Anfang, dann mit einem Menschen, der sich bewegt und mehrere Objekte greift bzw. verschiebt. Dabei befinden sich die Hände des Menschen stets oberhalb des Tisches.
9. Ein Szenario, bei dem sich von Anfang an ein Mensch in der Szene befindet.

Die ersten Testszenarien werden als statisch, die Szenarien 2 und 3 als dynamisch und die Szenarien 4 und 5 als semi-dynamisch bezeichnet. In den Szenarien 6, 7 und 8 werden statische und dynamische Aspekte fusioniert und sind deshalb als realistische Fälle anzusehen. In der letzten Szene wird ein Spezialfall ausgewertet. Es wird untersucht, ob der Mensch zuverlässig

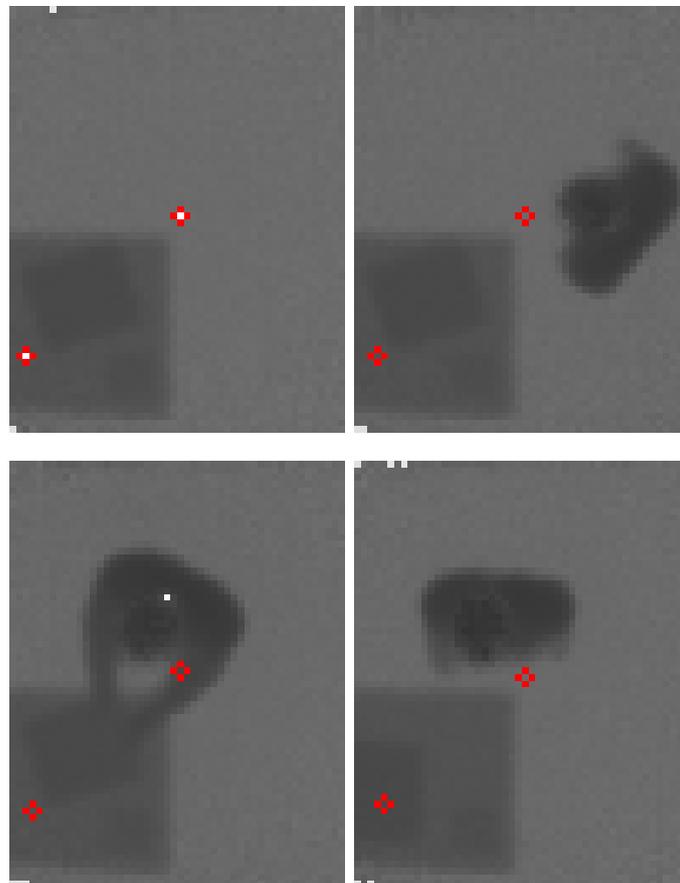


Abbildung 4.10: Ausgewählte Bilder der Experimente zu unterschiedlichen Zeitpunkten. Die unterschiedlichen Tiefendaten für die Bewertung wurden an den gekennzeichneten Stellen in den Bildern entnommen.

im Vordergrund repräsentiert wird, und wie der Tisch, sowie die Objekte, mit der Zeit als Hintergrund eingelernt werden. Um den Verlauf zu veranschaulichen, werden zwei Punkte (P1 und P2) aus der Bildfolge ausgewählt (siehe Abbildung 4.10) und deren korrespondierende Modellparameter aufgezeichnet.

Der Punkt P1 befindet sich mittig in den Bildern der aufgezeichneten Bildfolge. Es existieren keine weiteren Objekte. In den Abbildungen 4.11, 4.12 und 4.13 ist der Verlauf der Mittelwerte, Prioritäten und der Varianzen gut zu erkennen. Und stellt die Unterschiede zwischen den beiden Ansätzen klar heraus.

Der Punkt P2 ist links unten in der Ecke der Szene. Hier taucht der Tisch und die Objekte mit unterschiedlicher Höhe auf. Da das Verfahren von [Power und Schoones 2002] mit einer Lernrate von $\alpha_t = 1/t$ konvergiert, konvergieren die Parameter am Anfang langsamer. Deshalb werden der Tisch und die Objekte von dem gleichen Modell beschrieben und lassen sich nicht voneinander trennen.

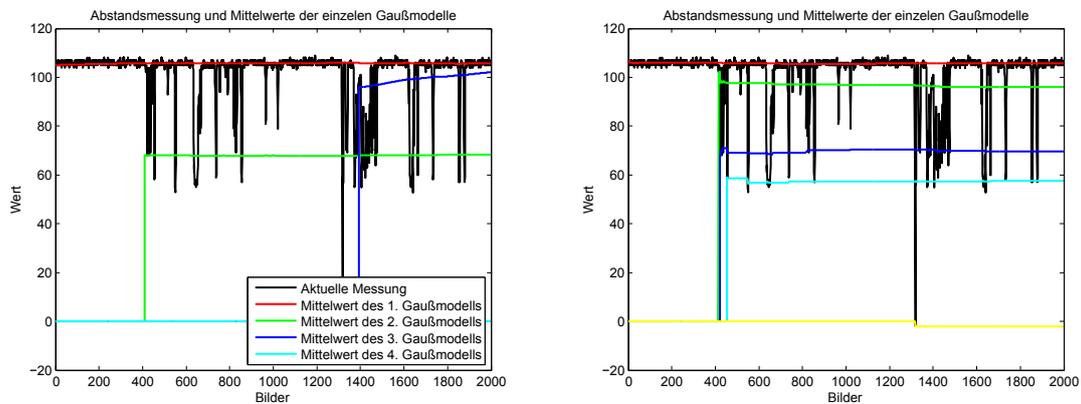


Abbildung 4.11: Vergleich der Mittelwert–Verläufe der einzelnen Gaußmodelle für den Punkt P1 zwischen dem Verfahren von [Stauffer und 1999] und dem hier weiterentwickelten Verfahren.

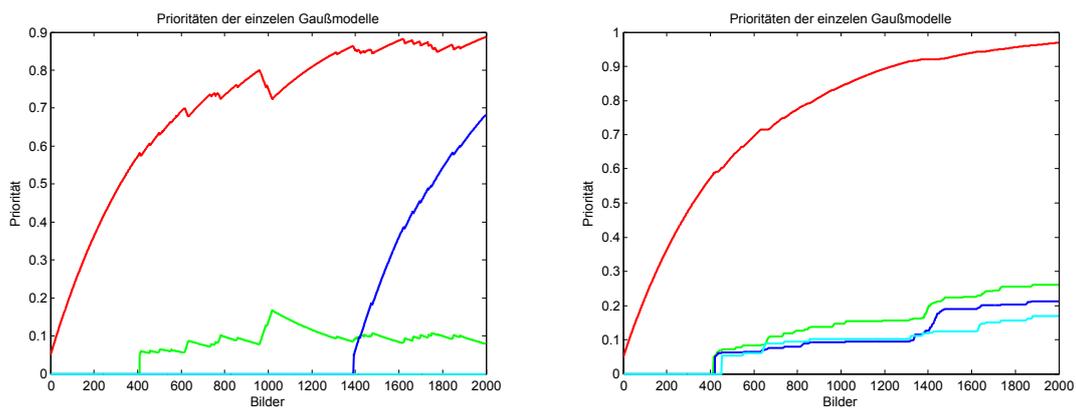


Abbildung 4.12: Vergleich des Prioritätenverlaufs der einzelnen Gaußmodelle im Fixpunkt 1 zwischen dem Verfahren von [Stauffer und 1999] und dem hier weiterentwickelten Verfahren.

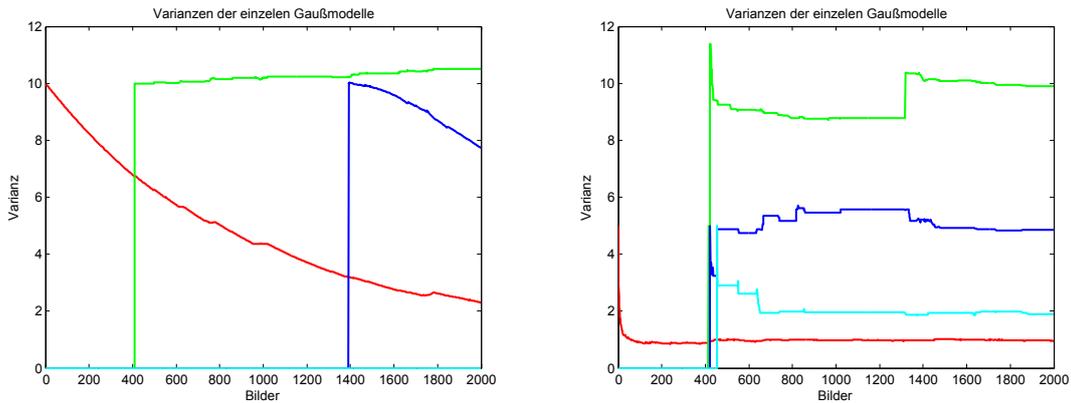


Abbildung 4.13: Vergleich des Varianzenverlaufs der einzelnen Gaußmodelle im Fixpunkt 1 zwischen dem Verfahren von [Stauffer und 1999] und dem hier weiterentwickelten Verfahren.

Auswertung für den allgemeinen Fall

Die für die Roboterzelle geeigneten Parameter werden empirisch ermittelt. In den Experimenten wird K mit den Werten 3, 5, 7 und 9 belegt. Die Ergebnisse der Experimente haben gezeigt, dass man keinen signifikanten Unterschied für $K > 5$ feststellen kann. Das heißt, in den meisten Fällen lässt sich die Umgebung durch weniger als vier Komponenten für das Hintergrundmodell ausreichend gut beschreiben. Eine Belegung von $K < 3$ führt allerdings nicht zu einem befriedigenden Modell, welches den Hintergrund beschreibt.

Da die Lerngeschwindigkeit die Konvergenz der Prioritäten ω direkt beeinflusst, kann man abschätzen, wie lange es dauert, bis die Priorität eines neu erscheinenden Objekts als zum Hintergrund zugehörig klassifiziert wird, bevor es als zum Vordergrund gehörend gezählt wird. Beispielsweise für eine Aufzeichnung mit 30 Bilder pro Sekunde. Falls die Lerngeschwindigkeit auf 0,002 gesetzt ist, benötigt man 347 Bilder, d.h. 11,64 s, um die Priorität von 0 auf 0,5 zu erhöhen. Falls die Lerngeschwindigkeit auf 0,001 gesetzt ist, sind 693 Bilder und 23,1 s notwendig.

Falls die Priorität eines Modells groß wird, bedeutet dies nicht, dass das Modell sich im Hintergrund befindet, da es durch andere Größen beeinflusst wird, wie bspw. durch ω/σ .

Im Verfahren von [Stauffer und 1999] ist die Konvergenz der Varianz σ vergleichsweise langsam gegenüber der im Rahmen dieser Arbeit eingesetzten Methode, was daran liegt, dass die Lerngeschwindigkeit einen festen Wert hat.

Im Verfahren von [Power und Schoones 2002] ist die Konvergenz der Varianz bei der Initialisierung viel schneller als im Verfahren von [Stauffer und 1999]. Für ein neu eingeführtes Objekt nach der Initialisierung benötigt das Verfahren allerdings zu viel Zeit.

Das im Rahmen dieser Arbeit entwickelte Verfahren wird mit den Intensitätsbildern der O3D-Kamera gekoppelt. Damit tritt ein äußerst robustes Verhalten mit einer schnellen Konvergenz ein, besonders für Objekte mit höheren Amplitudenwerten (wie bspw. für den Menschen). Diese Vorgehensweise hat sich in der Praxis, auch aufgrund einer effizienten Implementierung, als nützlich erwiesen. Die Verwendung eines Zählers ermöglicht die schnelle Anpassung an neu in der Szene auftauchende Objekte.

Bewertung spezieller Szenarien

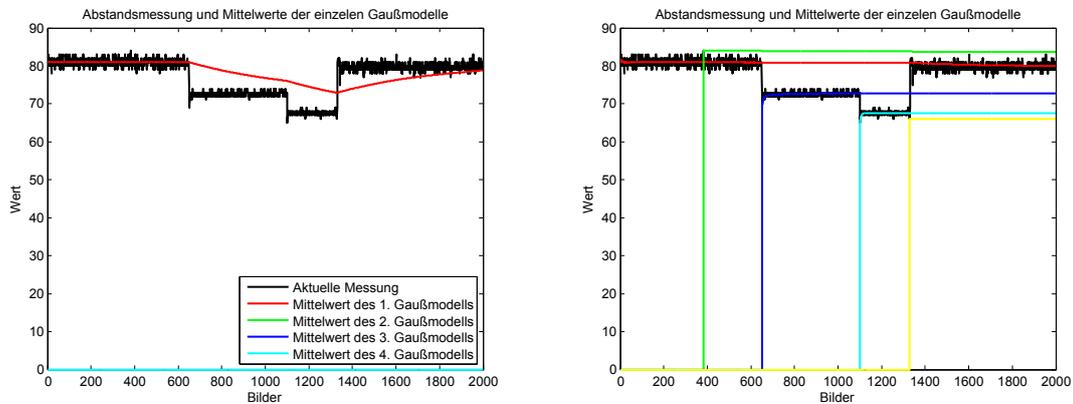


Abbildung 4.14: Vergleich des Mittelwertverlaufs der einzelnen Gaußmodelle im Fixpunkt 2 zwischen dem Verfahren von [Stauffer und 1999] und dem hier weiterentwickelten Verfahren.

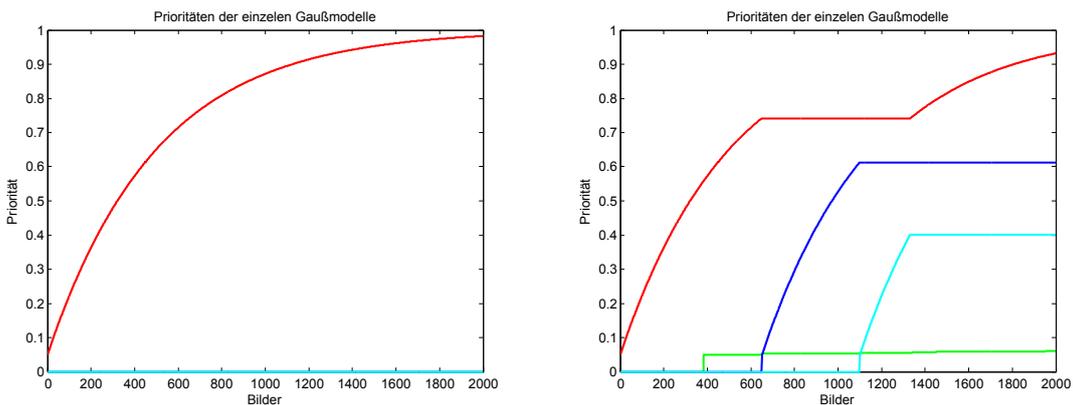


Abbildung 4.15: Vergleich des Prioritätenverlaufs der einzelnen Gaußmodelle im Fixpunkt 2 zwischen dem Verfahren von [Stauffer und 1999] und dem hier weiterentwickelten Verfahren.

Im Vergleich zu dem in [Stauffer und 1999] vorgestellten Verfahren eignet sich das im Rahmen dieser Arbeit entwickelte Verfahren deutlich besser. Das in [Stauffer und 1999] vorgeschlagene Verfahren kann zwar Bildpunkte korrespondierend zu Szeneobjekten des Vordergrundes nach einer bestimmten Trainingszeit segmentieren, was aber auch an der großzügigen Distanz zwischen den von der Kamera wahrgenommenen Körperteilen und dem Boden liegt (siehe Abbildung 4.11), aber die Objekte und die Arme, die sich über dem Tisch befinden, werden nach längere Zeit (über 2000 Bilder) nicht segmentiert. Die Ursache ist die langsame Konvergenz der Varianz (siehe Abbildung 4.13). In Abbildung 4.16 ist der Verlauf der Varianz der anderen Verfahren zu sehen. Die Objekte und die Arme lassen sich dank der schnelle Konvergenz der Varianz, nach knapp 300 Bildern sehr gut vom Tisch trennen. In Abbildung 4.12 ist deutlich zu erkennen, wie wenig differenziert die Prioritäten für das klassische Verfahren gegenüber

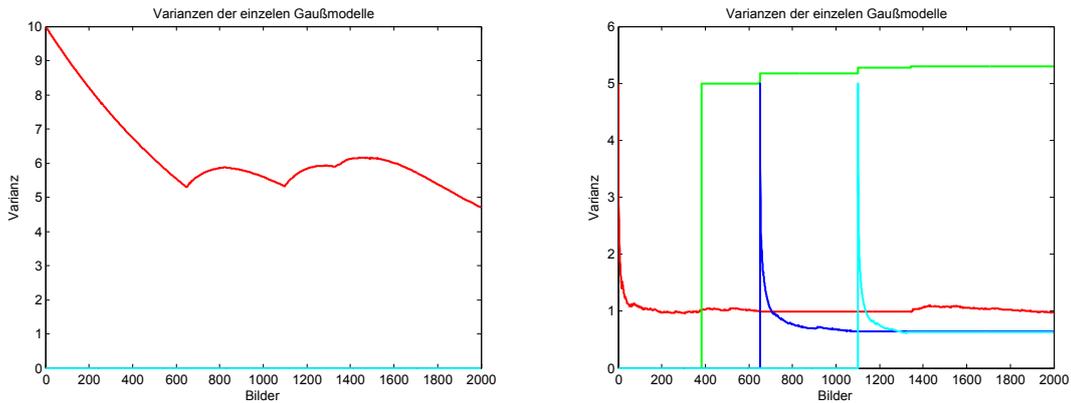


Abbildung 4.16: Vergleich des Varianzenverlaufs der einzelnen Gaußmodelle im Fixpunkt 2 zwischen dem Verfahren von [Stauffer und 1999] und dem hier weiterentwickelten Verfahren.

dem neuen Ansatz sich verteilen. In Abbildung 4.14 ist auch zu sehen, dass das Objekt und der Tisch besser modelliert werden, verglichen mit dem Verfahren von [Stauffer und 1999]. Man muss beachten, dass gerade am Anfang der Initialisierungsphase die Szene konstant bleibt. Ansonsten werden die Szeneobjekte schnell als Hintergrund gelernt und man benötigt anschließend mehr Zeit bis die Objekte wieder als zum Vordergrund zugehörig klassifiziert werden.

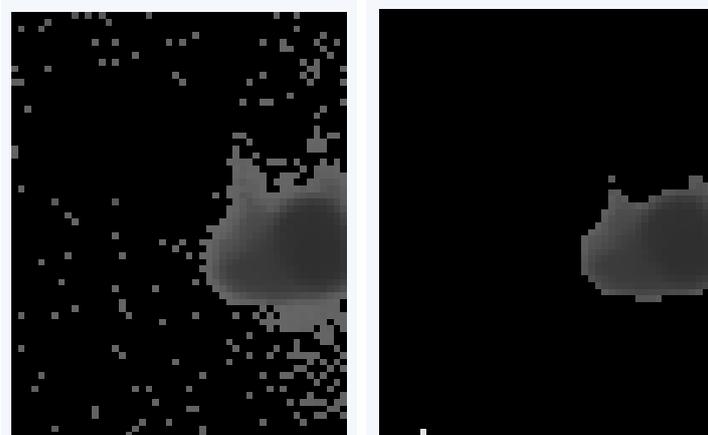


Abbildung 4.17: (Links) Hintergrundsabtraktion ohne Berücksichtigung der Intensitätsbilder. (Rechts) Hintergrundsabtraktion mit Berücksichtigung der Intensitätsbilder.

In Abbildung 4.17 wird das Rauschen in dem hier entwickelten Verfahren durch die zusätzliche Auswertung der Amplitudenbilder deutlich unterdrückt. Die Information, welche die Messung mit einer geringen Intensität als unsicherer einstuft, als vergleichsweise die Messungen mit einer höheren Intensität, beeinflussen die Konvergenz der Varianz. Beispielsweise hat der Boden wegen des größeren Abstands eine geringere Intensität und seine Varianz ist in einer Vielzahl von Fällen höher. Der Tisch und die Arme haben meist bessere Intensitätswerte und die Hin-

dergrundsubtraktion wird dadurch zuverlässiger. Der Randbereich, also dort wo es technisch bedingt häufiger zu Messfehlern kommt, hat ebenfalls eine hohe Varianz vorzuweisen. Die adaptiven Verfahren bieten trotzdem zuverlässige Ergebnisse.

Die positiven Eigenschaften des Verfahrens von [Lee 2005] bleiben erhalten. Probleme entstehen, wenn der Tisch mit einer Oberfläche beschaffen ist, die eine schlechte Remissionsfähigkeit besitzt, d.h. die Intensitätsbilder negativ beeinflusst. Der Abstand zwischen den Armen und dem Tisch ist relativ klein, weshalb die Arme nicht als Vordergrund erkannt werden.

Unterschiedliche Haarfarben können unterschiedliche Reflektionen erzeugen. Insbesondere schwarze Haare sind ungünstig. In diesem Fall entstehen für Bildpunkte, korrespondierend zum menschlichen Kopf, große Messfehler und kleine Amplituden. Da aber der Abstand zwischen dem Kopf und dem Tisch bzw. dem Boden relativ groß ist, hatte die Hintergrundsubtraktion in den durchgeführten Experimenten zuverlässig funktioniert.

4.2.5 Zusammenfassung

Die theoretischen Grundlagen und experimentellen Auswertungen sind für mehrere Verfahren vorgestellt worden. Eine neue, erweiterte Methode wurde eingeführt. Das hier vorgestellte Verfahren ist wegen seiner Adaptivität flexibel einsetzbar. Das Hintergrundmodell, insbesondere bei dem in dieser Arbeit entwickelten Verfahren, wird effizient an die Szene angepasst.

Die Integration der Intensitätsbilder der PMD-Kamera wirkt sich positiv aus, wodurch die Güte des Schätzverfahrens gesteigert werden konnte.

Man stellt fest, dass stochastische Verfahren in manchen Fällen trotz ihrer Vorteile Probleme haben können. Beispielsweise ist ein sich bewegender Mensch sehr gut vom Hintergrund zu trennen. Falls sich ein Mensch für ein hinreichend langes Zeitintervall nicht in der Szene bewegt, wird dieser automatisch zum Hintergrundmodell hinzugefügt. Ein ähnliches Problem existiert auch für die Arme. Falls ein Objekt weggenommen wird und die Arme sich nach kurzer Zeit in der gleichen Position bewegen, muss für die Bestimmung der Parameter, Punkte wie diese stets bedacht werden.

Solche Probleme sind dann zu lösen, falls zusätzliche Information a priori mitmodelliert werden. Für einen still stehenden Menschen sind die Rückkopplung der Ergebnisse der Bewegungsanalyse eine mögliche Lösung.

Eine im Rahmen dieser Arbeit favorisierte Vorgehensweise integriert das Wissen aus dem später beschriebenen Schätzprozess für die menschliche Kinematik in das System, so dass nachdem eine Bildpunktballung als zu einem Menschen gehörend identifiziert wird, diese Information verwendet wird und die Bildpunkte über die Bildfolge hinweg markiert werden. Damit kann eine Zuordnung zum Hintergrund nicht stattfinden.

4.3 Merkmale für die Bewegungsschätzung

In den folgenden drei Unterabschnitten werden verschiedene Merkmale für die im letzten Abschnitt vorgestellten Architekturen eingeführt und diskutiert.

Der Optische Fluss (OF) wird in beiden Architekturen eingesetzt. In der ersten Architektur wird er in einen Variationsansatz integriert und trägt maßgeblich dazu bei, die Dynamischen Konturen zu realisieren. Bei der zweiten Architektur liefert er wichtige Merkmale für die Posenschätzung einer menschlichen Kinematik. Er liefert dort wichtige Informationen über Bewegungsrichtungen und unterstützt auch bei der Prädiktion der menschlichen Bewegung.

Die Dynamischen Konturen basieren auf einem Modell, das zwei unterschiedliche Techniken der Bildverarbeitung und Bildfolgenauswertung in einer Modellgleichung vereint. Die hier vorgestellte Vorgehensweise hebt sich von bisherigen Ansätzen durch verschiedene, signifikante Verbesserungen ab. Im Rahmen der sicheren Mensch-Roboter-Kooperation wurde bisher kein Ansatz vorgestellt, welcher die hier vorgeschlagenen Vorgehensweise berücksichtigt.

Die zweite Architektur verwendet ein parametriertes, kinematisches Modell des Menschen. Das Modell muss in Echtzeit, d.h. im Takt der Kamera, parametrisiert werden. Die kinematischen Parameter, bzw. die Merkmale stammen aus der Bildfolgenauswertung. Der hier vorgestellte Ansatz setzt keine speziellen Voraussetzungen, z.B. in Form von Markern oder Farbinformationen voraus und hebt sich dadurch vom Stand der Technik und Forschung ab.

4.3.1 Grundlagen zur Schätzung des Optischen Flusses

Die beiden klassischen Methoden zur Schätzung des OF gehen auf die Arbeiten von [Horn und Schunck 1981] und [Lucas und Kanade 1981] zurück. Der Optische Fluss hat eine lange Tradition in der Bildfolgenauswertung. Im Rahmen dieser Arbeit soll eine Variante vorgestellt werden, welche an die Auswertung von Tiefenbildfolgen, und damit basierend auf der O3D-Kamera, angepasst ist.

Die hier vorgestellten Grundlagen zur effizienten Schätzung des OF stammen teilweise aus [Graf 2006], sind aber überarbeitet und stark ergänzt worden.

Effizient bedeutet in diesem Fall, dass sowohl die Geschwindigkeit für die Berechnung, als auch die Güte der Schätzung signifikant gesteigert werden konnte. Eine Leistungsschau verschiedener Ansätze aus der Vergangenheit findet man in [Barron u.a. 1994, Galvin u.a. 1998, Baker u.a. 2007].

In den letzten 30 Jahren wurden große Fortschritte bei der Schätzung des OF erzielt. Zwei wichtige Kriterien zur Beurteilung der Güte des OF sind die Dauer der Berechnung, die Winkelgenauigkeit und die Genauigkeit bzgl. der Amplitude.

Im Rahmen dieser Arbeit wird ein an die Bedürfnisse der Mensch-Roboter-Kooperation angepasstes Verfahren vorgestellt. Dabei wird eine Modifikation verschiedener in der jüngeren Vergangenheit vorgestellten Ansätze vorgenommen. Wesentliche neue Resultate haben sich dabei ergeben:

1. Die Einführung der konjugierten Gradienten-Methode als effiziente Löser (Laufzeit- und Parameterreduktion) in den Ansatz von [Brox u.a. 2004].
2. Die Integration eines adaptierten Strukturtensoransatzes in die Gleichungen zur Berechnung des OF (Anisotroper Datenterm).
3. Eine signifikante Senkung der Berechnungszeit durch effiziente Implementierung von Mehrgitteransätzen unter Ausnutzung moderner Grafikprozessoren.
4. Die Überprüfung der Tauglichkeit der neuen Ansätze bei Verwendung der [O3D-PMD]. Dazu wird die Gradientenkonstanzannahme in die Euler-Lagrange-Gleichungen eingeführt, wodurch eine Robustifizierung der Störungen, basierend auf Messrauschen der Tiefenwerte, erzielt wird.
5. Integration des OF in die Bewegungsschätzung (Kalman-Filterung und Prädiktion) eines Menschmodells.

Adaption des Ansatzes von [Brox u.a. 2004]

Einen interessantes Modell zur Schätzung des Optischen Flusses wird in [Brox u.a. 2004] vorgestellt, indem die Autoren nichtlineare Strafterme einführen. Allerdings ergeben sich bei dieser Vorgehensweise rechenaufwendige Zwischenschritte, aufgrund der durch das komplexere Modell induzierten Nichtlinearitäten.

Um Ausreißer weniger stark ins Gewicht fallen zu lassen, wird in [Brox u.a. 2004] zusätzlich eine nicht-quadratische Straffunktion Ψ eingeführt:

$$E_D(u, v) = \int_{\Omega} \Psi(|I(\mathbf{x} + \mathbf{w}) - I(\mathbf{x})|^2 + \gamma|\nabla I(\mathbf{x} + \mathbf{w}) - \nabla I(\mathbf{x})|^2) d\mathbf{x} \quad . \quad (4.48)$$

Es wird in [Brox u.a. 2004] vorgeschlagen, die Funktion $\Psi(s^2) = \sqrt{s^2 + \epsilon^2}$ zu verwenden, mit einer kleinen positiven Konstante, z.B. $\epsilon = 0.001$. Der noch fehlende Glättungsterm lautet folgendermaßen:

$$E_S(u, v) = \int_{\Omega} \Psi(|\nabla_3 u|^2 + |\nabla_3 v|^2) d\mathbf{x} \quad . \quad (4.49)$$

Das gesamte zu minimierende Energiefunktional lautet dann:

$$E(u, v) = E_D + \alpha E_S \quad . \quad (4.50)$$

Die im Folgenden aufgeführten Darstellungen basieren im Wesentlichen auf der Arbeit von [Brox u.a. 2004]. Am Ende dieses Abschnittes wird eine Erweiterung vorgestellt, welche es erlaubt, auf handelsüblichen Rechnern unter Verwendung des [O3D-PMD]-Sensors Bilder in Echtzeit, d.h. mit mehr als 22 Bilder pro Sekunde, zu verarbeiten.

Dem Variationsansatz folgend, muss eine Lösung für die Euler-Lagrange-Gleichungen

$$\begin{aligned} & \Psi'(I_z^2 + \gamma(I_{xz}^2 + I_{yz}^2)) \cdot (I_x I_z + \gamma(I_{xx} I_{xz} + I_{xy} I_{xz})) \\ & \quad - \alpha \operatorname{div} (\Psi'(|\nabla_3 u|^2 + |\nabla_3 v|^2) \nabla_3 u) = 0 \quad , \\ & \Psi'(I_z^2 + \gamma(I_{xz}^2 + I_{yz}^2)) \cdot (I_y I_z + \gamma(I_{yy} I_{yz} + I_{xy} I_{xz})) \\ & \quad - \alpha \operatorname{div} (\Psi'(|\nabla_3 u|^2 + |\nabla_3 v|^2) \nabla_3 v) = 0 \quad , \end{aligned} \quad (4.51)$$

gefunden werden. Durch bspw. I_x werden die partiellen Ableitungen der Bilddaten gekennzeichnet.

Diese Gleichungen sind nichtlinear in ihrem Argument $\mathbf{w} = (u, v, 1)^T$. Um ein lineares Gleichungssystem zu erhalten, ist die Benutzung von Fixpunktiterationen über \mathbf{w} der erste Schritt. Um den zuvor vorgestellten multiskalaren Ansatz (grob–nach–fein) zu implementieren, werden diese mit einer Unterabtastungsstrategie kombiniert. Unterabtastung einer Funktion mit möglichen lokalen Minima bzw. Maxima hat zur Folge, dass die Funktion geglättet wird und u.U. lokale Extrema bei der Lösung überwunden werden können. Hierdurch ist es möglich, eine bessere Approximation des globalen Minimum des Energiefunktional zu erhalten. Es ergibt sich \mathbf{w}^{k+1} als Lösung von

$$\begin{aligned} & \Psi'((I_z^{k+1})^2 + \gamma((I_{xz}^{k+1})^2 + (I_{yz}^{k+1})^2)) \cdot (I_x^k I_z^{k+1} + \gamma(I_{xx}^k I_{xz}^{k+1} + I_{xy}^k I_{yz}^{k+1})) \\ & \quad - \alpha \operatorname{div} (\Psi'(|\nabla_3 u^{k+1}|^2 + |\nabla_3 v^{k+1}|^2) \nabla_3 u^{k+1}) = 0 \quad , \\ & \Psi'((I_z^{k+1})^2 + \gamma((I_{xz}^{k+1})^2 + (I_{yz}^{k+1})^2)) \cdot (I_y^k I_z^{k+1} + \gamma(I_{yy}^k I_{yz}^{k+1} + I_{xy}^k I_{xz}^{k+1})) \\ & \quad - \alpha \operatorname{div} (\Psi'(|\nabla_3 u^{k+1}|^2 + |\nabla_3 v^{k+1}|^2) \nabla_3 v^{k+1}) = 0 \quad . \end{aligned} \quad (4.52)$$

Sobald ein Fixpunkt \mathbf{w}^* erreicht ist, wechselt man zur nächsten feineren Skala und benutzt die Auflösung als Initialisierung für die Fixpunktiterationen. Man hat hierdurch ein komplett implizites Schema für den Glättungsterm und eine teilimplizites Schema für den Datenterm. Implizite Schemata werden benutzt, um eine höhere Stabilität und eine schnellere Konvergenz zu erzielen. Trotzdem ist das neue System nichtlinear, wegen der nichtlinearen Funktion Ψ' und der Symbole I_*^{k+1} . Um die Nichtlinearität in I_*^{k+1} zu entfernen, wird die Taylorreihenentwicklung erster Ordnung

$$\begin{aligned} I_z^{k+1} &\approx I_z^k + I_x^k du^k + I_y^k dv^k \quad , \\ I_{xz}^{k+1} &\approx I_{xz}^k + I_{xx}^k du^k + I_{xy}^k dv^k \quad , \\ I_{yz}^{k+1} &\approx I_{yz}^k + I_{xy}^k du^k + I_{yy}^k dv^k \quad , \end{aligned} \quad (4.53)$$

benutzt, wobei $u^{k+1} = u^k + du^k$ und $v^{k+1} = v^k + dv^k$. Deshalb teilt man die Unbekannten u^{k+1}, v^{k+1} auf, in die Lösungen des vorherigen Iterationsschritts u^k, v^k und in die unbekanntes Inkrementierungen du^k, dv^k . Zur besseren Lesbarkeit sei

$$\begin{aligned} (\Psi')_D^k &:= \Psi' \left((I_z^k + I_x^k du^k + I_y^k dv^k)^2 \right. \\ &\quad \left. + \gamma((I_{xz}^k + I_{xx}^k du^k + I_{xy}^k dv^k)^2 + (I_{yz}^k + I_{xy}^k du^k + I_{yy}^k dv^k)^2) \right) \quad , \\ (\Psi')_S^k &:= \Psi' \left(|\nabla_3(u^k + du^k)|^2 + |\nabla_3(v^k + dv^k)|^2 \right) \quad , \end{aligned} \quad (4.54)$$

wobei $(\Psi')_D^k$ als ein Robustheitsfaktor im Datenterm und $(\Psi')_S^k$ als Diffusionskoeffizient im Glättungsterm angesehen werden kann. Mit Hilfe dieser Schreibweise kann das Gleichungssystem (4.52) geschrieben werden als

$$\begin{aligned} \mathbf{0} &= (\Psi')_D^k \cdot (I_x^k(I_z^k + I_x^k du^k + I_y^k dv^k)) \\ &\quad + \gamma(\Psi')_D^k \cdot (I_{xx}^k(I_{xz}^k + I_{xx}^k du^k + I_{xy}^k dv^k) + (I_{xy}^k(I_{yz}^k + I_{xy}^k du^k + I_{yy}^k dv^k))) \\ &\quad - \alpha \operatorname{div}((\Psi')_S^k \nabla_3(u^k + du^k)) \quad . \end{aligned} \quad (4.55)$$

Die zweite Gleichung ergibt sich in analoger Weise. Dies ist immer noch ein nichtlineares Gleichungssystem für ein festes k , allerdings jetzt mit den Unbekannten du^k, dv^k . Da die einzige Nichtlinearität in Ψ' auftritt, und Ψ selbst als konvex angenommen wird, ist auch das resultierende Optimierungsproblem ein konvexes Problem und enthält deshalb ein einziges Minimum als Lösung. Um nun die übrigbleibende Nichtlinearität in Ψ' zu entfernen, wird eine zweite innere Fixpunktiterationsschleife angewandt.

Voraussetzungen für das OF-Modell nach [Brox u.a. 2004] sind:

- Für die Initialisierung gilt: $du^{k,0} := 0, dv^{k,0} := 0$.
- Die Iterationsvariablen sind $du^{k,l}, dv^{k,l}$.
- Die robustifizierten Faktoren: $(\Psi')_D^{k,l}$ und $(\Psi')_S^{k,l}$.
- Außerdem der Diffusionskoeffizient aus Gleichung (4.54).

Damit lässt sich das lineare Gleichungssystem aus $du^{k,l+1}, dv^{k,l+1}$ für die erste Gleichung bestimmen:

$$\begin{aligned} \mathbf{0} &= (\Psi')_D^{k,l} \cdot (I_x^k(I_z^k + I_x^k du^{k,l+1} + I_y^k dv^{k,l+1})) \\ &\quad + \gamma I_{xx}^k (I_{xz}^k + I_{xx}^k du^{k,l+1} + I_{xy}^k dv^{k,l+1}) + \gamma I_{xy}^k (I_{yz}^k + I_{xy}^k du^{k,l+1} + I_{yy}^k dv^{k,l+1}) \\ &\quad - \alpha \operatorname{div}((\Psi')_S^{k,l} \nabla_3(u^k + du^{k,l+1})) \quad . \end{aligned} \quad (4.56)$$

Benutzt man nun Standarddiskretisierungsverfahren für die Ableitungen, dann wird das resultierende dünn besetzte lineare Gleichungssystem mit bisher bekannten numerischen Methoden (z.B. Gauß–Seidel oder SOR) gelöst. Die Ausdrücke vom Typ $I(\mathbf{x} + \mathbf{w}^k)$ werden mittels linearer Interpolation berechnet.

Grob–nach–Fein Strategien steigern in zweierlei Hinsicht die Effizienz von Methoden zur Berechnung des OF:

1. Zur Steigerung der Konvergenzgeschwindigkeit.
2. Um die Wahrscheinlichkeit, in einem lokalen Minimum stecken zu bleiben, zu verringern.

In [Brox u.a. 2004] wird empfohlen, $\gamma = 0$ zu setzen. Dies vereinfacht das Modell (siehe auch [Mémin und Perez 1998]). Es wird also lediglich eine räumliche Glättung angewandt. Deshalb können die Gleichungen (4.55) umgeschrieben werden:

$$\begin{aligned} (\Psi')_D^k \nabla I^k (\nabla I^k)^T \begin{pmatrix} du^k \\ dv^k \end{pmatrix} - \alpha \begin{pmatrix} \operatorname{div} ((\Psi')_S^k \nabla (u^k + du^k)) \\ \operatorname{div} ((\Psi')_S^k \nabla (v^k + dv^k)) \end{pmatrix} \\ = -(\Psi')_D^k I_z^k \nabla I^k \quad . \end{aligned} \quad (4.57)$$

Für festes k entspricht dieses Gleichungssystem den Euler–Lagrange–Gleichungen (siehe [Mémin und Perez 1998]). Es werden lediglich die inkrementellen Größen du und dv zwischen dem ersten und dem verzerrten zweiten Bild bestimmt. Die selben inkrementellen Größen erscheinen in den äußeren Fixpunktiterationen, um die Nichtlinearität der Konstanzannahme zu lösen. Also implementiert die Grob–nach–Fein–Strategie die Minimierung einer nichtlinearen Konstanzannahme mittels einer Fixpunktiteration über \mathbf{w} . Da zu u und v am groben Bildraaster nur eine Inkrementierung du und dv auf dem feinen Bildraaster berechnet werden musste, hatte die Schätzung ein Ausmaß von weniger als einen Bildpunkt pro Bild, unabhängig der eigentlichen Verschiebung. Es besteht die Möglichkeit, die Grob–nach–Fein–Strategie zu benutzen, um größere Verschiebungen zu behandeln.

Ein zweiter Ansatz, um größere Verschiebungen zu behandeln, gelingt durch die Einführung der nichtlinearen Konstanzannahme, da hier große Verschiebungen von Anfang an erlaubt sind. Denn die Nichtlinearität wird durch eine multimodale Funktion realisiert. Dort ist die Grob–nach–Fein–Strategie notwendig. Die Arbeit von [Brox u.a. 2004] zeigt, dass beide Strategien äquivalent sind und kommt zu dem Schluss, dass die Grob–nach–Fein–Strategie als eindeutiger Löser multimodaler Probleme gesehen werden kann.

Die in der Arbeit von [Brox u.a. 2004] vorgeschlagene Lösung der Fixpunktgleichungen gelingt durch Anwendung des SOR–Verfahrens (siehe [Golub und Van Loan 1996]) und hat deshalb schlechte Konvergenzgeschwindigkeiten. Eine Möglichkeit liegt darin, bidirektionale Mehrgitterlöser einzuführen. Dabei entsteht das Problem, dass weitere Parameter in den Schätzprozess integriert werden müssen. Die bisher diskutierten Ansätze, basierend auf Mehrgitterlösern, optimieren auf einzelne Bildfolgen. Für eine sichere Mensch–Roboter–Kooperation benötigt man allerdings ein flexibler einsetzbares Verfahren.

Deshalb wird im Rahmen dieser Arbeit eine bisher nicht beachtete Klasse von numerischen Lösern für die bisherigen Gleichungen eingeführt [Graf und Wörn 2008b]. Der neu vorgeschlagene Löser ist die Methode der konjugierten Gradienten (bspw. [Golub und Van Loan 1996]). Es ist bekannt, dass die Methode der konjugierten Gradienten hervorragende Konvergenzeigenschaften besitzen. Das Problem bei der Vorgehensweise von [Brox u.a. 2004] ist, dass die Gleichungen punktweise gelöst werden. Man ist jedoch in der Lage, mit etwas Aufwand die

diskretisierten Gleichungen als lineares Gleichungssystem aufzuschreiben. Dies ermöglicht den Zugang der konjugierten Gradientenmethode als effiziente Löser. Dabei gelten folgende Abkürzungen für die in Gleichung (4.60) dargestellte Matrix, bezogen auf die Gleichung (4.56):

- **uapp**
Argumente von $u(i)$,
- **vapp**
Argumente von $v(i)$,
- **uvapp**
Argumente von $v(i)$,
- **vuapp**
Argumente von $u(i)$.

Die Werte werden anschließend kompakt in dem Vektor **ss** zusammengefasst:

$$\mathbf{ss} = \begin{pmatrix} -pdfs(1 : 2, 0 : 2)[pos] \\ -pdfs(0 : 2, 1 : 2)[pos] \\ u[x] \\ uv[x] \\ -pdfs(2 : 2, 1 : 2)[pos] \\ -pdfs(1 : 2, 2 : 2)[pos] \\ -pdfs(1 : 2, 0 : 2)[pos] \\ -pdfs(0 : 2, 1 : 2)[pos] \\ vu[x] \\ v[x] \\ -pdfs(2 : 2, 1 : 2)[pos] \\ -pdfs(1 : 2, 2 : 2)[pos] \\ \vdots \end{pmatrix} . \quad (4.58)$$

Damit lässt sich die Systemmatrix A ableiten, wobei $pdfs$ für den in Matrixform geschriebenen Glättungsterm steht.

Die bei der Umformulierung entstehende Abbildungsmatrix hat $(12 \cdot h \cdot b)^2$ Einträge, wobei h und b für die Höhe und Breite der einzelnen Bilder stehen. Die Matrix besitzt eine Bandstruktur und ist zudem dünnbesetzt. Es sind weniger als 1 ‰ der Einträge ungleich Null. Die finale Matrix nimmt die in Gleichung (4.59) dargestellte Struktur an.

Der letztlich benötigte Vektor **b** besteht aus $2 \cdot h \cdot w$ Einträgen und setzt sich, für $i \leq h \cdot w$, folgendermaßen zusammen:

$$\begin{aligned} b[2 \cdot i] &= -(E_D * (I_x^k * I_z^k + \gamma \cdot (I_{xx} \cdot I_{xz} + I_{xy} \cdot I_{yz})) - t1) \quad , \\ b[2 \cdot i + 1] &= -(E_D * (I_y^k * I_z^k + \gamma \cdot (I_{xy} \cdot I_{xz} + I_{xy} \cdot I_{yz})) - t2) \quad . \end{aligned} \quad (4.61)$$

$$A = \begin{pmatrix}
u[x] & uv[x] & s3[x] & 0 & \dots & 0 & s4[x] & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 \\
vu[x] & v[x] & 0 & s3[x] & 0 & \dots & 0 & s4[x] & 0 & 0 & 0 & 0 & 0 & \dots & 0 \\
s2[x+1] & 0 & u[x+1] & uv[x+1] & s3[x+1] & 0 & \dots & s4[x+1] & 0 & 0 & 0 & 0 & 0 & \dots & 0 \\
0 & s2[x+1] & vu[x+1] & v[x+1] & 0 & u[x+2] & uv[x+2] & s3[x+2] & 0 & 0 & 0 & 0 & 0 & \dots & 0 \\
0 & 0 & 0 & s2[x+2] & \dots & 0 \\
\vdots & \vdots \\
0 & 0 & \dots & 0 & s1[n-2] & 0 & \dots & s2[n-2] & vu[n-2] & v[n-2] & 0 & s3[n-2] & 0 & \dots & 0 \\
0 & 0 & \dots & 0 & 0 & s1[n-1] & 0 & 0 & s2[n-1] & uv[n-1] & v[n-1] & s3[n-1] & 0 & \dots & 0 \\
0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & 0 & u[n-1] & uv[n-1] & 0 & s3[n-1] & 0 & 0 \\
0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & 0 & s1[n] & 0 & 0 & s3[n] & 0 & 0 \\
0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & 0 & s1[n] & 0 & 0 & 0 & 0 & s3[n]
\end{pmatrix}, \tag{4.59}$$

mit

$$\begin{aligned}
s1 &= -pdfs(2:2, 1:2), \\
s2 &= -pdfs(1:2, 2:2), \\
s3 &= -pdfs(3:2, 2:2), \\
s4 &= -pdfs(2:2, 3:2), \\
u &= uapp, \\
v &= vapp, \\
uv &= uvapp, \\
vu &= vuapp.
\end{aligned} \tag{4.60}$$

Für die konstanten Terme gilt dabei:

$$\begin{aligned}
t1 &= pdfs(2 : 2, 1 : 2) \cdot (u(2, 1) - u(2, 2)) + \\
&\quad pdfs(2 : 2, 3 : 2) \cdot (u(2, 3) - u(2, 2)) + \\
&\quad pdfs(1 : 2, 2 : 2) \cdot (u(1, 2) - u(2, 2)) + \\
&\quad pdfs(3 : 2, 2 : 2) \cdot (u(3, 2) - u(2, 2)) \\
t2 &= pdfs(2 : 2, 1 : 2) \cdot (v(2, 1) - v(2, 2)) + \\
&\quad pdfs(2 : 2, 3 : 2) \cdot (v(2, 3) - v(2, 2)) + \\
&\quad pdfs(1 : 2, 2 : 2) \cdot (v(1, 2) - v(2, 2)) + \\
&\quad pdfs(3 : 2, 2 : 2) \cdot (v(3, 2) - v(2, 2)) \quad ,
\end{aligned} \tag{4.62}$$

wobei die Ränder mit Nullen repräsentiert werden. Anschließend lässt sich das lineare Gleichungssystem $A\mathbf{x} = \mathbf{b}$ durch Anwendung der Methode der konjugierten Gradienten effizient behandeln [Graf und Wörn 2008b] (siehe Algorithmus 1).

Algorithmus 1: Adaptierte Methode der konjugierten Gradienten.

```

1  $\mathbf{x}_0 = \mathbf{0}$ 
2  $\mathbf{r}_0 = \mathbf{b} - A\mathbf{x}$ 
3  $k = 0$ 
4 while( $r_k > \epsilon$ )
5    $k = k + 1$ 
6    $\alpha_k = \mathbf{r}_{k-1}^T \mathbf{r}_{k-1} / \mathbf{r}_{k-1}^T A \mathbf{r}_{k-1}$ 
7    $\mathbf{x}_k = \mathbf{x}_{k-1} + \alpha \mathbf{r}_{k-1}$ 
8    $\mathbf{r}_k = \mathbf{b} - A\mathbf{x}_k$ 
9 endwhile

```

4.3.2 Ein Verfahren zur Bestimmung von dynamischen Konturen

Die Dynamischen Konturen wurden in [Graf und Wörn 2008a] eingeführt. Die Methode wird im Rahmen dieser Arbeit verwendet, um sich bewegende Personen im Gesichtsfeld der Kamera durch eine Kontur zu verfolgen. Es wird gezeigt werden, wie damit die sichere Mensch-Roboter-Kooperation erfüllt werden kann.

Die grundsätzliche Idee liegt darin, eine umhüllende Kurve um den Menschen zu legen und ausgehend von dieser, den kürzesten Abstand zum Roboter zu berechnen. Die sich daraus ergebenden Informationen werden in den später folgenden Abschnitten diskutiert.

Hier soll nun ein neues Modell vorgestellt werden, welches die effiziente Berechnung des Optischen Flusses, basierend auf der Methode der konjugierten Gradienten [Graf und Wörn 2008b] in die Modellgleichung der aktiven Konturen [Xu und Prince 2000] integriert. Die Verwendung der in [Xu und Prince 2000] vorgeschlagenen Vorgehensweise ist dadurch begründet, dass durch die von einem Menschen erzeugten Bewegungen konkave Konturen der menschlichen Silhouette generieren. Die klassische Arbeit von [Kass u.a. 1988] hat u.a. in diesem Punkt größere Nachteile.

Eine weitere wichtige Entscheidung ist die Integration eines OF-Schätzers, welcher in der Lage ist, große Verschiebungen handhaben zu können. Auch deshalb wird der im vorhergehenden Abschnitt modifizierte Ansatz von [Brox u.a. 2004] in den Schätzprozess integriert. Denn bei einer Abtastrate von derzeit 15 Bildern pro Sekunde kann es zu Verschiebungen von Strukturen

von mehreren Bildpunkten kommen. Des Weiteren wirkt der adaptierte OF-Schätzer wegen der Gradientenkonstanzannahme dem Messrauschen der Kamera entgegen. Es wird folgendes Vorgehen zur Konturverfolgung vorgeschlagen:

1. Nach einer Vordergrundsegmentierung, wie sie bereits im Rahmen der Arbeit vorgestellt wurde, wird eine Ellipse an Ballungen von Bildpunkten des Vordergrundes angepasst.
2. Die Methode von [Xu und Prince 2000] wird anschließend solange wiederholt, bis sie konvergiert (siehe dazu die Abbildung 4.18).
3. Berechnung des OF-Feldes mit der Methode nach [Brox u.a. 2004] und den konjugierten Gradienten.
4. Es wird die Methode der aktiven Konturen, unter Berücksichtigung des OF-Feldes als zusätzliches externes Potentialfeld, verwendet.
5. Der Vorgang wird ab Punkt 3 wiederholt, solange bis die Kontur konvergiert.

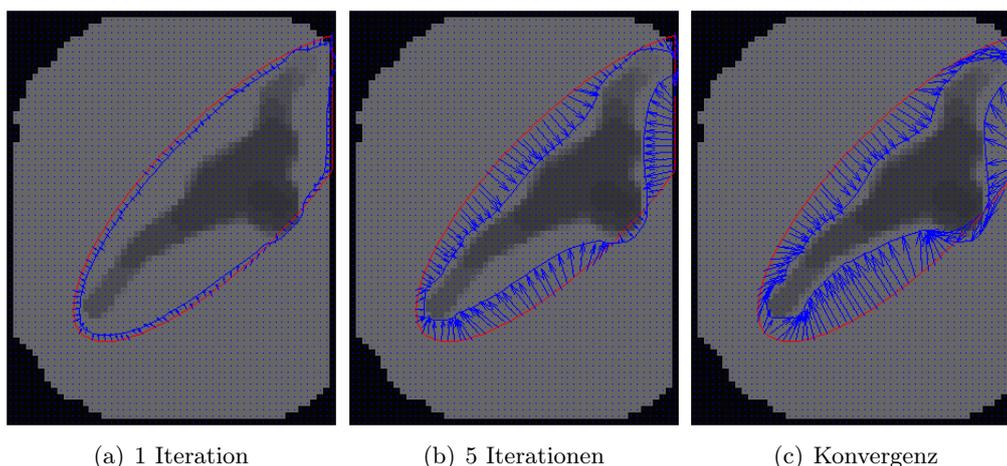


Abbildung 4.18: Anpassung einer parametrischen Kurve, ausgehend von einer Ellipse, an die Kontur einer menschlichen Silhouette.

Um die Vorgehensweise zu fundieren, werden im Folgenden einige technische Schritte erläutert. In [Kass u.a. 1988] wird ein Energiefunktional in Form eines Variationsansatzes vorgestellt:

$$E_C = \int_0^1 \frac{1}{2} \left(\alpha \|\mathbf{x}'(s)\|^2 + \beta \|\mathbf{x}''(s)\|^2 \right) + E_{ext}(\mathbf{x}(s)) ds \quad . \quad (4.63)$$

Dabei ist $x(s)$ eine parametrische Funktion, welche der Konturzug beschreibt. Der letzte Term dieser Gleichung E_{ext} realisiert eine externe Energiefunktion, welche üblicherweise durch signifikante Bildstrukturen, d.h. Kanten, Ecken oder sonstige Intensitätssprünge gebildet wird. Deshalb definieren [Xu und Prince 2000] folgende Energiefunktion:

$$E_{ext} = - \|K_\Sigma * \nabla I(x, y)\|^2 + \|K_\Sigma * I(x, y)\|^2 \quad . \quad (4.64)$$

Dabei wird durch I das Bild beschrieben und durch K_Σ eine Gaußfunktion mit der Kovarianzmatrix Σ .

Das Ziel ist die Minimierung des quadratischen Funktionals durch die Berechnung der zum Variationsansatz korrespondierenden Euler–Lagrange-Gleichung:

$$\alpha \mathbf{x}''(s) - \beta \mathbf{x}^{(4)}(s) - \nabla E_{ext} = \mathbf{0} \quad . \quad (4.65)$$

Um die Gleichung dynamisch zu gestalten, führt man eine Zeitvariable t ein:

$$\alpha \mathbf{x}''(s, t) - \beta \mathbf{x}^{(4)}(s, t) - \nabla E_{ext} = \mathbf{0} \quad . \quad (4.66)$$

Für hinreichend großes t gilt, dass $\mathbf{x}_t(s, t) = \mathbf{0}$ ist. Damit lässt sich folgende Gleichung herleiten:

$$\mathbf{x}_t(s, t) = \alpha \mathbf{x}''(s, t) - \beta \mathbf{x}^{(4)}(s, t) - \nabla E_{ext} \quad . \quad (4.67)$$

Bereits in [Kass u.a. 1988] wird gezeigt, wie durch Standardtechniken der Numerik diese Gleichung gelöst werden kann.

Da in der hier vorliegenden Arbeit der menschliche Körper aufgezeichnet und ausgewertet wird, entstehen Konkavitäten. Um dieses Problem zu lösen, wird in [Xu und Prince 2000] vorgeschlagen, ein Diffusionsschema für das Gradientenfeld zu verwenden, um dadurch eine externe Kraft einzuführen, welche in der Lage ist, die Konturlinie gegen konkave Bildstrukturen konvergieren zu lassen. Inspiriert durch die Arbeit von [Horn und Schunck 1981], wird folgendes Variationsproblem vorgeschlagen, welches gelöst werden muss:

$$E_{GV} = \int_{\Omega} \mu(u_x^2 + u_y^2 + v_x^2 + v_y^2) + \|\nabla f\| \|c - \nabla f\|^2 d\Omega \quad . \quad (4.68)$$

Mit Methoden der Variationsrechnung und durch Verwendung eines Gradientenabstiegsverfahrens ergibt sich das folgende Gleichungssystem:

$$u_t = \mu \Delta u - (u - f_x)(f_x^2 + f_y^2) \quad , \quad v_t = \mu \Delta v - (v - f_y)(f_x^2 + f_y^2) \quad . \quad (4.69)$$

Die Diffusion des Gradientenfeldes ist stark genug, um die Funktion $\mathbf{x}(s)$ gegen die Objektgrenzen konvergieren zu lassen. Ein Nachteil besteht darin, die Kontur in jedem Bild neu zu initialisieren, d.h. die Kontur muss in einer passenden Umgebung mit ausreichendem Abstand zum bisherigen Konturverlauf neu plaziert werden. Das bedeutet aber auch, dass kostbare Zeit bei der Berechnung verloren geht, da mit hohen Iterationszahlen gerechnet werden muss. Außerdem besteht die Gefahr, in einem lokalen Minima stecken zu bleiben. Berücksichtigt man, dass man es hier mit der Verfolgung eines menschlichen Körpers zu tun hat und dazu eine Kamera verwendet wird, die 15 Bilder pro Sekunde liefert, dann kann davon ausgegangen werden, dass sich zwischen zwei aufeinanderfolgenden Bildern Verschiebungen von lediglich wenigen Bildpunkten ergeben. Es liegt deshalb nahe, die neu zu initialisierende Kontur im Zeitschritt t in Abhängigkeit des Zeitschritts $t - 1$ zu platzieren. Bei dieser Vorgehensweise ergeben sich jedoch unerwünschte Divergenzen (siehe Abbildung 4.19)

Eine Lösung dieses Problem liegt in einer Neuformulierung des Variationsansatzes. Dazu soll der im Rahmen dieser Arbeit erweiterte OF-Schätzer als zusätzliche externe Kraft eingeführt werden (siehe auch [Graf und Wörn 2008a] und [Graf und Wörn 2008b]). Insbesondere die beiden Eigenschaften der Behandlung großer Verschiebungsraten durch Verwendung einer Grob–nach–Fein–Strategie und die Integration der Gradientenkonstanzannahme um den verrauschten Tiefenmesswerten entgegenzuwirken, führen zu sehr guten Resultaten. Einige

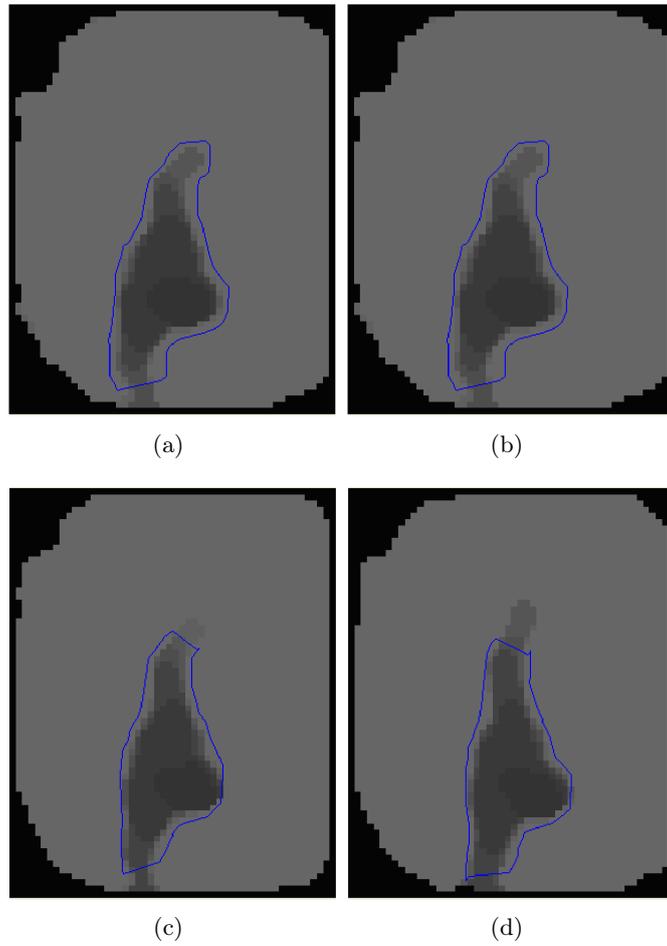


Abbildung 4.19: Divergender Fall, bei Initialisierung mit der Kontur des vorhergehenden Tiefenbildes ohne zusätzliche externe Kraft.

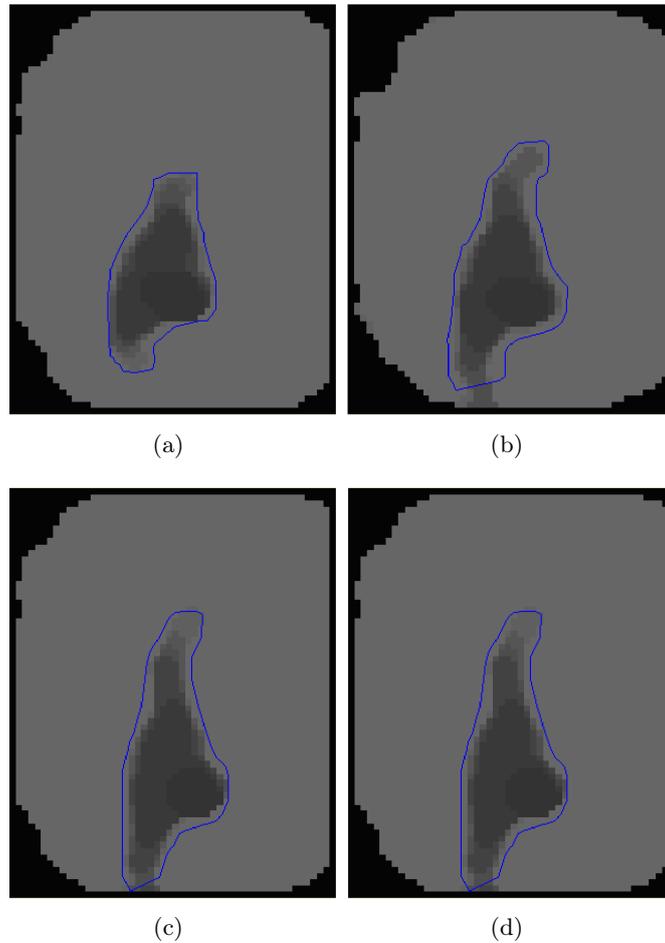


Abbildung 4.20: Konvergenter Fall, erzielt mit der in [Graf und Wörn 2008a] entwickelten Methode.

ausgewählte Beispiele einer ausgewerteten Tiefenbildfolge sieht man in den Abbildungen 4.20. Durch Einführung der Methode der konjugierten Gradienten ist es möglich, die Dynamischen Konturen in Echtzeit, d.h. mit mehr als 15 Bilder pro Sekunde, zu berechnen.

Die Prädiktion der Kontur für das nachfolgende Bild kann zu einer Konvergenz gegen ein lokales Minimum führen, was sich in einer divergierenden Kontur, bzgl. der Silhouette des Menschen, ausdrückt (siehe Abbildung 4.19). Deshalb wird in [Graf und Wörn 2008a] eine Erweiterung des Variationsansatzes vorgeschlagen:

$$E_{C'} = \int_0^1 \frac{1}{2} \left(\alpha \|\mathbf{x}'(s)\|^2 + \beta \|\mathbf{x}''(s)\|^2 \right) + E_{ext'}(\mathbf{x}(s)) ds \quad . \quad (4.70)$$

Die Änderung wird deutlich, wenn man die korrespondierende Gradientenabstiegsgleichung betrachtet:

$$\mathbf{x}_t(s, t) = \alpha \mathbf{x}''(s, t) - \beta \mathbf{x}^{(4)}(s, t) - \chi_i(t) E_{OF} + (1 - \chi_i(t)) (-\nabla E_{GV}) \quad . \quad (4.71)$$

In dieser Gleichung sind der Optische Fluss und das Gradientenvektorfeld durch eine konvexe Kombination verknüpft. Es gilt die Funktion $\chi_i \in [0, 1]$, mit $i \in \{1, 2\}$. Dabei ist χ eine von der Zeit t abhängige Funktion. Es wird vorgeschlagen, eine Treppenfunktion $\chi_i(t - \sigma)$ dafür einzusetzen. Eine andere Alternative ist die Verwendung einer Exponentialfunktion, wobei der zweite Parameter σ , hier den Übergang zwischen niedrigen Funktionswerten zu hohen Funktionswerten kontrolliert:

$$\chi_2(t - \sigma) = \exp\left\{-\frac{t^2}{\sigma^2}\right\} \quad . \quad (4.72)$$

Im weiteren Verlauf wird die Angabe des Parameters σ nicht explizit in den Gleichungen angegeben. Die Abbildung 4.21 illustriert exemplarisch beide Funktionen.

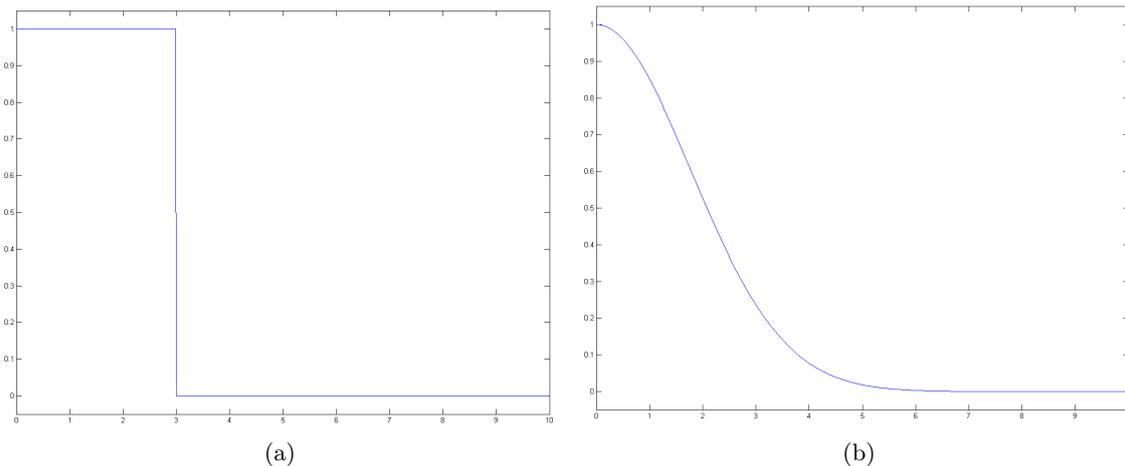


Abbildung 4.21: Visualisierung der Treppenfunktion (linke Abbildung) und der Exponentialfunktion mit $\sigma = 0.25$ (rechte Abbildung).

Die Idee bei Verwendung der hier vorgestellten Gleichung ist, dass in den ersten Iterationsschritten der Gradientenabstiegsgleichung (4.71) das externe Kraft abhängig vom OF-Feld

E_{OF} über das Gradientenvektorfeld E_{GV} dominiert. Das bedeutet, dass sich die Kontur zunächst entlang des OF-Feldes bewegt. Durch diese Vorgehensweise wird die Reinitialisierung der Kontur überflüssig, zumindest so, wie sie am Anfang dieses Abschnittes eingeführt wurde. Nach einigen Iterationsschritten, ändert die Funktion χ ihren Wert und veranlasst dadurch, dass das Gradientenvektorfeld stärker gewichtet wird. Damit kann die Kontur gegen die durch die Abbildung des menschlichen Körpers erzeugten Bildpunkte konvergieren. Die Abbildung 4.22 zeigt anhand ausgewählter Bilder die Vorgehensweise.

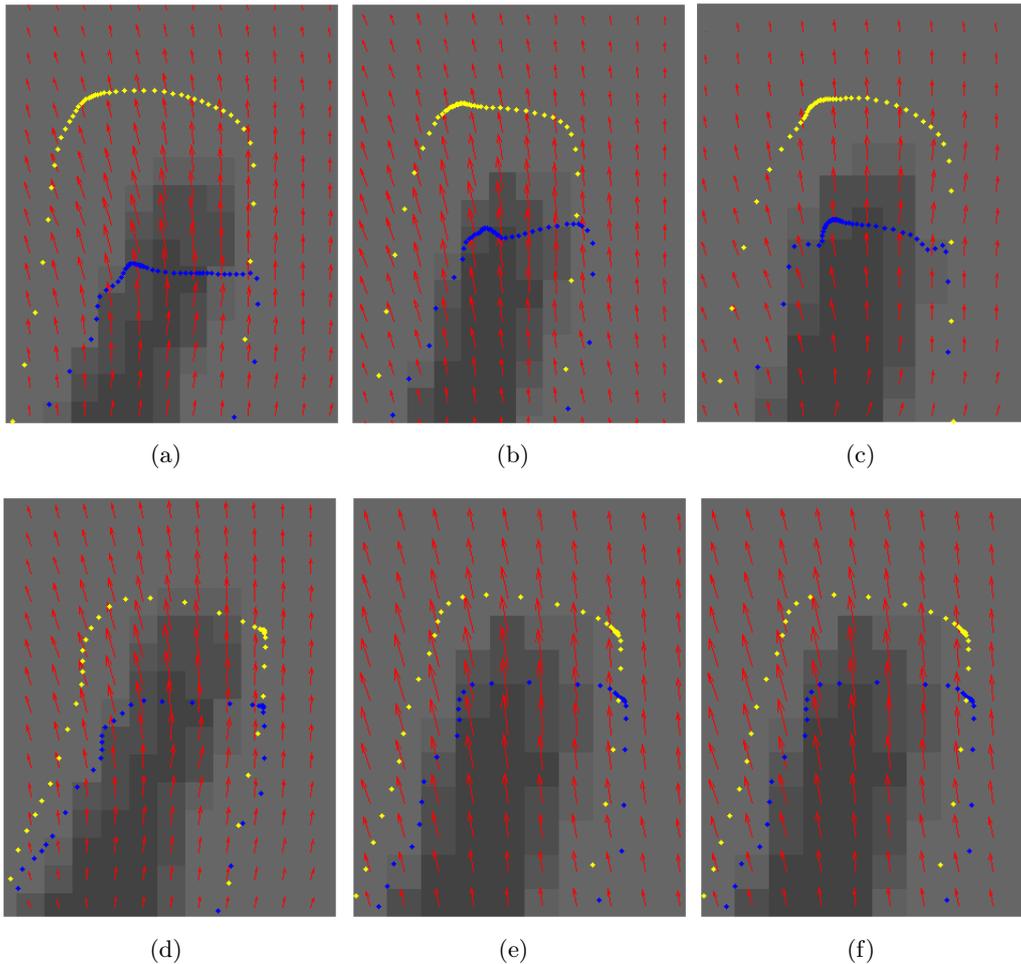


Abbildung 4.22: Ausgewählte Bilder, erzeugt durch die Bewegungen eines menschlichen Arms. Die blauen Punkte repräsentieren Knotenpunkte der Kontur des vorhergehenden Zeitschrittes. Die gelben Punkte repräsentieren korrespondierende Knotenpunkte der Kontur nach Anwendung der Gradientenabstiegsleichung. (Oben) Zwei Iterationen, d.h. $\chi_1(t-2)$. (Unten) Konvergenzfall.

In der Abbildung 4.22 zeigt sich, dass die alleinige Verwendung des OF als externe Kraft die Kontur zur Übersegmentierung tendieren lässt. Die alleinige Verwendung des Gradientenvektorfeldes führt nicht zu befriedigenden Ergebnisse und kann gravierende Folgen für die Sicherheit des Menschen haben. Die Berücksichtigung beider externen Kräfte und deren Kopplung durch eine konvexe Kombination, hat bei den durchgeführten Experimenten zu überzeugenden

Ergebnissen geführt. In Abbildung 4.23 werden einige ausgewählte Bilder gezeigt, die in einem komplexen Bewegungsablauf zu positiven Resultaten geführt hat. Die Bilder werden außerdem im Vorfeld mit dem orientierten Medianfilter vorverarbeitet, was die Güte der Schätzung zusätzlich stabilisierte.

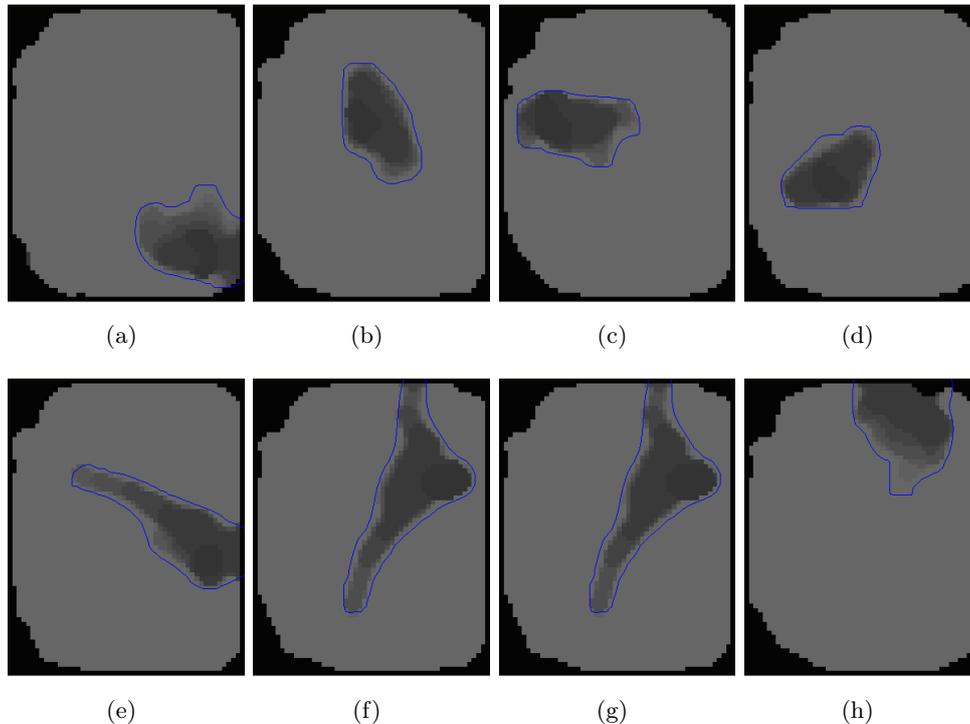


Abbildung 4.23: Ausgewählte Resultate aus einer Bildfolge mit einer sich in der Szene bewegendem Person.

Eine weiteres Beispiel für ein erfolgreiches Verfolgen einer sich bewegendem Person ist in den Abbildungen 4.24 zu sehen.

Zusätzlich ist eine Abstandsberechnung zur Werkzeugspitze des Roboters implementiert. Die Abstandsberechnung basiert auf einem einfachen Optimierungsansatz, welcher durch eine Grob–nach–Fein–Strategie realisiert ist. Es werden zunächst eine Auswahl von am kürzesten entfernten Punkte der aktiven Kontur zur Werkzeugspitze identifiziert. Anschließend wird aus den Paarkombinationen der Auswahl der kürzeste Abstand ermittelt.

Dieser einfache Demonstrator stellte eine erste Realisierung des in 3.1 vorgestellten Konzepts einer sicheren Mensch–Roboter–Kooperation dar und wurde in [Graf und Wörn 2008c] publiziert. Ausgehend von den bisher gewonnenen Informationen kann lediglich ein Vergleich mit dem bisherigen Stand der Forschung im Bereich der industriellen Robotik erzielt werden. Die systematische Gewinnung von Konturinformationen aus einem systematischen Variationsansatz ist gleichermaßen theoretisch und praktisch motiviert.

4.4 Ein OF–Schätzverfahren mit anisotropem Datenterm

Bei der Verschiebung eines Punktes in der Szene, ergibt sich eine entsprechende Verschiebung des korrespondierenden Punktes in der Bildebene (Abbildung 4.25). Werden die Verschiebun-

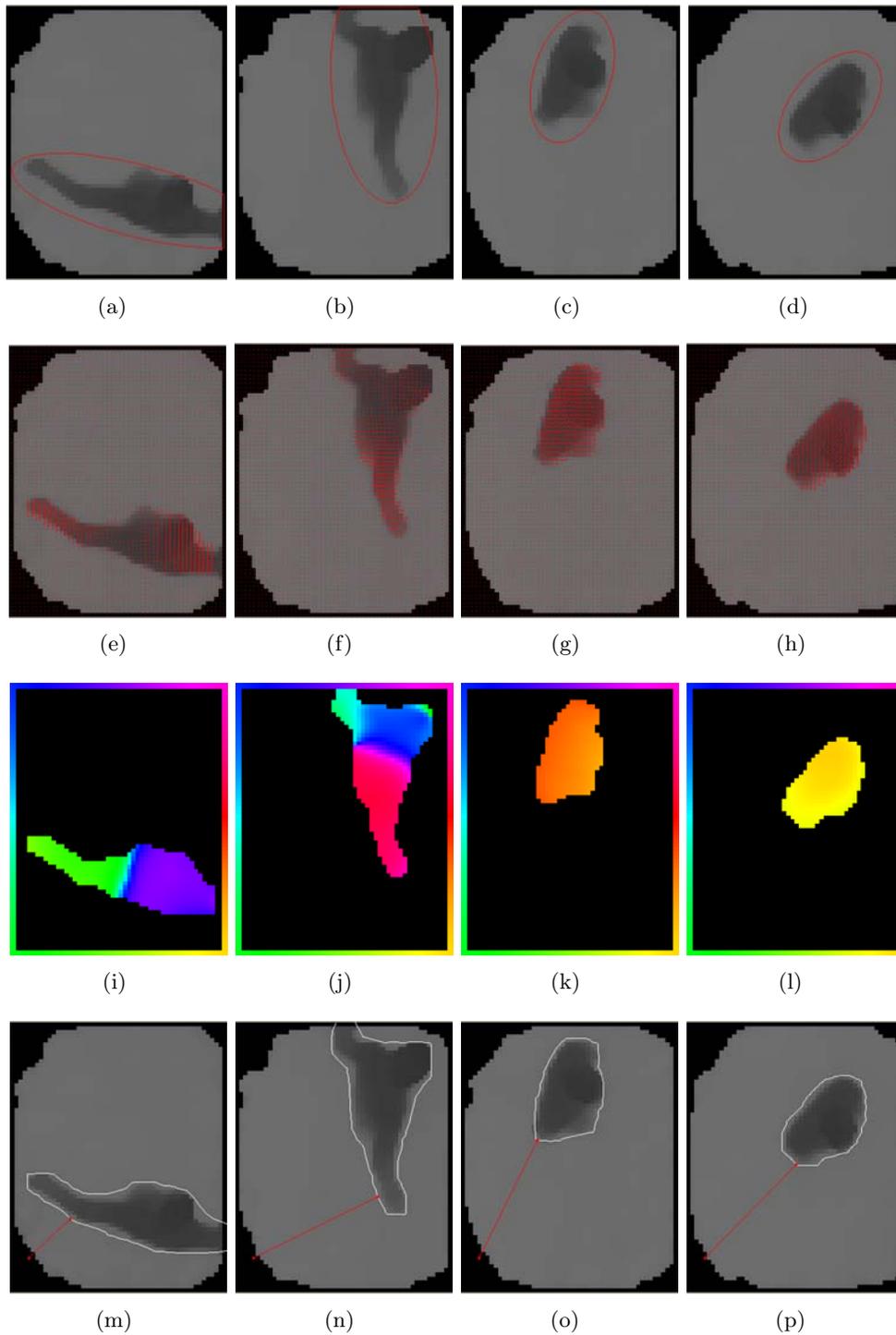


Abbildung 4.24: Ausgewählte Resultate aus einer Bildfolge mit einer sich in der Szene bewegendem Person.

gen durch Vektoren beschrieben, entsteht ein Vektorfeld, welches als Optisches Fluss-Feld bezeichnet wird. Durch die Berechnung des OFF können Informationen über die Bewegungsrichtung und die Verschiebungsrate von Szeneentitäten berechnet werden. Der so berechnete Wert entspricht nicht der tatsächlichen Bewegung in der Szene, sondern jeweils der Abbildung dieser Bewegung.

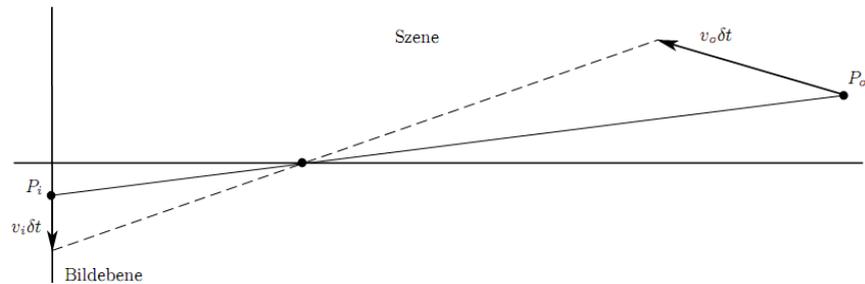


Abbildung 4.25: Projektion der Verschiebung eines Punktes in der Szene auf die Bildebene.

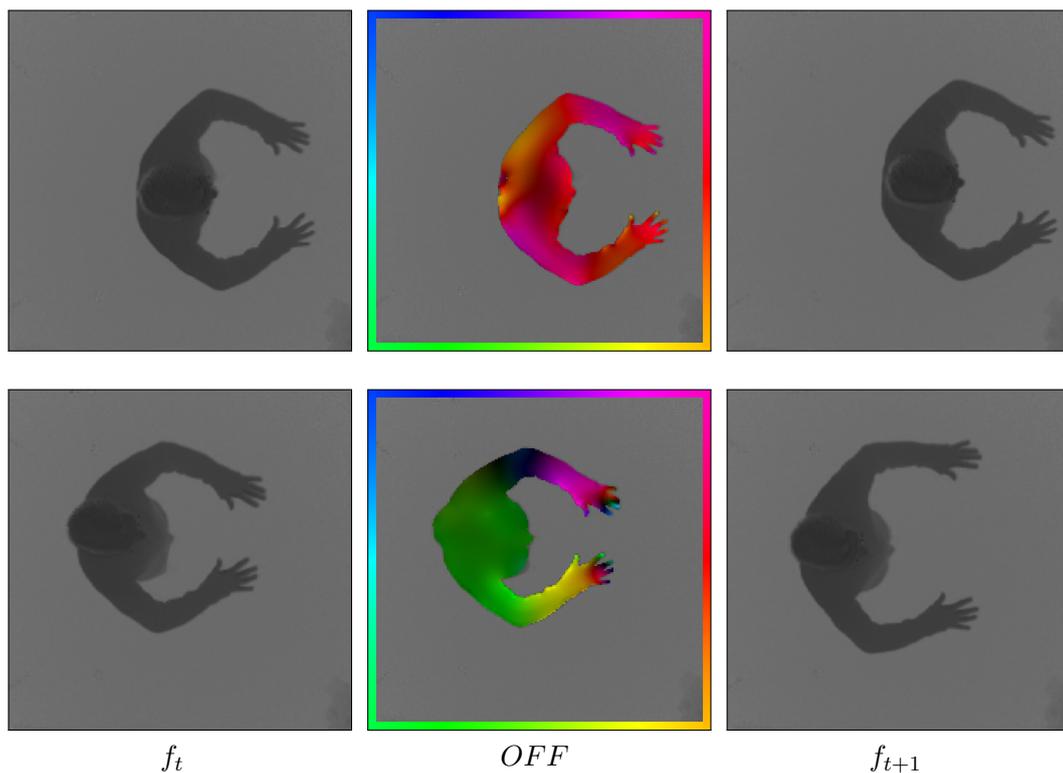


Abbildung 4.26: (Oben): Homogene Bewegung des Oberkörpers und der Arme. (Unten): Bewegung des Oberkörpers und der Arme in unterschiedliche Richtungen.

Möchte man Informationen über Bewegungen in der Szene gewinnen, so wird man mit zwei Problemen konfrontiert:

1. Die Berechnung des OF ist lediglich eine Schätzung des scheinbaren Verschiebungsfeldes von durch Intensitäten gebildeten Strukturen. Die Güte dieser Schätzung hängt vor

allem von der Qualität des gemessenen Signals und der ausgewählten Methode zur OF-Berechnung ab.

2. Die Ermittlung des OF kann je nach gewählter Methode und gewünschter Exaktheit unterschiedliche Berechnungszeit beanspruchen und damit die Anforderungen einer Echtzeitanwendung nicht erfüllen. Deshalb ist der Einsatz schneller Hardware oder die Anwendung effizienter numerischer Verfahren notwendig.

In den folgenden Abschnitten werden beide Probleme behandelt und weitere neuartige Lösungsvorschläge gemacht. Es wird zum einen eine Methode vorgestellt, welche eine Erweiterung bzw. Verbesserung eines bereits existierenden Ansatzes darstellt. Und zum anderen werden für die Erweiterung spezielle Hardware und effiziente Numerik verwandt, um Echtzeitbedingungen zu erfüllen.

4.4.1 Grundlagen

Im den folgenden Abschnitten werden zuerst die Grundlagen der kombiniert lokal-globalen Verfahren (CLG) erläutert. Anschließend wird eine Weiterentwicklung dieser Klasse von OF-Schätzverfahren vorgestellt, den sog. erweiterten lokal-globalen Verfahren (XCLG) [Graf u.a. 2010a].

CLG-Methode

Die Methode von [Lucas und Kanade 1981] gehört zu den Differentialverfahren. Es wird postuliert, dass die Intensitätswerte einer Entität konstant bleibt. Da dies auf ein schlecht gestelltes Problem führt (Aperturproblem), müssen zusätzliche Beschränkungen festgelegt werden. [Lucas und Kanade 1981] machen deshalb die Annahme, dass der OF in einer lokalen Umgebung ω konstant ist. Die mathematische Beschreibung liefert die folgende Funktion:

$$E_{LK}(\mathbf{u}) = K_{\omega} * ((f_x u_1 + f_y u_2 + f_t)^2) \quad . \quad (4.73)$$

Der Ausdruck K_{ω} bezeichnet eine zweidimensionale Gauß-Funktion mit der Standardabweichung ω . Es wird der Vektor gesucht, der in einem Punkt des Bildes, die Funktion (4.73) minimiert. Mit der Gauß-Funktion als Faltungsterm, erhält man eine gewichtete Integration der quadratischen Abweichung der Grauwertkonstanzannahme für die Nachbarschaft ω . Eine Lösung \mathbf{u} ist deshalb ein Minimum der Funktion (4.73). Damit gilt, dass die partiellen Ableitungen $\partial_{u_1} E_{LK}$ und $\partial_{u_2} E_{LK}$ gleich 0 sind. Durch Ableiten und Nullsetzen der Gleichung erhält man nach algebraischer Umformung das folgende lineare Gleichungssystem (LGS) [Bruhn u.a. 2005a]:

$$\begin{pmatrix} K_{\omega} * (f_x^2) & K_{\omega} * (f_x f_y) \\ K_{\omega} * (f_x f_y) & K_{\omega} * (f_y^2) \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = \begin{pmatrix} -K_{\omega} * (f_x f_t) \\ -K_{\omega} * (f_y f_t) \end{pmatrix} \quad . \quad (4.74)$$

Eine eindeutige Lösung existiert, falls die Systemmatrix invertierbar ist. In Gebieten, in welchen die Grauwerte konstant sind ist dies nicht möglich, da dort der Grauwertgradient bzw. die Bildinformation verschwindet. Das führt zu ausgedünnten Fluss-Feldern, eine generelle Eigenschaft von lokalen Verfahren. Um dies zu kompensieren ist ein weiterer Interpolationschritt notwendig. Dabei ist der lokale Ansatz aufgrund der Glättung des OFF in einer lokalen Nachbarschaft robust gegenüber Störungen.

Der Ansatz von [Horn und Schunck 1981] gehört zu den globalen Verfahren und liefert deshalb im Vergleich zu [Lucas und Kanade 1981] ein vollständiges OFF. Ebenso wie für den Ansatz von [Lucas und Kanade 1981] gilt die linearisierte OF-Beschränkungsgleichung, es wird also auch die Gültigkeit der Grauwertkonstanzannahme vorausgesetzt. Gesucht wird eine Funktion $\mathbf{u}(x_1, x_2, x_3) = (u_1(x_1, x_2, x_3), u_2(x_1, x_2, x_3))^T$, welche das Energiefunktional

$$E_{HS}(\mathbf{u}) = \int_{\Omega_2} \underbrace{(f_x u_1 + f_y u_2 + f_t)^2}_D + \alpha \underbrace{(|\nabla_2 u_1|^2 + |\nabla_2 u_2|^2)}_G dx \quad (4.75)$$

minimiert. Der Ausdruck D wird dabei als Datenterm und G als Glättungsterm bezeichnet. Die Grauwertkonstanzannahme steckt im Datenterm, wobei Abweichungen quadratisch bestraft werden. Dieser reicht allerdings nicht aus, um den OF zu bestimmen. Wegen des Aperturproblems ist die Lösung nicht eindeutig bestimmbar. In Bereichen, in denen keine strukturelle Information im Bild vorliegt, verschwindet der Gradient, was dazu führt, dass keine Aussage über die Lösung gemacht werden kann. Deshalb wird mit Hilfe des Glattheitsterms eine weitere Beschränkung eingeführt, welche durch den freien Parameter α gewichtet werden kann. Für den Ansatz von [Horn und Schunck 1981] werden dabei starke Gradienten des OF quadratisch bestraft, was im mathematischen Sinne einer Glattheitsforderung des OF entspricht. Diese Beschränkung führt an Stellen, in denen der Grauwertgradient nicht verschwindet dazu, dass eine eindeutige Lösung gefunden wird. An Stellen in denen der Grauwertgradient verschwindet, wird durch die globale Minimierung, Information von außen eingebracht (Füll-effekt), weshalb eine Schätzung des OFF in allen Bildpositionen prinzipiell möglich ist. In [Barron u.a. 1994] wird darauf hingewiesen, dass differentielle, globale Methoden weniger robust gegenüber Störungen sind als lokale Verfahren. Ein hybrider Ansatz ist deshalb eine Möglichkeit, von den Vorteilen beider Methoden zu profitieren, ohne die Nachteile in Kauf nehmen zu müssen.

Für den weiteren Verlauf werden folgende Notationen eingeführt:

$$\mathbf{u} := (u_1, u_2, 1)^T \quad (4.76)$$

$$|\nabla \mathbf{u}|^2 := (|\nabla_2 u_1|^2 + |\nabla_2 u_2|^2) \quad (4.77)$$

$$\nabla_3 f := (f_x, f_y, f_t)^T \quad (4.78)$$

$$J_\omega(\nabla_3 f) := K_\omega * (\nabla_3 f \nabla_3 f^T) \quad (4.79)$$

Mit der neu eingeführten Notation können der globale und der lokale Ansatz in einer einheitlichen Formulierung dargestellt und in einem gemeinsamen Energiefunktional zusammengefasst werden. Für die Lukas–Kanade–Methode ergibt sich die Funktion (4.73) zu:

$$E_{LK} = \mathbf{u}^T J_\omega(\nabla_3 f) \mathbf{u} \quad (4.80)$$

Das Energiefunktional (4.75) der Horn–Schunck–Methode kann in folgender Form geschrieben werden:

$$E_{HS}(\mathbf{u}) = \int_{\Omega_2} (\mathbf{u}^T J_0(\nabla_3 f) \mathbf{u} + \alpha (|\nabla_2 u_1|^2 + |\nabla_2 u_2|^2)) dx \quad (4.81)$$

Durch die neue Schreibweise, kann das Horn–Schunck–Funktional auf die hybride Form erweitert werden [Bruhn 2006]:

$$E_{CLG}(\mathbf{u}) = \int_{\Omega_2} (\mathbf{u}^T J_\omega(\nabla_3 f) \mathbf{u} + \alpha(|\nabla_2 u_1|^2 + |\nabla_2 u_2|^2)) \, d\mathbf{x} \quad . \quad (4.82)$$

Wobei lediglich die Matrix $J_0(\nabla_3 f)$ durch den Strukturtensor $J_\omega(\nabla_3 f)$ ersetzt wurde.

4.4.2 Die XCLG–Methode

Für die CLG–Methode [Bruhn u.a. 2005a] wird ein lokaler und ein globaler Ansatz zur OF–Schätzung in einem Variationsansatz vereinigt. Die XCLG–Methode erweitert diesen Ansatz durch einen punktweise angepassten Datenterm:

$$E_{XCLG}(\mathbf{u}) = \int_{\Omega_2} \underbrace{\mathbf{u}^T \tilde{J}(\nabla_3 f) \mathbf{u}}_D + \alpha \underbrace{(|\nabla_2 u_1|^2 + |\nabla_2 u_2|^2)}_G \, d\mathbf{x} \quad . \quad (4.83)$$

Der lokale Bereich der Grauwertkonstanz wird durch den räumlich–zeitlichen–Strukturtensor

$$\tilde{J}(\nabla_3 f) = \tilde{K} * (\nabla_3 f \nabla_3 f^T) = K_\omega^\alpha * (\nabla_3 f \nabla_3 f^T) \quad (4.84)$$

für jeden Bildpunkt festgelegt. Die Faltung bzw. Gewichtungsfunktion K_ω^α ist eine anisotrope Gauß–Funktion. Die Komponenten des Vektors $\omega = (\omega_1, \omega_2)^T$ bezeichnen die Standardabweichungen. Der Parameter α gibt die Orientierung der Funktion an.

Eine lokale Schätzung $\mathbf{u} = (u_1, u_2, 1)$ des OF soll den Datenterm D (4.83) minimieren. Deshalb kann die Forderung aufgestellt werden, dass die partiellen Ableitungen nach u_1 und u_2 verschwinden. Dies führt auf das folgende Gleichungssystem:

$$\underbrace{\begin{pmatrix} \tilde{K} * (f_x f_x) & \tilde{K} * (f_x f_y) \\ \tilde{K} * (f_x f_y) & \tilde{K} * (f_y f_y) \end{pmatrix}}_{=: J_\omega(\nabla_2 f)} \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = \begin{pmatrix} -\tilde{K} * (f_x f_t) \\ -\tilde{K} * (f_y f_t) \end{pmatrix} \quad . \quad (4.85)$$

Die 2×2 Matrix $J_\omega(\nabla_2 f)$ der linken Seite wird als räumlicher Strukturtensor bezeichnet. Nach der Eigenwertzerlegung

$$\begin{pmatrix} \tilde{K} * (f_x f_x) & \tilde{K} * (f_x f_y) \\ \tilde{K} * (f_x f_y) & \tilde{K} * (f_y f_y) \end{pmatrix} = (\mathbf{v}_1, \mathbf{v}_2) \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} (\mathbf{v}_1, \mathbf{v}_2)^T \quad , \quad (4.86)$$

des Strukturtenors können Aussagen für die lokale Lösung \mathbf{u} getroffen werden. Es werden drei Fälle in Abhängigkeit des Rangs bzw. der Eigenwerte des Strukturtenors unterschieden:

1. Der Strukturtensor besitzt den vollen Rang und ist deshalb invertierbar. Die beiden Eigenwerte λ_1 und λ_2 sind ungleich Null. Das bedeutet, dass mindestens ein Paar $(\nabla_2 f(x_{1_1}, x_{2_1}), \nabla_2 f(x_{1_2}, x_{2_2}))$ linear unabhängig ist. Insgesamt folgt daraus, dass eine eindeutige lokale Lösung \mathbf{u} existiert.
2. Der Strukturtensor hat den Rang 1 und ist deshalb nicht invertierbar. Einer der beiden Eigenwerte ist 0. Alle Gradienten $\nabla_2 f$ sind linear abhängig sind. Mindestens ein Gradient ist $\nabla_2 f \neq \mathbf{0}$, so dass Informationen in Form von signifikanter Strukturierung in dem Signal verfügbar sind. Es ergeben sich unendlich viele lokale Lösungen, die alle linear abhängig sind und auf einer Geraden zu Liegen kommen.

3. Der Strukturtensor hat den Rang 0 und ist deshalb nicht invertierbar. Die Eigenwerte und alle Gradienten in der lokalen Umgebung sind Null. Damit ist keine Informationen vorhanden und keine Aussage über die lokale Lösung möglich.

Angepasster Strukturtensor

In [Bruhn u.a. 2005a] wird für die Erstellung des räumlichen-zeitlichen-Strukturtenors $J_\omega(\nabla_3 f) = K_\omega * (\nabla_3 f \nabla_3 f^T)$ eine isotrope Gauß-Funktion K_ω mit der Standardabweichung ω verwendet. Die Standardabweichung wird als ein fester Parameter für alle Positionen im Bild gewählt, spezielle lokale Eigenschaften werden nicht betrachtet. Eine hier vorgeschlagene Verbesserung berücksichtigt diese Tatsache, indem die Kovarianzmatrix der Funktion \tilde{K} in geeigneter Weise an die zugrundeliegende Struktur adaptiert wird (siehe Abbildung 4.27). Wichtige lokale Eigenschaften für die OF-Schätzung sind bspw. Ecken und Kanten. Diese enthalten die Informationen über die Struktur von Objekten und Bereichen und ermöglichen so erst eine Zuordnung von korrespondierenden Punkten in aufeinanderfolgende Bilder.

Ermittlung von Ecken und Kanten

Damit eine Anpassung an diese Strukturen möglich wird, müssen in einem ersten Schritt die Bereiche gefunden werden, welche jeweils zu einer Kante bzw. zu einer Ecke gehören.

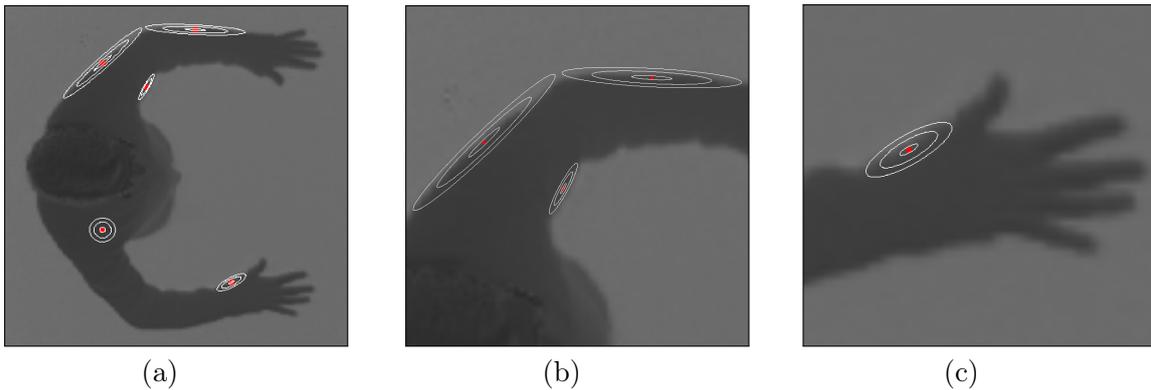


Abbildung 4.27: (a) – (c): An lokale Strukturen angepasste Faltungsfunktionen \tilde{K} .

Lokale Strukturanalyse Um Informationen über die lokalen strukturellen Eigenschaften zu erhalten, wird für jeden Bildpunkt der räumliche-ST $J_\sigma(\nabla_2 f)$ erstellt und eine Eigenwertzerlegung durchgeführt:

$$\begin{pmatrix} K_\sigma * (f_x f_x) & K_\sigma * (f_x f_y) \\ K_\sigma * (f_x f_y) & K_\sigma * (f_y f_y) \end{pmatrix} = (\mathbf{v}_1, \mathbf{v}_2) \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} (\mathbf{v}_1, \mathbf{v}_2)^T \quad . \quad (4.87)$$

Da der ST eine symmetrische positiv semidefinite 2×2 Matrix ist, existiert stets eine Eigenwertzerlegung. Die Eigenvektoren bilden eine orthonormale Basis. Eine anschließende Analyse der zugehörigen Eigenwerte λ_1 und λ_2 liefert Informationen über die Struktur, in einem durch σ festgelegten Bereich. Es werden drei Fälle unterschieden, wobei o.B.d.A. angenommen wird, dass λ_1 jeweils der Größere der beiden Eigenwerte ist:

1. $\lambda_1 > \lambda_2 \gg 0$: Es gibt Grauwertgradienten, die linear unabhängig sind. In diesem Fall gehört der Bildpunkt zu einer Ecke.
2. $\lambda_1 \gg \lambda_2 \approx 0$: Die Grauwertgradienten sind linear abhängig. Deshalb gehört der Punkt zu einer Kante.
3. $\lambda_1 \approx \lambda_2 \approx 0$: Alle Grauwertgradienten verschwinden bzw. liegen unterhalb eines geringen Schwellwertes. Der Punkt liegt in einem homogenen Bereich.

Der Eigenvektor \mathbf{v}_1 gibt die Richtung der größten Intensitätsvariationen, der Eigenvektor \mathbf{v}_2 gibt die Richtung der evtl. kleineren Intensitätsvariation an. Dadurch ist es möglich, die Zugehörigkeit einzelner Bildpunkte zu bestimmen.

Ballungsanalyse

Eine Zugehörigkeit einzelner benachbarter Bildpunkte zu einer gemeinsamen Struktur kann quantitativ mit Hilfe eines Ähnlichkeitsmaßes bzw. Distanzmaßes bestimmt werden. Eine Maßfunktion hängt hier von den Eigenwerten bzw. den korrespondierenden Eigenvektoren ab:

$$d(p_1, p_2) = \sqrt{(p_{1x} - p_{2x})^2 + (p_{1y} - p_{2y})^2 + w_\alpha(p_{1\alpha} - p_{2\alpha})^2} \quad . \quad (4.88)$$

Die Ballungsanalyse findet aufgrund der Beschaffenheit von $d(p_1, p_2)$ in einem Euklidischen Raum (3D) statt. Dabei sind p_1 und p_2 Punkte in der Bilddomäne und p_{n_x} bzw. p_{n_y} bezeichnen die Komponenten. Der Winkel p_{n_α} des Punktes p_n ergibt sich aus:

$$p_{n_\alpha} = \text{atan2}(\mathbf{v}_{1_y}, \mathbf{v}_{1_x}) \quad . \quad (4.89)$$

Die Winkeldifferenz wird mit w_α gewichtet. Dadurch besteht die Möglichkeit, kleine Variationen der Orientierung innerhalb der Struktur zuzulassen (Abbildung 4.28).

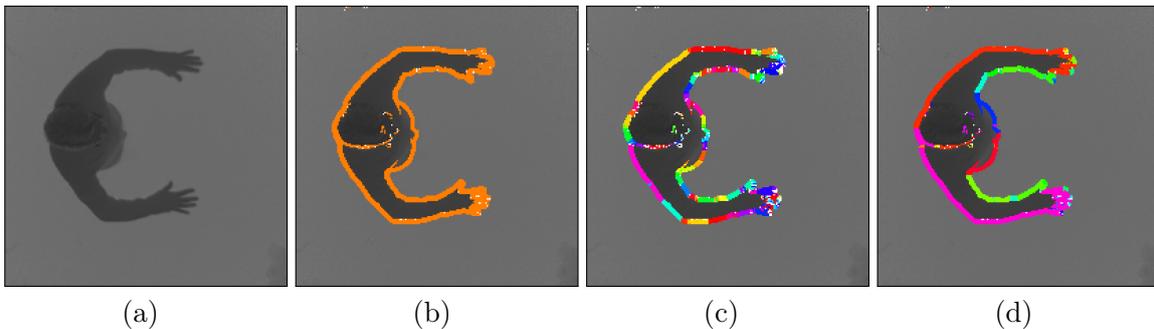


Abbildung 4.28: (b) Punkte in orange beschreiben initiale Punktmenge des DBSCAN Algorithmus. (c) Starke Gewichtung der Winkeldifferenz. (d) Der Gewichtungsparameter ω_α wurde sehr klein gewählt.

Durch das Distanzmaß kann für jeden Punkt die Zugehörigkeit zur gleichen Struktur bestimmt werden. Dazu wird ein Ballungsverfahren angewandt, welches Punktmengen findet, die eine einheitliche Struktur bzgl. des Distanzmaßes haben. Es sollen Kanten und Ecken gefunden werden. Die Menge der Punkte, wird durch ein Kriterium getestet, welches die Bildpunkte aus der Analyse ausschließt, die zu homogenen Bereichen gehören. Die Eigenwerte λ_1 und λ_2 des räumlichen-ST müssen folgende Voraussetzung erfüllen:

$$(\lambda_1 > t_1) \quad \vee \quad ((\lambda_1 > t_2) \wedge (\lambda_2 > t_2)) \quad . \quad (4.90)$$

Durch die Schwellwerte t_1 und t_2 werden zu geringe Intensitätsvariationen von der Analyse ausgeschlossen (Abbildung 4.29).

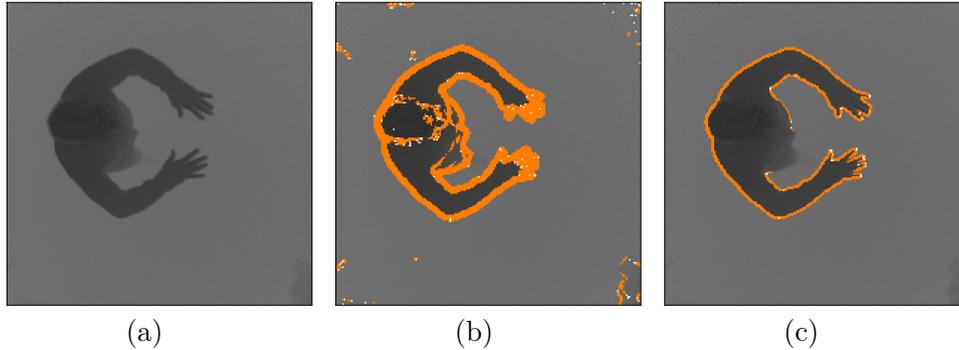


Abbildung 4.29: (a) Original. In (b), (c) wurden unterschiedliche Schwellwerte t_1 und t_2 festgelegt. Dadurch konnten mehr oder weniger strukturierte Bildpunkte identifiziert werden, die später der Ballungsanalyse übergeben werden.

Das DBSCAN–Verfahren von [Ester et. al. 1996] ist ein für diese Aufgabe effizientes Verfahren. Im Hinblick auf die XCLG–Methode sind folgende Eigenschaften des Verfahrens von Vorteil:

- Die Anzahl der Ballungen muss vorab nicht bekannt sein.
- Störungen in Form von Rauschen wird in der Modellierung berücksichtigt.
- Das Verfahren benötigt lediglich zwei Parameter.
- Eine Parallelisierung des Verfahrens ist möglich.

Das Verfahren gehört zur Klasse der dichtebasierten Ballungsverfahren. Dabei werden Ballungen durch Festlegung einer bestimmten Dichte bzgl. der Erfüllung eines Distanzmaßes akzeptiert. Die Punktemengen welche diese Dichte unterschreiten, werden zurückgewiesen und deshalb nicht als Ballung akzeptiert. Deshalb wird die Forderung aufgestellt, dass in einer Nachbarschaft eines Ballungspunktes eine bestimmte Anzahl von Punkten existieren muss. Eine notwendige Hilfsfunktion ist:

$$N_\epsilon(p) = \{q \in D | d(p, q) \leq \epsilon\} \quad . \quad (4.91)$$

Die Funktion $d(p, q)$ ist die Distanzfunktion (4.88), D steht für die Menge *aller* Punkte. Der Parameter ϵ legt die maximale Distanz fest. Damit ist $N_\epsilon(p)$ die Menge der Punkte, welche bzgl. des Distanzmaßes d in der ϵ –Nachbarschaft von p liegen. Für die Forderung, dass die Mächtigkeit der Menge $N_\epsilon(p) \geq \text{MinPts}$ ist, mit MinPts einem Schwellwert ist, wird nicht berücksichtigt, dass Randpunkte eine kleinere ϵ –Nachbarschaft besitzen als Punkte innerhalb der Ballung. Eine Anpassung des Parameters ϵ wird nur eine Klasse von Punkten berücksichtigen. Deshalb wird für jeden Punkt p einer Ballung C gefordert, dass ein Punkt $q \in C$ existiert, in dessen Nachbarschaft sich p befindet. Gleichzeitig soll gelten, dass die Anzahl der Punkte in der Nachbarschaft größer ist als der Parameter MinPts .

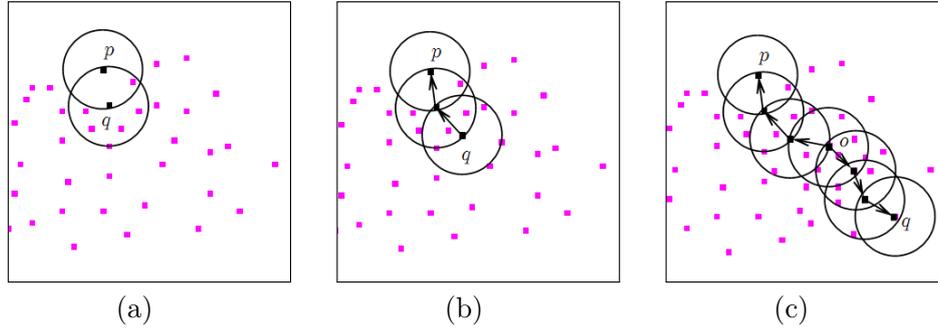


Abbildung 4.30: (a) Darstellungen zur Dichtezugehörigkeit (Definition 1), (b) Dichteerreichbarkeit (Definition 2) und Dichteverbundenheit (Definition 3).

Die im folgende angegebenen Definitionen sind adaptiert nach [Ester et. al. 1996]:

Definition 1:

1. $p \in N_\epsilon(q)$
2. $|N_\epsilon(q)| \geq MinPts$.

Der Punkt p bzgl. q gehört genau dann zur Dichte, wenn beide Voraussetzungen erfüllt sind (siehe Abbildung 4.30 (a)).

Definition 2: Wenn eine Kette von Punkten p_1, \dots, p_n , $p_1 = q, p_n = p$ existiert, so dass alle p_{i+1} direkt dichte erreichbar sind von p_i , dann ist p_n dichte erreichbar⁶ von p_1 (Abbildung 4.30 (b)).

Da zwei Randpunkte eventuell nicht dichte erreichbar sind, wird eine weitere Eigenschaft eingeführt.

Definition 3: Ein Punkt p ist dichte verbunden⁷ mit einem Punkt q , falls es einen Punkt o gibt, mit dem sowohl p als auch q dichte erreichbar sind (Abbildung 4.30 (c)).

Mit Hilfe der Definitionen 1 bis 3, können die dichte basierten Eigenschaften für eine Ballung festgelegt werden.

Definition 4: Für eine Menge D von Punkten, ist eine Ballung C bzgl. der Parameter ϵ und $MinPts$ eine nichtleere Teilmenge von D für die gilt:

1. $\forall p, q$: wenn $p \in C$ und q ist dichte-erreichbar von p bzgl. ϵ und $MinPts$, dann gilt $q \in C$ (Maximalität).
2. $\forall p, q \in C$: p ist dichte-verbunden mit q bzgl. ϵ und $MinPts$ (Verbundenheit).

Definition 5: C_1, \dots, C_k seien Ballungen der Punktmenge D bzgl. der Parameter ϵ_i und $MinPts_i$, für $i = 1, \dots, k$. Das Rauschen ist definiert als die Menge der Punkte aus D , welche nicht zu einer Ballung C_i gehören. Unter Verwendung der Definitionen 4, 5 und Festlegung der Parameter ϵ und $MinPts$, können die Ballungen bzw. das Rauschen in einer Menge von Punkten D gefunden werden.

⁶density reachable

⁷density connected

Schätzung der Strukturtensor-Parameter

Falls durch die Ballungsanalyse eine Zuordnung der Punkte zu bestimmten Ballungen (z.B. Ecken und Kanten) erfolgt ist, können die Parameter des räumlich-zeitlichen-ST (4.84) bestimmt werden. Die Anpassung betrifft nur die Faltungsfunktion \tilde{K} , welche an lokale Strukturen angepasst wird. Dazu werden für jeden Punkt der Ballung die Parameter für eine anisotrope Gauß-Funktion bestimmt. Die Parametrierung erfolgt über die Standardabweichungen ρ_1 und ρ_2 , sowie die Orientierung α .

Orientierung und Ausbreitung der Ballungen

Um die Eigenschaften der Ballung zu bestimmen, wird zuerst für jede Ballung eine Approximation der Kovarianzmatrix C und des Mittelwertes \bar{p} der Ballungspunkte berechnet:

$$\bar{p} = \frac{1}{N} \sum_{i=1}^N p_i \quad (4.92)$$

$$C = \frac{1}{N-1} \sum_{i=1}^N (p_i - \bar{p})(p_i - \bar{p})^T \quad (4.93)$$

Der Wert \bar{p} und die Matrix C sind erwartungstreue Schätzungen des Mittelwertes bzw. der Kovarianzmatrix. Eine Eigenwertzerlegung von C liefert die zugehörigen Eigenwerte λ_1 und λ_2 bzw. Eigenvektoren \mathbf{v}_1 und \mathbf{v}_2 . Die Eigenvektoren definieren ein orthonormales Koordinatensystem, welches die Orientierung der Ballung approximiert (Abbildung 4.31(b)). Um die Ausbreitung der Ballungspunktmenge zu schätzen, werden diese als Ausgabe eines Zufallprozesses behandelt. Die Punkte werden dabei als unabhängig-normalverteilte Zufallsvariablen angenommen:

$$p_i \sim N(\bar{p}, C) \quad , \quad \forall p_i \in D_c \quad (4.94)$$

Wird eine Koordinatentransformation in die Hauptachsen vorgenommen, respektive des Mittelwertes, spricht man von einer Hauptachsen- bzw. Karhunen-Loève-Transformation. Die so transformierten Punkte sind im statistischen Sinne unkorreliert und die Kovarianzmatrix C_{trans} hat die folgende Form:

$$C_{\text{trans}} = \begin{pmatrix} \rho_1 & 0 \\ 0 & \rho_2 \end{pmatrix} \quad (4.95)$$

Es kann gezeigt werden, dass ca. 99,7% der Dichte innerhalb des Intervalls $[-3\sigma, +3\sigma]$ liegen. Da die Varianzen bzgl. der transformierten Punkte bekannt sind, gilt für die Ausbreitung der Ballungspunkte:

$$a_i \approx w_\rho \sqrt{\lambda_i} \quad , \quad i \in \{1, 2\} \quad (4.96)$$

Es wird eine an den Hauptachsen der Ballung ausgerichtete Ellipse angepasst, so dass die Mehrzahl aller Punkte innerhalb der Ballung liegen (Abbildung 4.31 (b)).

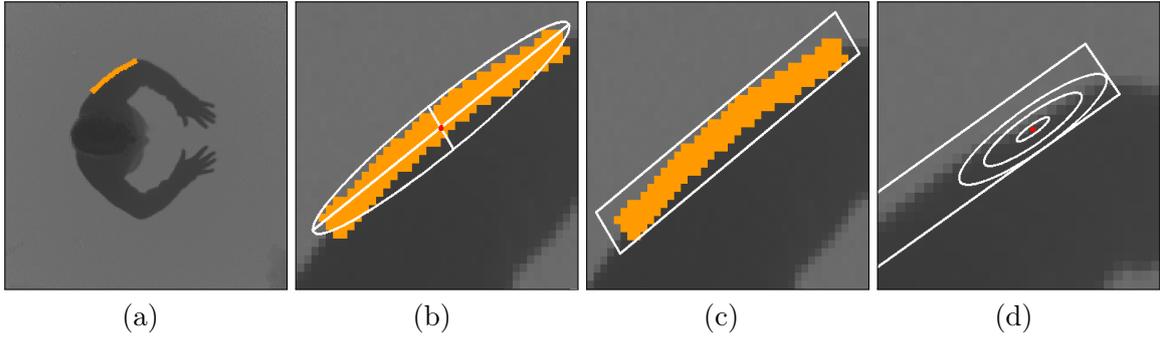


Abbildung 4.31: (a) Mit DBSCAN gefundene Kante. (b) Schätzung der Orientierung und Ausbreitung. (c) Umschließendes Rechteck der Ballung. (d) Bestimmung der Faltungsfunktion \tilde{K} für einen Ballungspunkt.

Punktweise Parameterschätzung

Mit Hilfe der Ballungseigenschaften kann für jeden Ballungspunkt eine Schätzung der Parameter erfolgen. Ein optimales Ergebnis ist eine Faltungsfunktion \tilde{K} , welche vollständig innerhalb der Ballung liegt, und gleichzeitig maximal ist (Abbildung 4.29 (d)). Daher gilt für die Parameter des transformierten Punktes p :

$$\begin{aligned}
 \omega_1 &= \frac{|a_1 - p_1|}{\omega_\rho} \quad , \\
 \omega_2 &= \frac{|a_2 - p_2|}{\omega_\rho} \quad , \\
 \alpha &= \arctan\left(\frac{v_{1y}}{v_{1x}}\right) \quad .
 \end{aligned} \tag{4.97}$$

Die Komponenten p_1 und p_2 bezeichnen die Position des transformierten Punktes p , bzgl. der ersten und zweiten Hauptachse.

Mit Hilfe der so geschätzten Parameter ist es möglich, \tilde{K} an den Ecken und Kanten anzupassen. Für Punkte, welche nicht zu Ballungen gehören, wird für alle ω_1 , ω_2 und α ein Standardwert festgelegt. Durch die anschließende Faltung des Datenterms mit \tilde{K} , wird der räumlich-zeitliche-ST erzeugt.

Konstante Gradientenbeträge

In dem Energiefunktional können neben der Intensitätskonstanz noch weitere Forderungen eingebracht werden. Die Intensitätskonstanztannahme hat den Vorteil, dass sie rotationsunabhängig ist. Der Nachteil ist, dass die Annahme bzgl. Helligkeitsänderungen nicht erfüllt ist. Eine weitere Forderung ist, dass sich der Gradient nicht ändert. Die Annahme ist unempfindlich bzgl. Helligkeitsvariationen, starke Rotationsbewegungen sind mit dieser Forderung nicht erfasst. Wird statt des Gradienten dessen Betrag als konstant angenommen, erhält man eine Forderung, die unempfindlich gegenüber Helligkeitsvariationen und starken Rotationsbewegungen ist. Die mathematische Notation hat die Form:

$$|\nabla_2 f(x, y, t)| = |\nabla_2 f(x + u_1, y + u_2, t + 1)| \quad . \tag{4.98}$$

Mit der Annahme, dass die Verschiebungen klein sind und dass die Bildfolge zeitlich und räumlich hinreichend kleine Differenzen aufweist, kann die Taylor-Reihenentwicklung erster Ordnung angewandt werden. Das führt auf die linearisierte Form:

$$\nabla_3|\nabla_2 f|\mathbf{u} = 0 \quad (4.99)$$

mit $\mathbf{u} = (u_1, u_2, 1)^T$.

Um die Forderung in die OF-Schätzung einzubringen, wird (4.99) quadriert und additiv in das Energiefunktional aufgenommen. Da die lokale Umgebung ebenfalls durch die Funktion \tilde{K} bestimmt wird, werden beide Forderungen in einem Datenterm zusammengefasst:

$$\begin{aligned} \tilde{D}_\eta &= (1 - \eta) \underbrace{\mathbf{u}^T \tilde{K} * (\nabla_3 f \nabla_3 f^T) \mathbf{u}}_{\tilde{K} * (\nabla_3 f \mathbf{u})^2} + \eta \underbrace{\mathbf{u}^T \tilde{K} * ((\nabla_3 |\nabla_2 f|)(\nabla_3 |\nabla_2 f|)^T) \mathbf{u}}_{\tilde{K} * (\nabla_3 |\nabla_2 f| \mathbf{u})^2} \\ &= \mathbf{u}^T \left((1 - \eta) \tilde{K} * (\nabla_3 f \nabla_3 f^T) + \eta \tilde{K} * ((\nabla_3 |\nabla_2 f|)(\nabla_3 |\nabla_2 f|)^T) \right) \mathbf{u} \\ &= \mathbf{u}^T \left(\tilde{K} * ((1 - \eta) \nabla_3 f \nabla_3 f^T + \eta (\nabla_3 |\nabla_2 f|)(\nabla_3 |\nabla_2 f|)^T) \right) \mathbf{u} \\ &= \mathbf{u}^T \tilde{J}(\nabla_3 f, \nabla_3 |\nabla_2 f|) \mathbf{u} \end{aligned} \quad (4.100)$$

Der Parameter $\eta \in [0, 1]$ gewichtet die Forderungen in der konvexen Summe. Das resultierende Energiefunktional hat die Form:

$$E_{XCLG}(\mathbf{u}) = \int_{\Omega_2} \underbrace{\mathbf{u}^T \tilde{J}(\nabla_3 f, \nabla_3 |\nabla_2 f|) \mathbf{u}}_{\tilde{D}_\eta} + \alpha \underbrace{(|\nabla_2 u_1|^2 + |\nabla_2 u_2|^2)}_G \, d\mathbf{x} \quad (4.101)$$

Angepasster Ableitungsoperator

Zur Bestimmung des räumlichen Gradienten werden die Ableitungen nach beiden Bildkoordinaten geschätzt. Da die Schätzung anfällig gegenüber Störungen ist, wird das Bild vorher mit einem Tiefpassfilter gefaltet. Die anschließende Differentiation des geglätteten Bildes ist äquivalent zur Faltung des Bildes mit dem differenzierten Tiefpassfilter:

$$\begin{aligned} \frac{\partial}{\partial x} (t(\mathbf{x}) * f(\mathbf{x})) &= \frac{\partial}{\partial x} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(\boldsymbol{\alpha}) t(\mathbf{x} - \boldsymbol{\alpha}) \, d\boldsymbol{\alpha} \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(\boldsymbol{\alpha}) \frac{\partial t}{\partial x}(\mathbf{x} - \boldsymbol{\alpha}) \, d\boldsymbol{\alpha} \\ &= \frac{\partial t}{\partial x}(\mathbf{x}) * f(\mathbf{x}) \quad . \end{aligned} \quad (4.102)$$

Die Funktion t bezeichnet die Funktion des Tiefpassfilters.

Für die XCLG-Methode wird zur Bestimmung der räumlichen Ableitungen ein differenzierter Tiefpassfilter verwendet. Der Gradient $\nabla_2 f$ wird durch isotrope differenzierte Gauß-Funktionen bestimmt (Abbildung 4.32 (a), (b)). Strukturen, die nicht horizontal oder vertikal

verlaufen, werden dabei geglättet. Dadurch werden die partiellen Ableitungen nicht richtig bzw. unterschätzt.

Da die Faltungsfunktion \tilde{K} an lokale Strukturen angepasst wird (Abbildung 4.32 (c)), ist der daraus gebildete Ableitungsoperator ebenfalls an die Struktur angepasst (Abbildung 4.32 (d)). Die zur Bestimmung des Gradienten $\tilde{\nabla}_2 f$ angepassten Ableitungsoperatoren erhalten daher die lokale Struktur.

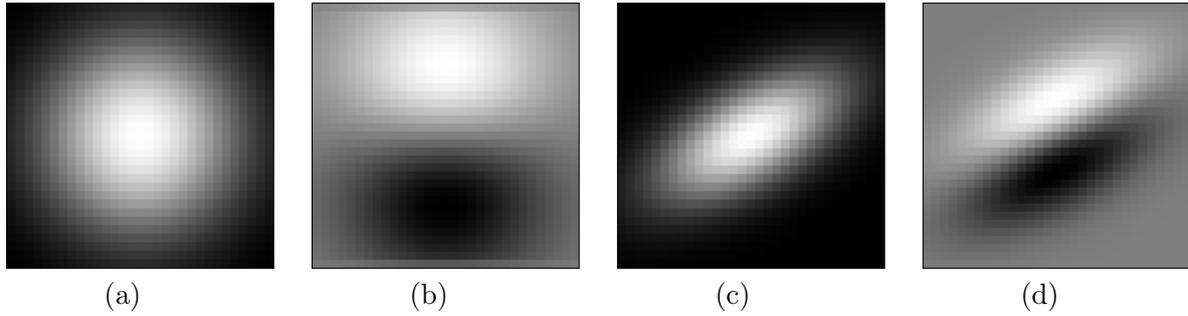


Abbildung 4.32: (a), (b): Diskrete, isotrope Gauß-Funktion und der daraus gebildete Ableitungsoperator. (c), (d): Diskrete, anisotrope gedrehte Gauß-Funktion und der daraus gebildete Ableitungsoperator.

Das resultierende Energiefunktional hat die Form:

$$E_{XCLG}(\mathbf{u}) = \int_{\Omega_2} \underbrace{\mathbf{u}^T \tilde{J}(\nabla_3 f, \nabla_3 |\tilde{\nabla}_2 f|)}_{\tilde{D}_\eta} \mathbf{u} + \alpha \underbrace{(|\nabla_2 u_1|^2 + |\nabla_2 u_2|^2)}_G dx \quad . \quad (4.103)$$

4.4.3 Effiziente Lösung durch Mehrgitter-Verfahren

Da die XCLG-Methode für den Kontext der Mensch-Roboter-Kooperation entwickelt wurde, ist der Berechnungsaufwand zur Bestimmung des Optischen Flussfeldes in jedem Zeitschritt ein sehr wichtiger Faktor. Die Schätzungen des OF müssen, damit sie sinnvoll eingesetzt werden können, innerhalb des Kameraktaktes berechnet werden. Eine geeignete Klasse zur Behandlung dieser Problematik bieten die sog. Mehrgitter-Verfahren [Hackbusch 1994, Wesseling 2004, Bruhn 2006].

Mitte der 80er Jahre wurden diese Verfahren erstmals zur OF-Berechnung eingesetzt [Glazer 1984, Terzopoulos 1986, Enkelmann 1987]. Die Ergebnisse wiesen daraufhin, dass Mehrgitter-Verfahren nicht für diese Art von partiellen Differentialgleichungen geeignet waren. Ende der 90er Jahre wurde das Mehrgitter-Konzept wieder erfolgreich aufgegriffen, mit angepassten Entwurfsansätzen [Zini u.a. 1997, Kalmoun und Rüde 2003]. Seither haben sich diese Verfahren, für Echtzeitberechnungen als Standard etabliert.

Die Anwendung der Verfahren ist vergleichsweise schwierig. Für den grundlegenden Aufbau des Verfahrens muss ein Glättungsverfahren gewählt und die Abfolge der Gitterweiten festgelegt werden. Parameter, wie bspw. die Anzahl der Iterationen für die Vor- bzw. Nachglättung, werden durch Experimente an das spezielle Problem angepasst.

Vom Energiefunktional zum LGS

Zur Schätzung des OF mit Hilfe der XCLG-Methode werden Funktionen $u_1(x, y, t)$ und $u_2(x, y, t)$ gesucht, welche das Energiefunktional in Gleichung (4.103) minimieren. Dabei steht $f(x, y, t)$ für eine Bildfolge, (x, y) beschreibt den Ort in einer rechteckigen Bilddomäne $\Omega_2 \subset \mathbb{R}^2$. Die Komponente $t \in [0, T]$ steht für die Zeit. Der Term $\tilde{J}(\nabla_3 f, \nabla_3 |\tilde{\nabla}_2 f|)$ bezeichnet den räumlich-zeitlichen-ST:

$$\tilde{J}(\nabla_3 f, \nabla_3 |\tilde{\nabla}_2 f|) = \tilde{K} * \left((1 - \eta) \nabla_3 f \nabla_3 f^T + \eta (\nabla_3 |\tilde{\nabla}_2 f|) (\nabla_3 |\tilde{\nabla}_2 f|)^T \right) \quad (4.104)$$

Eine Lösung des Energiefunktional (4.103) wird erreicht durch Lösung der korrespondierenden Euler-Lagrange-Gleichungen:

$$0 = \Delta u_1 - \frac{1}{\alpha} \left(\tilde{J}_{11} u_1 + \tilde{J}_{12} u_2 + \tilde{J}_{13} \right) \quad , \quad (4.105)$$

$$0 = \Delta u_2 - \frac{1}{\alpha} \left(\tilde{J}_{21} u_1 + \tilde{J}_{22} u_2 + \tilde{J}_{23} \right) \quad . \quad (4.106)$$

Wie leicht zu sehen ist, sind die durch das Energiefunktional (4.103) entstandenen Euler-Lagrange-Gleichungen (4.105) und (4.106) unabhängig von der Form bzw. Art der Faltungsfunktion \tilde{K} .

Durch die Diskretisierung der OF-Schätzung u_1 und u_2 und Anwendung der Finiten-Differenzen-Methode, beschreibt man den Laplace-Operator Δ :

$$\Delta u = u_{xx} + u_{yy} \quad (4.107)$$

durch die Approximation:

$$\Delta u = \frac{u_{i+1,j} - 2u_{ij} + u_{i-1,j}}{h^2} + \frac{u_{i,j+1} - 2u_{ij} + u_{i,j-1}}{h^2} \quad . \quad (4.108)$$

Die diskreten Stellen, an denen die Funktionen u_1 und u_2 berechnet werden, wird durch ein 2D-Gitter mit den Gitterweiten h_x und h_y festgelegt. Für die OF-Berechnung mithilfe der XCLG-Methode wurde lediglich ein Parameter h zur Diskretisierung gewählt.

Durch Einsetzen der Approximation (4.108) erhält man das LGS⁸:

$$0 = \sum_{j \in N(i)} \frac{u_j - u_i}{h^2} - \frac{1}{\alpha} \left(\tilde{J}_{11i} u_i + \tilde{J}_{12i} v_i + \tilde{J}_{13i} \right) \quad , \quad (4.109)$$

$$0 = \sum_{j \in N(i)} \frac{v_j - v_i}{h^2} - \frac{1}{\alpha} \left(\tilde{J}_{21i} u_i + \tilde{J}_{22i} v_i + \tilde{J}_{23i} \right) \quad . \quad (4.110)$$

Der Index i , mit $i = 1..N$ und $N = \text{Bildweite} * \text{Bildhöhe}$, steht für die Bildposition, wobei die Positionen zeilenweise angeordnet sind. Die Nachbarschaft $N(i)$ entspricht den Positionen bzgl. der Approximation des Laplace-Operators (4.108). Die horizontalen und vertikalen

⁸Lineares-Gleichungs-System

Fixpunktiteration

Eine Möglichkeit ein LGS

$$\mathbf{Ax} = \mathbf{b}, \quad \mathbf{A} \in \mathbb{R}^{n \times n}, \quad \mathbf{x}, \mathbf{b} \in \mathbb{R}^n, \quad (4.112)$$

numerisch zu lösen, geschieht durch Anwendung iterativer Verfahren. Das Jacobi- und das Gauß-Seidel-Verfahren gehören bspw. zu dieser Klasse. Zur Herleitung wird in einem ersten Schritt das LGS in eine lineare Fixpunktform überführt.

$$\mathbf{x} = \mathbf{T}\mathbf{x} + \mathbf{c}, \quad \mathbf{T} \in \mathbb{R}^{n \times n}, \quad \mathbf{c} \in \mathbb{R}^n. \quad (4.113)$$

Dazu wird (4.112) zeilenweise nach der jeweiligen Unbekannten aufgelöst:

$$x_i = -\frac{1}{a_{ii}} \left[\sum_{\substack{j=1 \\ j \neq i}}^n a_{ij} x_j - b_i \right], \quad i = 1, 2, \dots, n. \quad (4.114)$$

Mit Hilfe des Banachschen-Fixpunktsatzes [Schwarz & Köckler 2009] kann gezeigt werden, dass eine Iterationsfolge $\mathbf{x}^{k+1} = \mathbf{T}\mathbf{x}^{(k)} + \mathbf{c}$ unter bestimmten Voraussetzungen gegen einen eindeutigen Fixpunkt \mathbf{x} konvergiert. Die Lösung der Fixpunktgleichung (4.113) ist auch eine Lösung des LGS (4.112).

Für das Jacobi-Verfahren ergibt sich die Iterationsvorschrift zu:

$$x_i^{k+1} = -\frac{1}{a_{ii}} \left[\sum_{\substack{j=1 \\ j \neq i}}^n a_{ij} x_j^k - b_i \right], \quad i = 1, 2, \dots, n. \quad (4.115)$$

Werden statt des alten iterierten Vektors \mathbf{x}^k die jeweils neu berechnete Komponenten x_j^{k+1} verwendet, führt das auf die Iterationsvorschrift für das Gauß-Seidel-Verfahren:

$$x_i^{k+1} = -\frac{1}{a_{ii}} \left[\sum_{j=1}^{i-1} a_{ij} x_j^{k+1} + \sum_{j=i+1}^n a_{ij} x_j^k - b_i \right], \quad i = 1, 2, \dots, n. \quad (4.116)$$

Für praktische Fälle konnte gezeigt werden [Schwarz & Köckler 2009], dass das Konvergenzverhalten wesentlich verbessert werden kann, falls die Korrektur

$$\Delta x_i^{k+1} = x_i^{k+1} - x_i^k \quad (4.117)$$

mit einem Relaxationsfaktor ω multipliziert und zu \mathbf{x}^k addiert wird. Die resultierende Iterationsvorschrift (SOR⁹-Verfahren) lautet:

$$\begin{aligned} x_i^{k+1} &= x_i^k + \omega \cdot \Delta x_i^{k+1} \\ &= (1 - \omega)x_i^k - \frac{\omega}{a_{ii}} \left[\sum_{j=1}^{i-1} a_{ij} x_j^{k+1} + \sum_{j=i+1}^n a_{ij} x_j^k - b_i \right]. \end{aligned} \quad (4.118)$$

Eine Anwendung auf das resultierende LGS (4.111) der XCLG-Methode, führt auf die Iterationsvorschrift:

⁹engl. successive overrelaxation

$$\begin{aligned}
u_i^{k+1} &= (1 - \omega)u_i^k + \omega \frac{\sum_{j \in N^-(i)} u_j^{k+1} + \sum_{j \in N^+(i)} u_j^k - \frac{h^2}{\alpha} \left(\tilde{J}_{12i} v_i^k + \tilde{J}_{13i} \right)}{4 + \frac{h^2}{\alpha} \tilde{J}_{11i}} , \\
v_i^{k+1} &= (1 - \omega)v_i^k + \omega \frac{\sum_{j \in N^-(i)} v_j^{k+1} + \sum_{j \in N^+(i)} v_j^k - \frac{h^2}{\alpha} \left(\tilde{J}_{21i} u_i^{k+1} + \tilde{J}_{23i} \right)}{4 + \frac{h^2}{\alpha} \tilde{J}_{22i}} ,
\end{aligned}$$

mit

$$\begin{aligned}
N^-(i) &= j \in N(i) | j < i \quad , \\
N^+(i) &= j \in N(i) | j > i \quad .
\end{aligned} \tag{4.119}$$

Glättungseigenschaft iterativer Verfahren

Ist ein LGS wie im Fall der XCLG-Methode aus der Lösung partieller Differentialgleichungen entstanden, so haben bestimmte iterative Verfahren wie das Gauß-Seidel oder das Jacobi-Verfahren spezielle Eigenschaften.

Es wurde gezeigt, dass der Fehler \mathbf{e} nach wenigen Iterationsschritten schnell, und in den folgenden Iterationen vergleichsweise signifikant langsamer abnimmt [Briggs u.a. 2000].

Der Fehler ist definiert, als die Differenz der wahren Lösung \mathbf{s} und der Schätzung \mathbf{v} des linearen Gleichungssystems:

$$\mathbf{e} = \mathbf{s} - \mathbf{v} \quad . \tag{4.120}$$

Eine generelle Untersuchung des Verhaltens bestimmter iterativer Verfahren zur Lösung allgemeiner partieller Differentialgleichungen, kann in [Briggs u.a. 2000] gefunden werden.

Zwei-Gitter Schema

In [Briggs u.a. 2000] wird gezeigt, dass bspw. das Gauß-Seidel-Verfahren schnell konvergieren, solange der Fehler hochfrequente Anteile enthält. Das Zwei-Gitter-Schema nutzt dies aus, um auch nach den ersten Iterationen eine hohe Konvergenzrate zu erhalten.

Vorab wird Folgendes festgelegt:

$$\mathbf{Ax} = \mathbf{b} \quad , \tag{4.121}$$

$$\mathbf{r} = \mathbf{b} - \mathbf{Ax} \quad , \tag{4.122}$$

$$\mathbf{Ae} = \mathbf{r} \quad . \tag{4.123}$$

Der Wert \mathbf{r} steht für das Residuum, die Gleichung (4.122) wird als Residuungleichung bezeichnet. Gleichung (4.123) ist das zu lösende lineare Gleichungssystem.

Ablauf des Verfahrens

Im ersten Schritt wird mit einer geringen Anzahl Iterationen der Fehler geglättet, also die hochfrequenten Anteile beseitigt. Weitere Iterationen würden nur wenig Verbesserung des Fehlers bewirken, da dieser hauptsächlich niederfrequente Anteile besitzt. Deshalb wird der Fehler mit Hilfe der Residuums Gleichung (4.123) auf einem gröberem Gitter geschätzt. Durch die Abbildung des LGS auf ein gröberes Gitter, enthält der Fehler wieder hochfrequente Anteile und kann deshalb schneller gegen die Lösung konvergieren. Ist die Lösung der Residuums Gleichung ausreichend genau geschätzt worden, wird sie auf das feinere Gitter interpoliert und im Korrekturschritt auf die Schätzung \mathbf{v} des ursprünglichen LGS addiert, wodurch die meisten niederfrequenten Anteile des Fehlers beseitigt werden. Allerdings fängt man sich wieder hochfrequente Teile ein. Deshalb wird in einem letzten Schritt nachgeglättet.

Algorithmus 2: Zwei-Gitter Schema

- 1 n_1 mal auf dem LGS $A^h \mathbf{x}^h = \mathbf{b}^h$ glätten, mit der initialen Schätzung \mathbf{v}^h .
 - 2 Restriktion des Residuums: $\mathbf{r}^{2h} = \mathbf{I}^{h \rightarrow 2h}(\mathbf{b}^h - A^h \mathbf{v}^h)$.
 - 3 Löse $A^{2h} \mathbf{e}^{2h} = \mathbf{r}^{2h}$, $\mathbf{e}_{init}^{2h} = \mathbf{0}$.
 - 4 Korrektur: $\mathbf{v}^h \leftarrow \mathbf{v}^h + \mathbf{I}^{2h \rightarrow h} \mathbf{e}^{2h}$.
 - 5 n_2 mal auf dem LGS $A^h \mathbf{x}^h = \mathbf{b}^h$ nachglätten, mit der initialen Schätzung \mathbf{v}^h .
-

Hochgestellte Indizes bezeichnen die aktuelle Gitterweite.

Eine Übertragung des ursprünglichen LGS würde keinen Sinn machen, da die Lösung sehr wahrscheinlich hochfrequente Anteile enthält. Eine Restriktion auf das gröbere Gitter würde die Lösung wegen der Verletzung des Abtasttheorems verfälschen.

In dem Zwei-Gitter-Schema müssen Transfers zwischen den unterschiedlichen Gittern durchgeführt werden. Für diese Aufgabe müssen geeignete Operatoren festgelegt werden.

Restriktion

Um den Fehler mit Hilfe der Residuums Gleichung auf einem gröberem Gitter Ω^{2h} zu berechnen, muss das Residuum auf das gröbere Gitter überführt werden. Dazu wird eine Restriktion des Residuums \mathbf{r} mit Hilfe des Restriktionsoperators $I^{h \rightarrow 2h}$ durchgeführt.

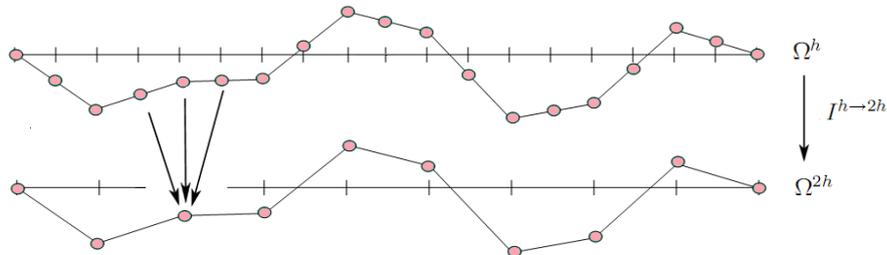


Abbildung 4.33: Restriktion mit der Methode der vollen Gewichtung (1D).

Eine Restriktionsvariante ist die sogenannte Methode der vollen Gewichtung¹⁰ (Abbildung 4.33), welches für $\mathbf{v}^{2h} = I^{h \rightarrow 2h} \mathbf{v}^h$ definiert ist durch:

$$v_i^{2h} = \frac{1}{4}(v_{2i-1}^h + 2v_{2i}^h + v_{2i+1}^h) \quad 1 \leq i \leq \frac{n}{2} - 1 \quad (4.124)$$

im eindimensionalen Fall, und respektive für den zweidimensionalen Fall durch:

$$v_{ij}^{2h} = \frac{1}{16} \left((v_{2i+1,2j+1}^h + v_{2i-1,2j-1}^h + v_{2i-1,2j+1}^h + v_{2i+1,2j-1}^h) \right. \\ \left. + 2(v_{2i+1,2j}^h + v_{2i-1,2j}^h + v_{2i,2j+1}^h + v_{2i,2j-1}^h) \right. \\ \left. + 4v_{2i,2j}^h \right), \quad 1 \leq i, j \leq \frac{n}{2} - 1. \quad (4.125)$$

An den Rändern wird eine abgewandelte Form des Operators verwendet.

Prolongation

Nachdem der Fehler \mathbf{e} auf dem größeren Gitter Ω^{2h} mit ausreichender Genauigkeit berechnet wurde, muss das Ergebnis wieder auf das feinere Gitter Ω^h überführt werden. Dazu wird eine Interpolation mit Hilfe des Interpolationsoperators $I^{2h \rightarrow h}$ durchgeführt.

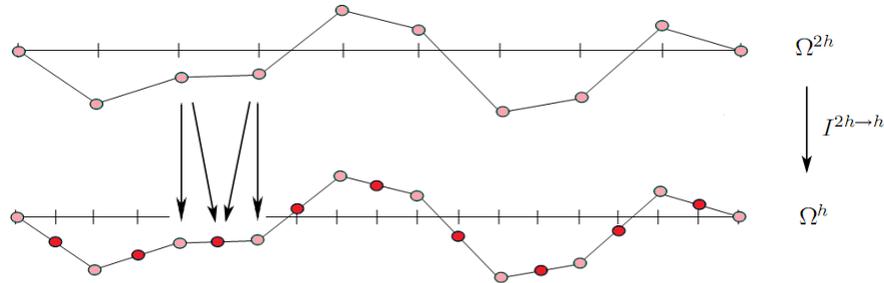


Abbildung 4.34: Lineare Interpolation (1D).

Speziell für Mehrgitterverfahren eignet sich neben anderen Operatoren, die lineare Interpolation (Abbildung 4.34) [Briggs u.a. 2000]. Diese ist für $\mathbf{v}^h = I^{2h \rightarrow h} \mathbf{v}^{2h}$ definiert durch:

$$v_{2i}^h = v_i^{2h}, \\ v_{2i+1}^h = \frac{1}{2}(v_i^{2h} + v_{i+1}^{2h}), \quad 0 \leq i \leq \frac{n}{2} - 1 \quad (4.126)$$

im eindimensionalen Fall und respektive für den zweidimensionalen Fall durch:

¹⁰engl. full weighting

$$\begin{aligned}
v_{2i,2j}^h &= v_{ij}^{2h}, \\
v_{2i+1,2j}^h &= \frac{1}{2}(v_{ij}^{2h} + v_{i+1,j}^{2h}), \\
v_{2i,2j+1}^h &= \frac{1}{2}(v_{ij}^{2h} + v_{i,j+1}^{2h}), \\
v_{2i+1,2j+1}^h &= \frac{1}{4}(v_{ij}^{2h} + v_{i+1,j}^{2h} + v_{i,j+1}^{2h} + v_{i+1,j+1}^{2h}), \quad 0 \leq i, j \leq \frac{n}{2} - 1. \quad (4.127)
\end{aligned}$$

An den Rändern muss der Operator ebenfalls angepasst werden.

Anwendung auf die OF-Berechnung

Die Lösung des Energiefunktionals (4.103) der XCLG-Methode, führt auf die Lösung der Euler-Lagrange-Gleichungen (4.105) und (4.106). Dabei handelt es sich um gekoppelte PDGen. Aus der numerischen Approximation entsteht das LGS (4.111), dessen Lösung die horizontalen und vertikalen Komponenten u_1 und u_2 des OF enthält. Da die PDGen gekoppelt sind, ergibt sich die punktweise zusammengesetzte Anordnung der OF-Komponenten im Lösungsvektor (4.111). Deshalb muss die Restriktion des Residuums \mathbf{r}^h und die Interpolation des Fehlers \mathbf{e}^{2h} für jede Komponente getrennt in 2D durchgeführt werden. Das Ergebnis wird im Anschluss wieder in gleicher Anordnung zusammengesetzt.

Des Weiteren wird bei dem Wechsel auf unterschiedliche Gitter die Systemmatrix A^h transformiert. Dabei ist zu beachten, dass der Faktor h in der Systemmatrix an die jeweilige Gitterweite angepasst wird.

Mehrgitter Schema

Das Zwei-Gitter-Schema lässt sich intuitiv erweitern. Dazu wird für die Berechnung des Fehlers, mit Hilfe der Residuungsgleichung (4.123) ebenfalls ein Zwei-Gitter-Schema verwendet. Dadurch entsteht eine Struktur mit rekursiver Berechnungsvorschrift, welche als V-Zyklus bezeichnet wird [Briggs u.a. 2000]:

Algorithmus 3: Rekursiver V-Zyklus

$$\mathbf{v}^h \leftarrow V^h(\mathbf{v}^h, \mathbf{b}^h) :$$

1 n_1 mal auf dem LGS $A^h \mathbf{x}^h = \mathbf{b}^h$ glätten, mit der initialen Schätzung \mathbf{v}^h .

2 **if** $t_{rek} < t_{max}$ **then**

3 Restriktion des Residuums: $\mathbf{r}^{2h} \leftarrow I^{h \rightarrow 2h}(\mathbf{b}^h - A^h \mathbf{v}^h)$.

4 Initialisierung der Schätzung: $\mathbf{v}^{2h} \leftarrow \mathbf{0}$.

5 t_{rek} um 1 erhöhen.

6 Rekursion: $\mathbf{v}^{2h} \leftarrow V^{2h}(\mathbf{v}^{2h}, \mathbf{r}^{2h})$.

7 Korrektur: $\mathbf{v}^h \leftarrow \mathbf{v}^h + I^{2h \rightarrow h} \mathbf{v}^{2h}$.

8 **end**

9 n_2 mal auf dem LGS $A^h \mathbf{x}^h = \mathbf{b}^h$ nachglätten, mit der initialen Schätzung \mathbf{v}^h .

Mit t_{rek} wird die aktuelle Rekursionstiefe bezeichnet. Die Konstante t_{max} steht für die maximale Rekursionstiefe, wobei diese nach oben durch die Samplingrate der Daten beschränkt

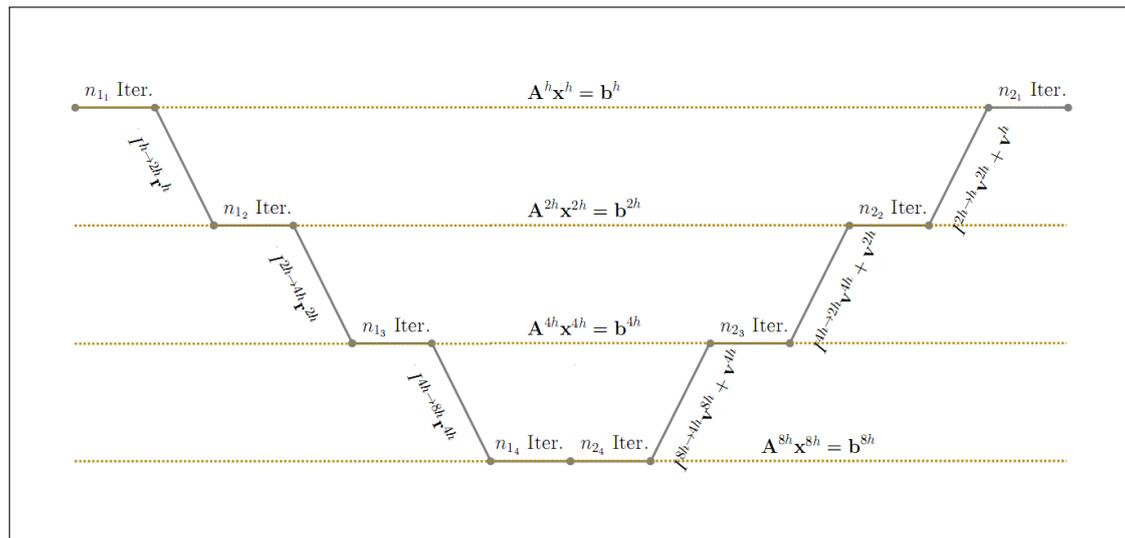


Abbildung 4.35: V-Zyklus mit der maximalen rekursiven Tiefe: $t_{max} = 3$.

ist.

Eine weitere rekursive Mehrgitter Form ist der μ -Zyklus. Die rekursive Berechnungsvorschrift ist eine Verallgemeinerung des V-Zyklus [Briggs u.a. 2000]:

Der Fehler wird in jedem Rekursionsschritt μ mal rekursiv berechnet. Als initiale Schätzung des Fehlers wird jeweils das Ergebnis der vorangegangenen Rekursion verwendet.

Für die Optische Fluss Schätzung werden hauptsächlich μ -Zyklen mit $\mu = 1$ (Abbildung 4.35) und $\mu = 2$ (Abbildung 4.36) angewendet.

Algorithmus 4: Rekursiver μ -Zyklus

$$\mathbf{v}^h \leftarrow M\mu^h(\mathbf{v}^h, \mathbf{b}^h) :$$

1 n_1 mal auf dem LGS $A^h \mathbf{x}^h = \mathbf{b}^h$ glätten, mit der initialen Schätzung \mathbf{v}^h .

2 **if** $t_{rek} < t_{max}$ **then**

3 Restriktion des Residuums: $\mathbf{r}^{2h} \leftarrow I^{h \rightarrow 2h}(\mathbf{b}^h - A^h \mathbf{v}^h)$.

4 Initialisierung der Schätzung: $\mathbf{v}^{2h} \leftarrow \mathbf{0}$.

5 t_{rek} um 1 erhöhen.

6 Rekursion: $\mathbf{v}^{2h} \leftarrow M\mu^{2h}(\mathbf{v}^{2h}, \mathbf{r}^{2h})$ μ mal .

7 Korrektur: $\mathbf{v}^h \leftarrow \mathbf{v}^h + I^{2h \rightarrow h} \mathbf{v}^{2h}$.

8 **end**

9 n_2 mal auf dem LGS $A^h \mathbf{x}^h = \mathbf{b}^h$ nachglätten, mit der initialen Schätzung \mathbf{v}^h .

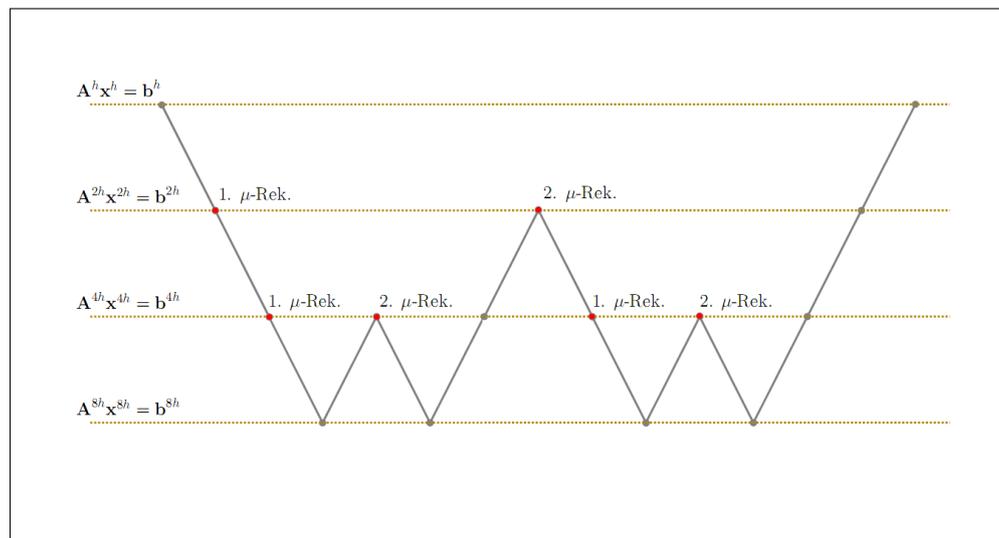


Abbildung 4.36: Ablaufdiagramm des μ -Zyklus mit der maximalen rekursiven Tiefe: $t_{max} = 3$ und $\mu = 2$.

4.4.4 3D Optischer Fluss

In MAROCO werden 3D-Kameras verwendet, um Informationen über die Szene zu erhalten. Diese basieren auf der PMD Technologie und berechnen Tiefenmessungen mithilfe des Lichtflugzeitverfahren. Es kommen zwei Typen zum Einsatz:

1. Ein *Cam Cube* der Firma PMD Technologies. Die Kamera liefert Tiefenmessungen mit einer Auflösung von 204×204 . Die Kamera liefert zusätzlich zu den Tiefenwerten auch die zugehörigen Helligkeitswerte. Die OF-Schätzung erfolgt dementsprechend auf den Grauwertinformationen.
2. Eine Kamera der Firma IFM, mit einer Auflösung von 50×64 . Im Gegensatz zum *Cam Cube* werden keine Helligkeitsinformationen zu den Tiefenmessungen geliefert. Der OF wird deshalb auf den grauwertkodierten Tiefendaten berechnet. Problematisch ist, dass der variationelle Ansatz auf der OFBG beruht. Durch eine ausreichend hohe Abtastfrequenz können trotzdem gute Schätzungen erzielt werden.

Durch die OF-Schätzung erfolgt eine Registrierung von Punkten, welche in Messungen zum Zeitpunkt t und $t + 1$ erscheinen. Um den OF bzw. die Verschiebung im Raum zu berechnen, müssen die Positionen der zugeordneten Punkte im Raum bestimmt werden. Die Differenz der Positionen entspricht dann der Verschiebung im Raum.

Im Folgenden werden vereinfachende Annahmen bzgl. des Kameramodells gemacht. Dazu wird ein lineares Modell verwendet, bei dem die Verzerrung des Bildes nicht beachtet wird.

Positionsbestimmung

Zur Positionsbestimmung eines Punktes werden zuerst die Winkel α_x und α_y (Abbildung 4.37) berechnet. Dies geschieht unabhängig voneinander, da Verschiebungen in der Bildebene, in x bzw. y Richtung, linear unabhängig voneinander sind:

$$\begin{aligned}f_x &= (x - o_x) \cdot s_x \quad , \\f_y &= (y - o_y) \cdot s_y \quad , \\ \alpha_x &= \arctan\left(\frac{d_b}{f_x}\right) \quad , \\ \alpha_y &= \arctan\left(\frac{d_b}{f_y}\right) \quad .\end{aligned}\tag{4.128}$$

Die Position $(f_x, f_y)^T$ in der Bildebene, wird aus der Pixelposition $(x, y)^T$, der Pixelgröße s_x bzw. s_y und dem Bildursprung $(o_x, o_y)^T$ berechnet. Der Abstand der Bildebene zum Kameraursprung wird mit d_b bezeichnet (Abbildung 4.38).

Durch Projektion des Punktes in die xz - und yz -Ebene (Abbildung 4.38), kann die Position berechnet werden:

$$\begin{aligned}p_x &= -\frac{d}{\tan(\alpha_x)} = -\frac{d \cdot f_x}{d_b} \quad , \\p_y &= -\frac{d}{\tan(\alpha_y)} = -\frac{d \cdot f_y}{d_b} \quad . \\p_z &= d \quad .\end{aligned}\tag{4.129}$$

Berechnung des OF

Wenn für beide Punkte die Positionen bestimmt wurden, wird der OF definiert als:

$$\begin{aligned}u_x &= p_{2x} - p_{1x} \quad , \\u_y &= p_{2y} - p_{1y} \quad . \\u_z &= p_{2z} - p_{1z} \quad .\end{aligned}\tag{4.130}$$

Die berechneten Punktpositionen und der OF beziehen sich auf das Koordinatensystem der Kamera. In Abbildung 4.39 sind ausgewählte Bilder mit berechnetem 3D-OF zu sehen.

Berechnungen auf Grafikkartenprozessoren

Um Berechnungen in Echtzeit zu ermöglichen wurde die XCLG-Methode auf die Grafikkarte portiert. Als Hardware wird eine NVIDIA GTX 285 Grafikkarte mit 265 Stream-Prozessoren eingesetzt. Die GTX Serie ist mit der CUDA Architektur ausgestattet. Bei der GPU Programmierung wird mit der Hochsprache wie C/C++ gearbeitet. Programmierung auf Treiber Ebene ist möglich, aber nicht erforderlich.

Der Fokus der Arbeit lag nicht auf der Parallelisierung oder der Übertragung der Berechnungen auf die Grafikkarte. Die parallelisierten Schritte wurden nicht optimiert.

Faltung

Faltungen lassen sich direkt parallelisieren. Dabei berechnet je ein Thread die Faltung in einem Punkt. Die Faltungsmasken werden nicht gespeichert, sondern zur Laufzeit in dem jeweiligen Thread erstellt. Dies ist speziell für die Berechnung des örtlich-zeitlichen-ST von Vorteil.

Ballungsanalyse

Die Ballungsanalyse wurde in einer hybriden Form auf der CPU und GPU implementiert. Auf der GPU wird der örtliche-ST berechnet, die Eigenwertzerlegung durchgeführt und die Eigenschaften der einzelnen Punkte bestimmt. Die Berechnung der Distanzmatrix und die Auswahl der möglichen Ballungspunkte läuft ebenfalls auf der Grafikkarte. Auf der CPU wird der DBSCAN-Algorithmus ausgeführt. Die punktweise Parameterbestimmung der Faltungsfunktion \tilde{K} wurde vollständig auf die Grafikkarte portiert.

Mehrgitter Methode

Die Mehrgitter-Methode wird vollständig auf der GPU ausgeführt. Die Iterationsschritte des *SOR*-Verfahrens, welche zur Vor/- und Nachglättung verwendet werden, wurden nicht parallelisiert. Die Berechnungsreihenfolge der einzelnen Updates x_i^{k+1} ist zufällig. Dabei werden zur Berechnung aktuelle Werte x_j^{k+1} verwendet, falls diese bereits von einem anderen Thread berechnet wurden. Von daher unterscheidet sich diese Variante von dem Jacobi-Verfahren und erreicht bessere Konvergenzraten.

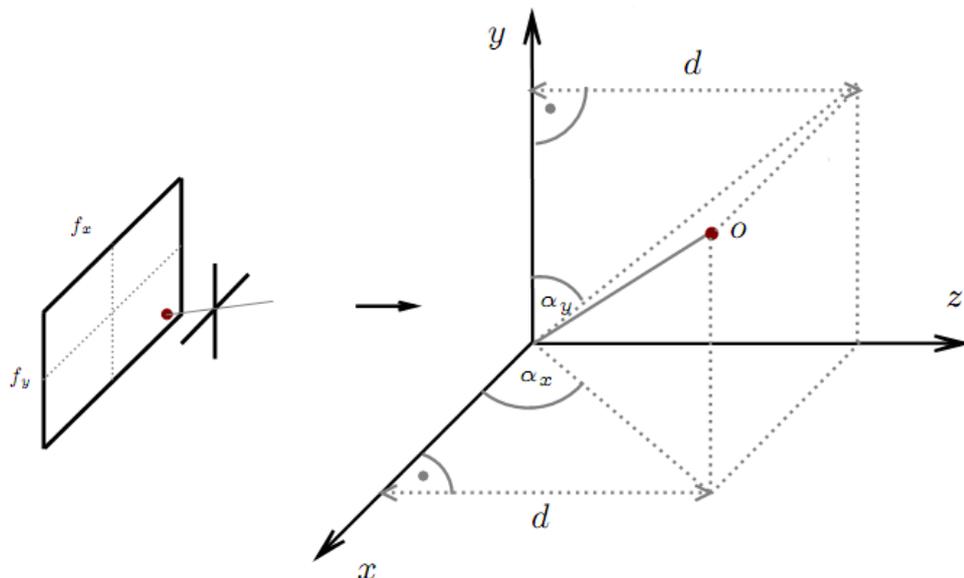


Abbildung 4.37: Position eines Punktes im 3D-Koordinatensystem der Kamera.

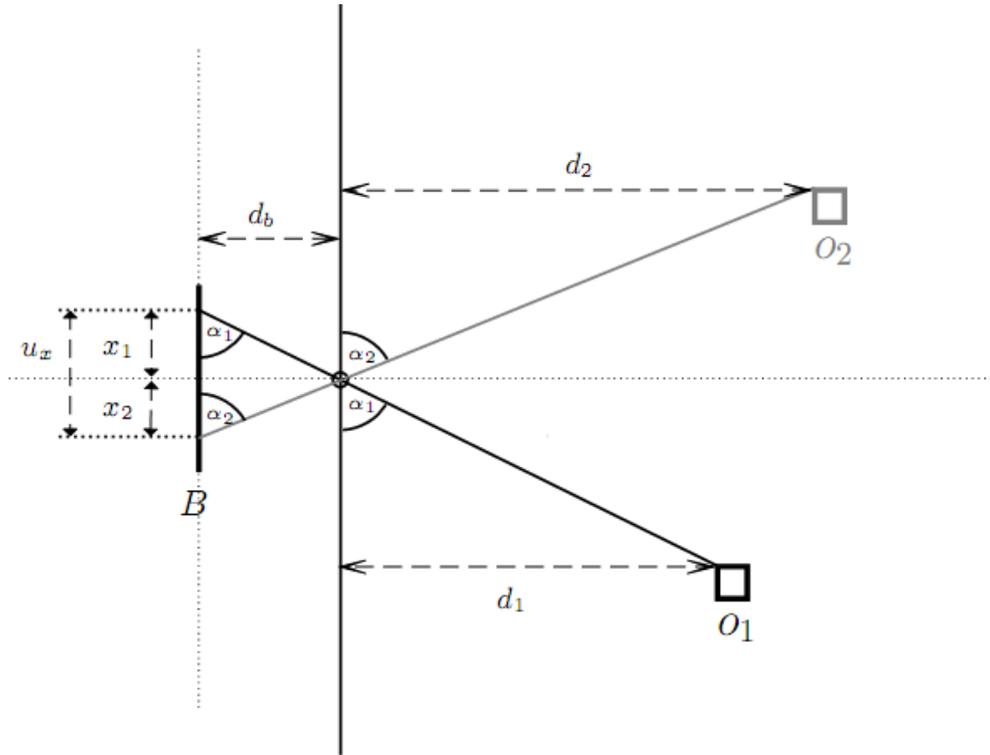


Abbildung 4.38: Verschiebung eines Objekts im 2D-Koordinatensystem der Kamera.

4.4.5 Experimente und Diskussion

In den folgenden beiden Unterabschnitten wird die hier vorgestellte XCLG-Variante zur Schätzung des Optischen Flusses mit der CLG-Variante verglichen. Dabei sind die systematische Unterschätzung durch die CLG-Methode und die Dauer der Laufzeit der Verfahren von besonderem Interesse.

Vergleich CLG/- XCLG-Methode

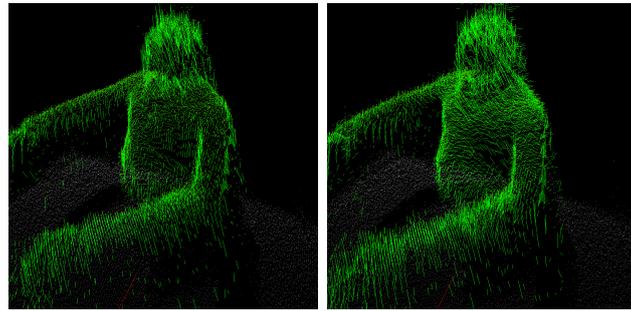
Die Abbildungen 4.41 und 4.43 zeigen die OF-Schätzungen der CLG/- und XCLG-Methode. Die roten Pfeile beschreiben die OF-Schätzung der CLG-Methode, grüne Pfeile die der XCLG-Methode. Es fällt auf, dass die OF-Vektoren der XCLG-Methode an vielen Stellen einen größeren Betrag haben.

Die Verteilung der Betragsdifferenzen ist in Abbildung 4.40 (f) und 4.42 (f) visualisiert. Die Färbung entspricht den Farben der Vektoren. In grünen Bereichen gilt bzgl. der Euklidischen Norm: $\|\mathbf{u}_{CLG}\|_2 < \|\mathbf{u}_{XCLG}\|_2$. In roten Bereichen gilt: $\|\mathbf{u}_{CLG}\|_2 > \|\mathbf{u}_{XCLG}\|_2$. Die Helligkeit wurde bzgl. der maximalen Differenz d_{\max} normalisiert.

Der Ausschnitt in den Abbildungen 4.41 und 4.43 entspricht jeweils der rechteckigen Markierung in den Abbildungen 4.40 (f) und 4.42 (f).

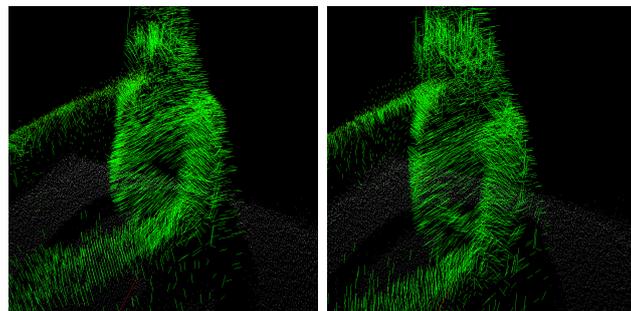
Die Tiefenbilder der PMD-Kamera (Abbildung 4.40 (a), (b)) wurden vor der OF-Berechnung mit einem Tiefpass-Filter bearbeitet. Dazu wurden beide Frames mit einer Gaußfunktion mit der Standardabweichung σ gefaltet.

Die Standardabweichung der Faltungsfunktion, welche den lokalen Bereich der OF-Schätzung



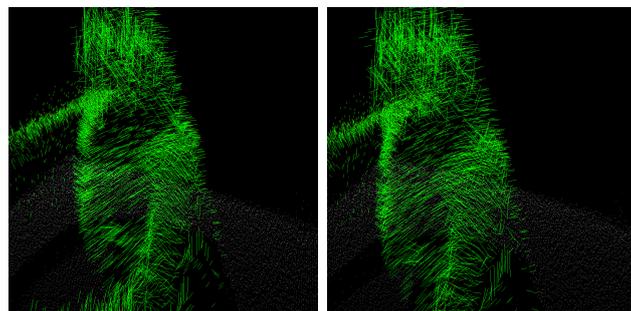
(a)

(b)



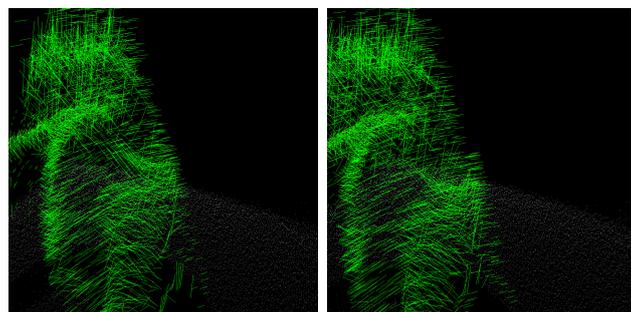
(c)

(d)



(e)

(f)



(g)

(h)

Abbildung 4.39: Ausgewählte Bilder einer Bildfolge zur Visualisierung des 3D-OF.

der CLG-Methode definiert, wird mit ρ bezeichnet.

Die OF-Schätzungen wurden mit Hilfe des *SOR*-Verfahrens durchgeführt. Die Berechnung wurde nach 1000 Iterationsschritten abgebrochen.

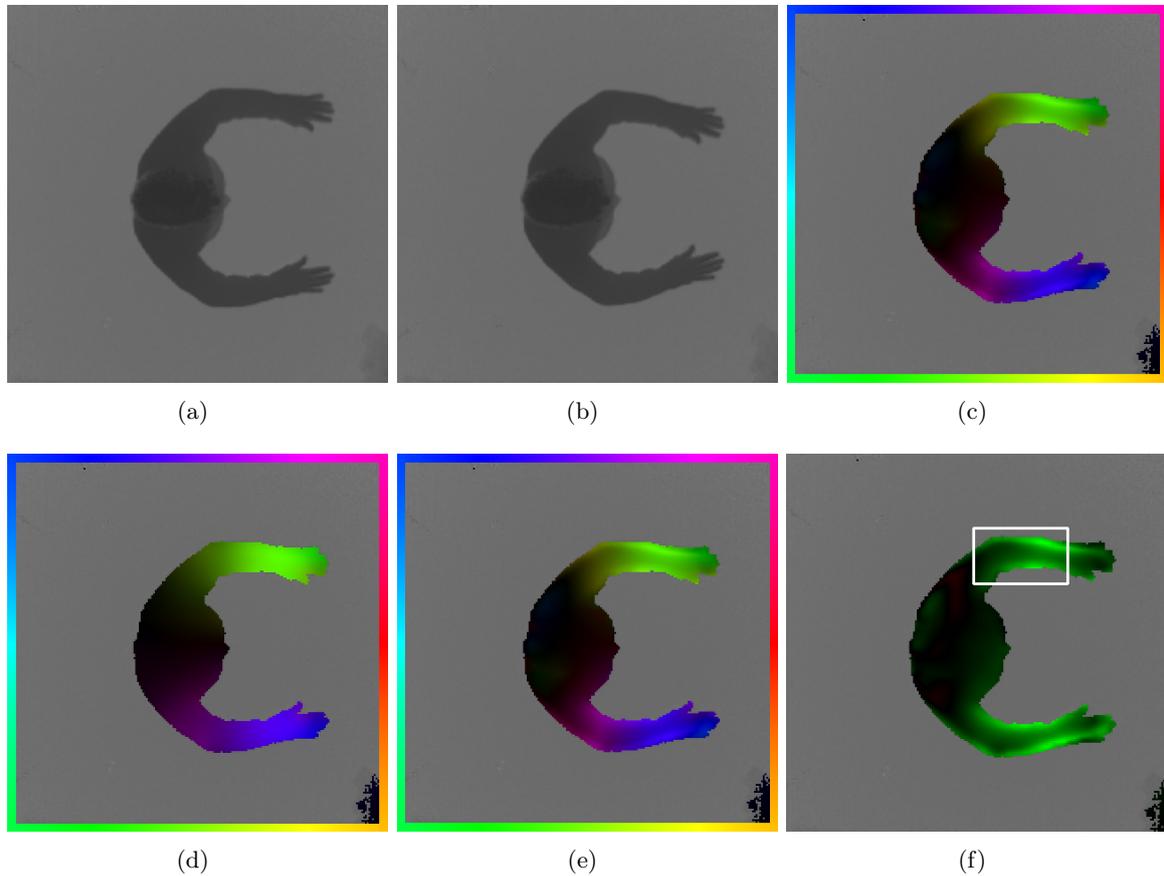


Abbildung 4.40: Stillstehender Oberkörper, mit sich aufeinander zu bewegender Arme. (a) Frame zum Zeitpunkt t . (b) Frame zum Zeitpunkt $t + 1$. (c) Ergebnis der Ballungsanalyse. (d) OF-Schätzung der CLG-Methode, mit $\rho = 2.5$. (e) OF-Schätzung der XCLG-Methode. Beide Schätzungen wurden nachträglich bzgl. des Hintergrunds segmentiert. (f) Visualisierung der Verteilung der Betragsdifferenzen, mit $d = 1.26$.

Benchmark

Die folgenden Tabellen zeigen eine Übersicht der Berechnungszeiten der CLG/- und XCLG-Methode. Dabei werden drei Bildfolgen verwendet:

1. Tiefenbilder einer PMD-Kamera mit einer Auflösung von 50×64 .
2. Grauwert/- und Tiefenbilder einer PMD-Kamera mit einer Auflösung von 204×204 .
3. Die Yosemite Sequenz¹¹ mit einer Auflösung von 316×252 .

¹¹<http://www.cs.brown.edu/black//Sequences/yosemite.tar.gz>

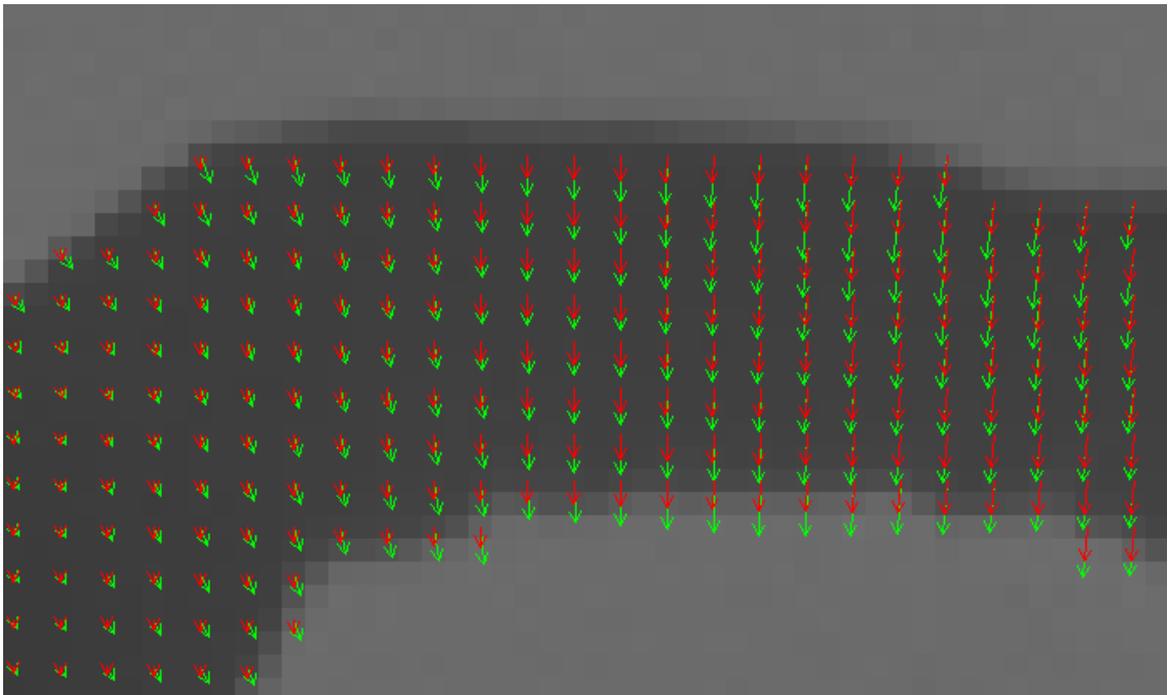


Abbildung 4.41: Vergrößerter Ausschnitt der OF-Schätzung. Der Bereich entspricht Abbildung 4.40 (f). Die roten Pfeile beschreiben die Schätzung der CLG-Methode, die grünen Pfeile die der XCLG-Methode. Der OF wurde an jeder zweiten Position in horizontale und vertikale Richtung eingezeichnet.

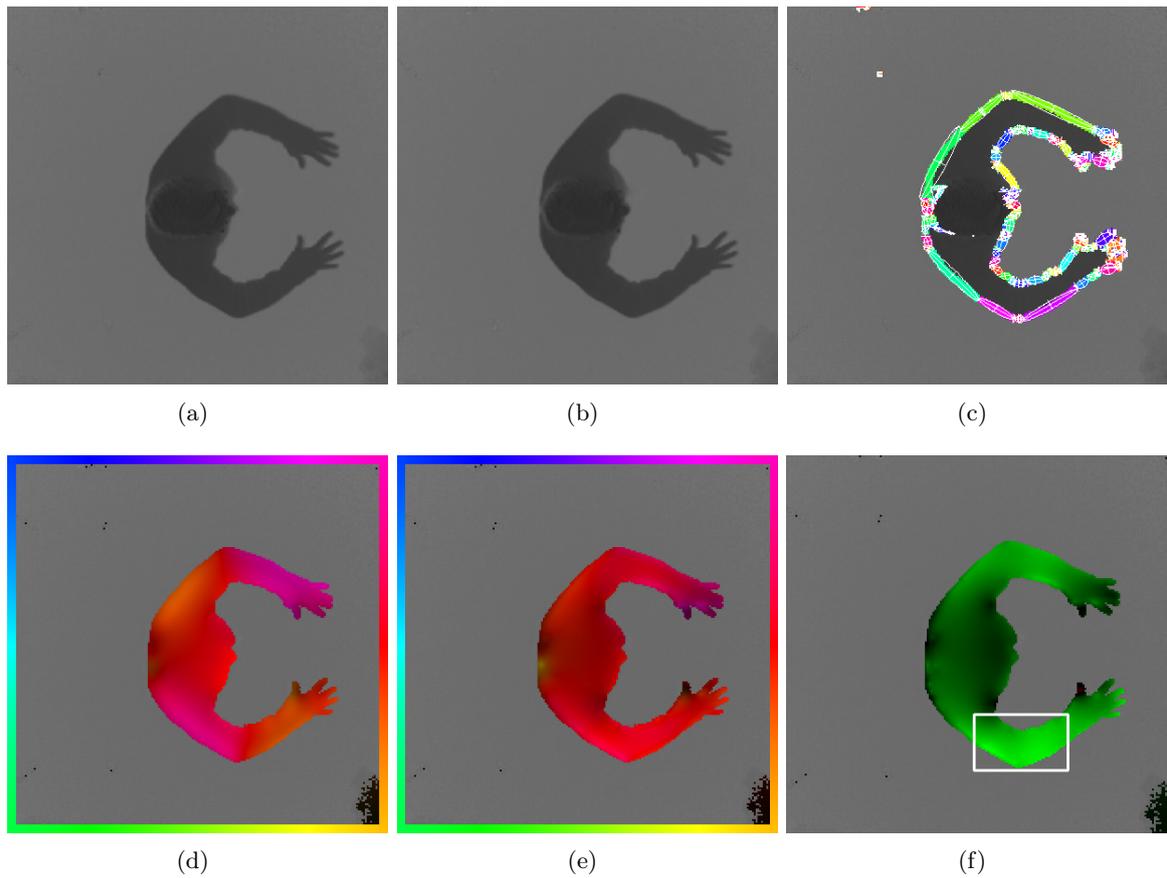


Abbildung 4.42: Homogene Bewegung des gesamten Körpers. (a) Frame zum Zeitpunkt t . (b) Frame zum Zeitpunkt $t + 1$. (c) Ergebnis der Ballungsanalyse. (d) OF-Schätzung der CLG-Methode, mit $\rho = 2.5$. (e) OF-Schätzung der XCLG-Methode. Beide Schätzungen wurden nachträglich bzgl. des Hintergrunds segmentiert. (f) Visualisierung der Verteilung der Betragsdifferenzen, mit $d = 1.47$.

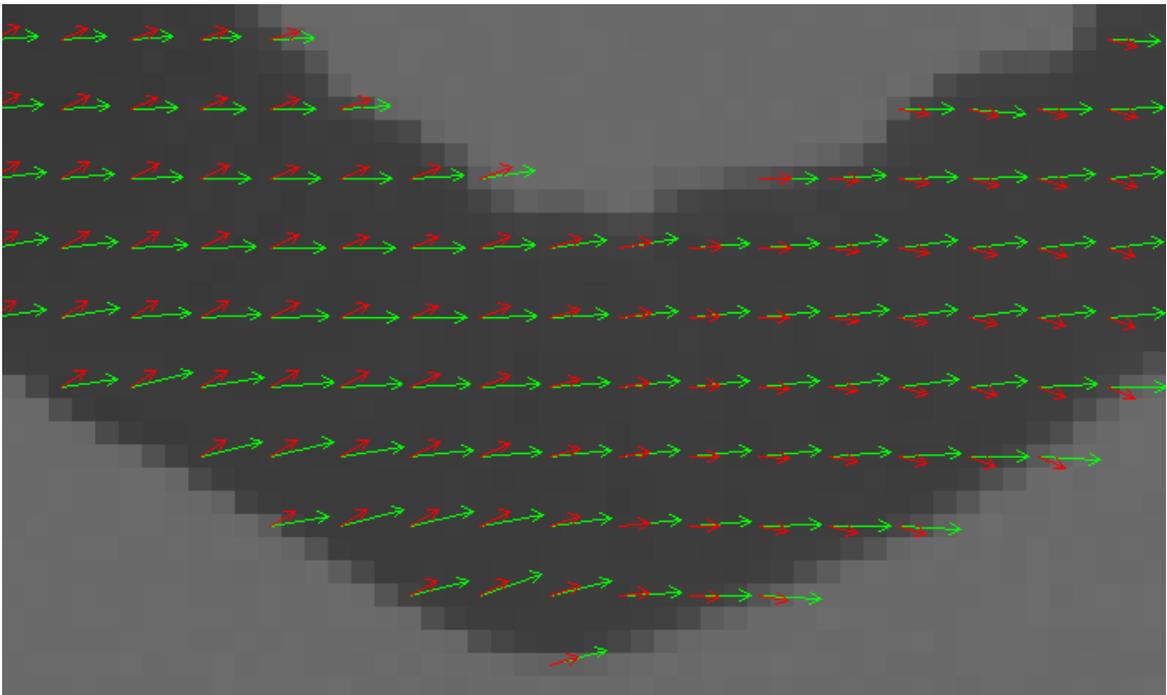


Abbildung 4.43: Vergrößerter Ausschnitt der OF-Schätzung. Der Bereich entspricht Abbildung 4.42 (f). Die roten Pfeile beschreiben die Schätzung der CLG-Methode, die grünen Pfeile die der XCLG-Methode. Der OF wurde an jeder dritten Position in horizontale und vertikale Richtung eingezeichnet.

Die Yosemite-Sequenz ist eine übliche Testsequenz im Bereich der Evaluierung des Optischen Flusses. Für das hier entwickelte Verfahren ist diese Sequenz von keiner großen Bedeutung, da es sich um eine Aufzeichnung einer nicht-stationären Kamera handelt, im Gegensatz zur stationären Kamera im Falle der sicheren MRK.

Die Berechnungen wurden auf einem PC mit einer 3.0 GHz Dual Core CPU, 4.0 GByte RAM und einer NVIDIA GTX 260 Grafikkarte durchgeführt. Alle Programmdurchläufe wurden auf der CPU und der GPU durchgeführt. Die Details zur Implementierung auf der CPU und Details bzgl. der GPU Implementierung werden in den entsprechenden Kapiteln diskutiert.

Die Tabellen 4.3 und 4.4 enthalten die Benchmarks der OF-Schätzung mithilfe der CLG- und XCLG-Methode. Bei der Berechnung mit dem SOR-Verfahren wurden je 1000 Iterationen ausgeführt. Bei der Vor/- und Nachglättung des Mehrgitter-Verfahren wurde je ein Iterationsschritt durchgeführt. Als Rekursionstiefe wurde $t_{\max} = 4$ gewählt.

	PMD _{50×64}				PMD _{204×204}			
	CLG		XCLG		CLG		XCLG	
	SOR	MG	SOR	MG	SOR	MG	SOR	MG
CPU	164	2	172	11	2150	25	2225	85
GPU	20	5	23	8	89	20	102	33

Tabelle 4.3: Benchmark: OF-Berechnung der PMD Tiefenbilder.

	Yosemite			
	CLG		XCLG	
	SOR	MG	SOR	MG
CPU	4020	48	4260	295
GPU	164	32	220	85

Tabelle 4.4: Benchmark: OF-Berechnung der Yosemite Sequenz.

Bei der Vorverarbeitung werden die Bilddaten mit einer isotropen Gauß-Funktion gefaltet. Die Tabelle 4.5 zeigt die entsprechenden Berechnungszeiten. Die Ausbreitung der Faltungsfunktion ist jeweils in Pixeln angegeben.

Die Ballungsanalyse wird auf einer Punktmenge der Mächtigkeit P ausgeführt. Das Ergebnis sind Cluster, deren Anzahl mit C bezeichnet wird. In den nachfolgenden Tabellen 4.6 und 4.7 ist die Abhängigkeit der Berechnungszeiten von diesen Werten abgebildet.

Die Erstellung des örtlich-zeitlichen-ST \tilde{J} beinhaltet mehrere Schritte. Zum einen die Analyse der Cluster zur Bestimmung der Parameter der Faltungsfunktion \tilde{K} , zum anderen die Faltung der Datenterme $\mathbf{M1}$ und $\mathbf{M4}$ mit \tilde{K} . Beide Schritte sind abhängig von der Anzahl der Cluster C und Mächtigkeit der initialen Punktmenge P (siehe Tabellen 4.8 und 4.9).

	PMD _{50×64}		PMD _{204×204}		Yosemite	
	5 × 5	15 × 15	5 × 5	15 × 15	5 × 5	15 × 15
CPU	0.1	0.2	0.8	1.8	1.5	3.4
GPU	0.1	0.6	1.4	4.3	1.8	7.7

Tabelle 4.5: Benchmark: Vorverarbeitung

	PMD _{50×64}	PMD _{204×204}
	500 [P] – 55 [C]	1900 [P] – 140 [C]
CPU	5.7	24
GPU	1.9	8.8

Tabelle 4.6: Benchmark: Ballungsanalyse der PMD Tiefenbilder

	Yosemite		
	20000 [P] – 1800 [C]	10000 [P] – 1200 [C]	2000 [P] – 230 [C]
CPU	270	125	24
GPU	65	36	12

Tabelle 4.7: Benchmark: Ballungsanalyse der Yosemite-Sequenz

	PMD _{50×64}	PMD _{204×204}
	500 [P] – 55 [C]	1900 [P] – 140 [C]
CPU	4	37
GPU	2	15

Tabelle 4.8: Benchmark: ST-Erstellung der PMD Tiefmessungen

	Yosemite		
	20000 [P] – 1800 [C]	10000 [P] – 1200 [C]	2000 [P] – 230 [C]
CPU	350	130	26
GPU	62	39	20

Tabelle 4.9: Benchmark: ST-Erstellung der Yosemite Sequenz

Im letzten Schritt der OF-Schätzung wird ein LGS gelöst, welches aus den vorangegangenen Berechnungen entstanden ist. Die Laufzeit der Berechnung ist dabei abhängig von der Auflösung der Bilddaten und der Auswahl des Lösungsverfahrens. In dieser Arbeit wurden dazu das SOR- und das Mehrgitter-Verfahren verwendet. Die Tabellen 4.10 und 4.11 zeigen die Benchmarks mit unterschiedlicher Anzahl von Iterationsschritten. Die Anzahl der Iterationsschritte des Mehrgitter-Verfahrens, zur Vor/- und Nachglättung, liegen zwischen 1 und 4. Als Rekursionstiefe wurde $t_{max} = 4$ gewählt.

	PMD_{50×64}				PMD_{204×204}			
	<i>SOR</i>			<i>MG</i>	<i>SOR</i>			<i>MG</i>
	<i>100</i>	<i>1000</i>	<i>5000</i>	<i>1 → 4</i>	<i>100</i>	<i>1000</i>	<i>5000</i>	<i>1 → 4</i>
CPU	16	163	811	1 → 2	214	2190	10735	12 → 30
GPU	2	16	84	~ 2	7	74	369	~ 4.5

Tabelle 4.10: Benchmark: Lösung des LGS – PMD Messungen

	Yosemite			
	<i>SOR</i>			<i>MG</i>
	<i>100</i>	<i>1000</i>	<i>5000</i>	<i>1 → 4</i>
CPU	423	4210	19730	23 → 56
GPU	14	137	686	~ 6.5

Tabelle 4.11: Benchmark: Lösung des LGS – Yosemite Sequenz

Kapitel 5

Robuste Merkmale zur Kinematik– und Risikoschätzung

In diesem Kapitel werden Merkmale für die teilweise Rekonstruktion einer menschlichen Kinematik und Merkmale für die später im nächsten Kapitel 6 folgende Risikoschätzung eingeführt. Es werden zuerst Merkmale besprochen, die es erlauben die Orientierung des Oberkörpers aus den vorverarbeiteten Bildern der Tiefenbildfolge zu schätzen. Anschließend werden die Kopforientierung und die Stellung der Schulter bestimmt. Danach folgt ein Abschnitt über die Bestimmung der Armparameter einer entsprechenden Kinematik.

Im nächsten Abschnitt wird ein im Rahmen dieser Arbeit modifizierter Kalman–Filter Rahmenwerk eingeführt, wodurch robustere Schätzergebnisse für die Rekonstruktion der Kinematik möglich werden. Darüber hinaus wird ein Hüllkörpermodell eingeführt, welches durch Aneinanderreihung von Kugeln entlang der geschätzten Kinematik hervorgeht. Das Kugelmodell wird dazu verwendet die Ergebnisse der Bewegungs– und der Kinematikschätzung experimentell zu bewerten indem durch Anwendung von geometrischen Formalismen eine Rückprojektion in den Bildbereich der Kamera stattfindet, wodurch sich die originalen Bilddaten mit dem projizierten Kinematikmodell auf die Bildebene der Kamera vergleichen lassen und damit eine Aussage über den Verdeckungsgrad quantifizieren lässt.

5.1 Merkmale für die Risikoquantifizierung

Die bisher dargestellte Verfolgung von Personen basierte auf der systematischen Auswertung von Tiefenbildfolgen durch die Methode der dynamischen Konturen. Diese Vorgehensweise wurde motiviert durch das in Abschnitt 3.1 vorgestellte Konzept. Es ist möglich, auf dieser Grundlage eine sichere Mensch–Roboter–Interaktion zu realisieren (siehe [Graf und Wörn 2008c]). Allerdings stößt man bei dieser Vorgehensweise schnell an Grenzen, insbesondere beim Versuch, eine komplexe Kooperation zu realisieren. Deshalb wird im Abschnitt 3.2 vorgeschlagen, die menschliche Pose durch Auswertung der Tiefenbilder zu rekonstruieren. Die Vorteile werden später in Abschnitt 6 diskutiert. Hier soll die Extraktion wichtiger Merkmale zur Rekonstruktion der menschlichen Pose, zumindest für ausgewählte kinematische Parameter, erklärt werden. Schließlich ist es möglich, markerlos und in Echtzeit, die Orientierung des Kopfes, des Oberkörpers, sowie die Position der Person im Raum, aber insbesondere auch die wichtigen Gelenkwinkelstellungen der Arme zu schätzen.

Mit nur einer einzelnen Tiefenbildkamera ist es offensichtlich nicht möglich, die gesamte

menschliche Pose zu rekonstruieren. Deshalb ist es sinnvoll, abhängig von der speziellen Anforderung der Aufgabenstellung, sich darüber klar zu werden, welche kinematischen Parameter von großer Bedeutung sind.

Für den Fall der Mensch–Roboter–Kooperation und Interaktion wird davon ausgegangen, dass sich der Mensch aufrecht in der Szene bewegt. Der erste Parameter, den man schätzen möchte, ist die Position des Menschen im Raum. Des Weiteren ist es von großer Bedeutung, die Armstellung rekonstruieren zu können. Von großem Vorteil ist auch die Kenntnis über die Oberkörperorientierung und evtl. die Orientierung des Gesichtsfeldes.

Es wurde deshalb für den Aufbau des Demonstrators entschieden, die Kamera an der Decke anzubringen. Dadurch besteht die Möglichkeit, die eben aufgezählten kinematischen Punkte aus den Tiefenbildern zu ermitteln. Außerdem besteht bei dieser Anordnung der Kamera die Möglichkeit, klassische Methoden der Bildverarbeitung und –auswertung auf die Bildfolgen anzuwenden.

5.1.1 Lokalisation von menschlichen Körpern

Die Position einer aufrecht gehenden Person lässt sich mit Hilfe einer Tiefenbildkamera schätzen, indem zuerst die Lage der Schulter ermittelt wird. Stößt man anschließend virtuell eine Achse ausgehend vom Schwerpunkt der Schulter auf den Boden, so dass die Achse den Boden orthogonal trifft, lässt sich auf diese Weise die Position der Person im Raum ermitteln.

5.1.2 Lage von Kopf und Schulter

Es wurde bereits angedeutet, dass ein wichtiges Merkmal für eine sichere Mensch–Roboter–Interaktion der Abstand des Menschen zum Roboter ist. Die meisten Forschungsgruppen im Bereich der sicheren Mensch–Roboter–Kooperation haben dies erkannt und realisieren eine Architektur, welche es ermöglicht, in Abhängigkeit des Abstandes die Arbeitsgeschwindigkeit des Roboters anzupassen. Die Vorgehensweise dieser Arbeit geht einen Schritt weiter, indem versucht wird, neben der Distanz zwischen Mensch und Roboter weitere Merkmale zu berücksichtigen. Wenn eine Kooperation stattfinden soll, muss der Roboter deshalb mehr Informationen erhalten. Es wird deshalb vorgeschlagen, zusätzlich die Orientierung des Kopfes und der Schulter zu ermitteln.

Die beiden Merkmale dienen als Grundlage, um zu ermitteln, ob der Mensch dem Roboter seine Aufmerksamkeit schenkt. Es besteht ein großer Unterschied in Bezug auf das Risiko für den Menschen, ob man sich nahe am Roboter befindet und dem Roboter zugewandt ist, oder ob man nahe am Roboter steht und der Oberkörper und das Gesichtsfeld vom Roboter abgewandt sind.

Diese Vorgehensweise soll später bei der Risikoquantifizierung in Abschnitt 6 genauer beschrieben werden.

Die Orientierung von Kopf und Schulter lassen sich durch eine Hauptachsenanalyse bestimmen. Dazu müssen Bildpunkte, korrespondierend zum Kopf oder der Schulter, segmentiert werden. Anschließend werden die sich ergebenden Bildpunktballungen durch Ellipsen beschrieben. Dann kann aus den Hauptachsen der berechneten Ellipsen die Orientierung des entsprechenden Körperteils bestimmt werden. Dabei gelten folgende Regeln:

- Die Kopforientierung wird bestimmt durch eine der Bildpunktballungen der Bildpunkte korrespondierend zum Kopf angenäherten Ellipse. Die Hauptachse der Ellipse, korre-

spondierend zum *größten* Eigenwert der Kovarianz der angepassten Ellipse, repräsentiert die Schätzung der Kopforientierung.

- Die Orientierung der Schulter wird bestimmt durch eine der Bildpunktballungen der Schulterbildpunkte angenäherten Ellipse. Die Hauptachse der Ellipse, korrespondierend zum *kleinsten* Eigenwert der Kovarianz der angepassten Ellipse, repräsentiert die Schätzung der Schulter–Orientierung.

Die für die Berechnung der Ellipsen notwendigen Bildpunktballungen des Kopfes und der Schultern werden durch Schwellwerte generiert. Die Schwellwerte sind angepasst an die Körpergröße der Person. Die Größe der Person wird initial beim Betreten der Zelle aus den Tiefenmessungen der Kamera ermittelt.

Die Berechnung der Ellipsen zur Beschreibung der Bildpunktballungen basiert auf Standard-techniken der Bildanalyse. Zunächst müssen dafür die Bilder binarisiert werden. Anschließend werden die Bildpunktballungen als Verteilungen interpretiert. Im Rahmen dieser Arbeit gelingt die Berechnung der Ellipsen unter Verwendung von zentralen Momenten. Dazu folgende Gleichung:

$$m_{pq} = \iint_{\Omega} f(x, y) \cdot x^p y^q dx dy \quad . \quad (5.1)$$

Die Funktion $f(\cdot, \cdot)$ dient normalerweise dazu, physikalische Eigenschaften einer Verteilung zu modellieren. Hier handelt es sich um eine konstante Funktion mit $f(x, y) = 1$, für alle definierten x und y . Die Momente der Ordnung Null beschreiben gerade das Flächenintegral des Bildbereiches Ω . Das Zentrum des Bildes lässt sich bestimmen durch $\left(\frac{m_{10}}{m_{00}}, \frac{m_{01}}{m_{00}}\right)$.

Die zentralen Momente werden beschrieben durch:

$$u_{pq} = \iint_{\Omega} f(x, y) \cdot (x - \bar{x})^p \cdot (y - \bar{y})^q dx dy \quad , \quad (5.2)$$

wobei gilt, dass

$$\begin{aligned} u_{00} &= m_{00} \\ u_{10} &= m_{01} = 0 \\ u_{20} &= m_{20} - \frac{m_{10}^2}{m_{00}} \\ u_{02} &= m_{02} - \frac{m_{01}^2}{m_{00}} \\ u_{11} &= m_{11} - \frac{m_{10}m_{01}}{m_{00}} \quad . \end{aligned} \quad (5.3)$$

Die zentralen Momente können für die Berechnung der angepassten Ellipse verwendet werden. Zur Vereinfachung folgende Berechnung vorab:

$$\sigma = \sqrt{4u_{11}^2 + (u_{20} - u_{02})^2} \quad . \quad (5.4)$$

Dann lässt sich die 1. und 2. Hauptachse bestimmen durch:

$$\begin{aligned} a_1 &= \sqrt{\frac{2(u_{20} + u_{02} + \sigma)}{u_{11}}} \quad , \\ a_2 &= \sqrt{\frac{2(u_{20} + u_{02} - \sigma)}{u_{11}}} \quad . \end{aligned}$$

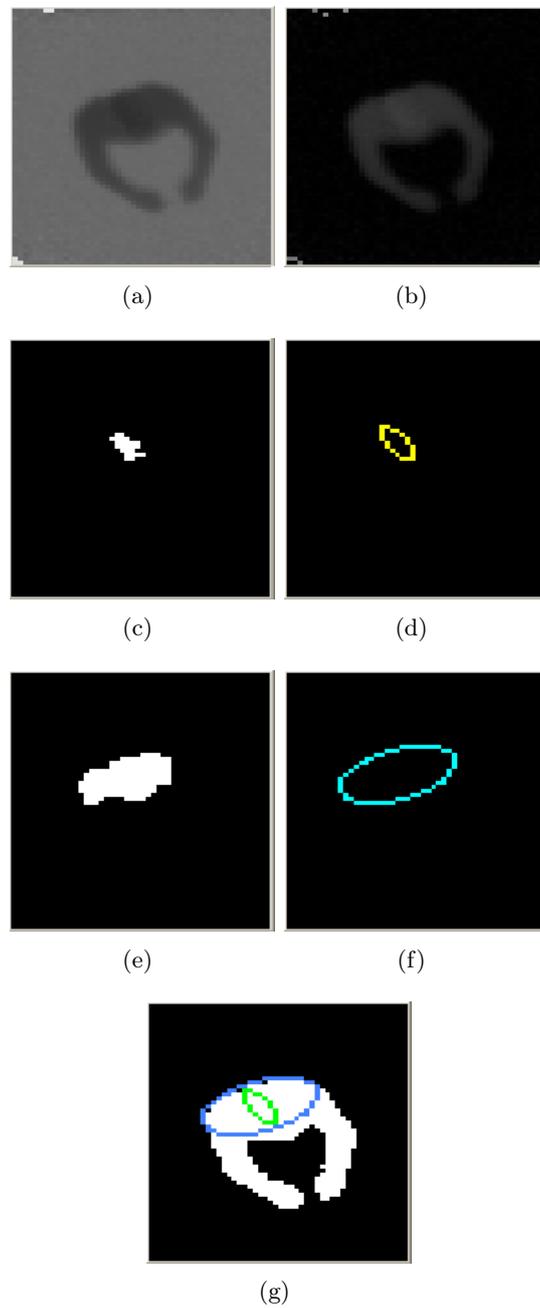


Abbildung 5.1: Beispiel zur Berechnung der Orientierung des Kopfes und der Schultern. In (a) und (b) ist das Originalbild und das segmentierte Bild zu sehen. In (c) bzw. (e) und (d) bzw. (f) sind die Bildpunkte des segmentierten Körperteils und die an die Daten angepasste Ellipse zu sehen. Abbildung (g) fusioniert die Resultate.

Die Orientierung lässt sich durch

$$\Phi = \arctan\left(\frac{2u_{11}}{u_{20} - z_{02}}\right) \quad (5.5)$$

bestimmen.

Um eine korrekte Darstellung der Ellipse zu erzielen, müssen die Eigenwerte renormalisiert werden.

5.1.3 Kinematische Parameter der Arme

Für die Berechnung der Merkmale zur Bestimmung der Stellung der Arme, wird eine andere Vorgehensweise gewählt. Dabei wird eine Strategie zur Auflösung von sich verdeckenden Armen mitberücksichtigt. Es wird ein Verfahren vorgestellt, welches zusätzliche Informationen aus den extrahierten Merkmalen der Arme verwendet und anschließend die Problemstellung auf einen Graphen abbildet. Die Lösung wird dann auf dem Graphen berechnet.

Die wichtigen Merkmale für die Arme sind die Schulterpunkte, die Armbeugen und die Handpunkte. Sind die Merkmale zur Beschreibung der Körperpunkte gefunden, kann durch lineare Interpolation, bis auf Rotation, die Konfiguration der Arme geschätzt werden. Dies gelingt insbesondere durch die Integration der Tiefenwerte.

Da der Schulterbereich, beschrieben durch eine Ellipse, bekannt ist, lassen sich die Bildpunkte der Arme isolieren. Im Extremfall hat man zwei zusammenhängende Bildpunktballungen zur Verfügung, welche global durch einen Skelettierungsansatz vorverarbeitet werden. Für die Skelettierung wird ein in [Lam u.a. 1994] vorgestelltes Verfahren verwendet. Dieses Verfahren arbeitet zuverlässig, aber insbesondere kann es effizient implementiert werden.

Das Problem bei Verfahren der Skelettierung ist, dass im Allgemeinen eine Nachbearbeitung notwendig wird, da durch die Skelettierung sogenannte Bärte (Ausläufer die aus mehreren zusammenhängenden Bildpunkten bestehen) erzeugt werden. Die Bärte sind ein Problem für die genaue Schätzung der Armkonfiguration, da die Schätzung auf Merkmale wie End- und Kreuzungspunkte basiert und sich so unerwünschte Mehrdeutigkeiten ergeben.

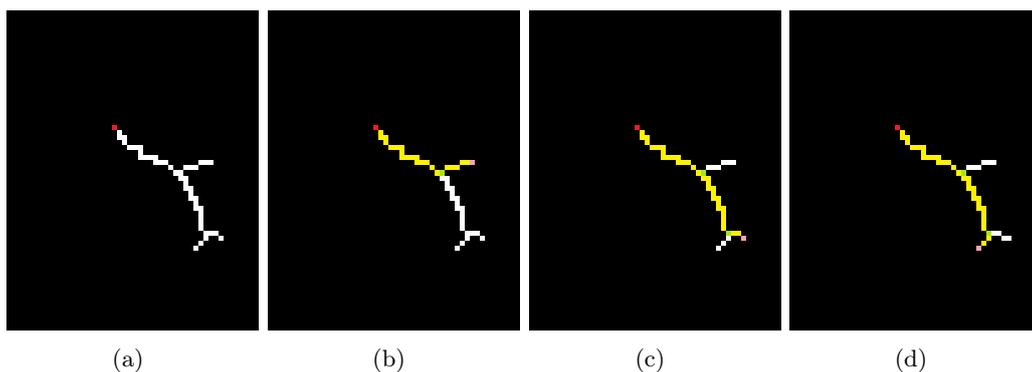


Abbildung 5.2: Sich ergebende Teilpfade nach der Skelettierung durch Einsatz des rekursiven Pfadfinders.

Um diese Mehrdeutigkeiten zu lösen, wird deshalb ein rekursives Suchverfahren, welches im Rahmen der Arbeit entwickelt wurde, eingesetzt. Das Verfahren arbeitet auf einer zusammenhängenden Menge von Bildpunkten.

Wegen stark gestörter Schätzungen kommt es vor, dass für einen Arm mehrere Endpunkte, bzw. Schulter- und Handpunkte existieren. Um eine geeignete Auswahl treffen zu können, müssen alle möglichen Pfade bekannt sein. Dazu werden, ausgehend von einem beliebigen Endpunkt, alle möglichen Pfade rekursiv durchlaufen und gespeichert (Abbildung 5.2). Das Verfahren speichert alle Pfade und Kreuzungspunkte.

Es wird angenommen, dass die Bildpunkte, korrespondierend zum Arm, einen zyklenfreien Graph bilden. Falls diese Annahme verletzt ist, bricht die rekursive Pfadsuche ab. In diesem Fall wird die letzte gültige Armkonfiguration verwendet. Details dazu werden im Abschnitt 5.2 geklärt.

Bevor das Verfahren startet, müssen die Richtungen kodiert werden (siehe Abbildung 5.3).

0	1	2
3	x	5
6	7	8

Abbildung 5.3: Richtungskodierung.

Außerdem werden noch Masken für die Bestimmung der Richtung benötigt (siehe dazu Abbildung 5.4).

<table border="1" style="border-collapse: collapse; width: 30px; height: 30px;"> <tr><td style="text-align: center;">-</td><td style="text-align: center;">1</td><td style="text-align: center;">-</td></tr> <tr><td style="text-align: center;">-</td><td style="text-align: center;">x</td><td style="text-align: center;">-</td></tr> <tr><td style="text-align: center;">-</td><td style="text-align: center;">-</td><td style="text-align: center;">-</td></tr> </table>	-	1	-	-	x	-	-	-	-	<table border="1" style="border-collapse: collapse; width: 30px; height: 30px;"> <tr><td style="text-align: center;">-</td><td style="text-align: center;">-</td><td style="text-align: center;">-</td></tr> <tr><td style="text-align: center;">-</td><td style="text-align: center;">x</td><td style="text-align: center;">1</td></tr> <tr><td style="text-align: center;">-</td><td style="text-align: center;">-</td><td style="text-align: center;">-</td></tr> </table>	-	-	-	-	x	1	-	-	-	<table border="1" style="border-collapse: collapse; width: 30px; height: 30px;"> <tr><td style="text-align: center;">-</td><td style="text-align: center;">-</td><td style="text-align: center;">-</td></tr> <tr><td style="text-align: center;">-</td><td style="text-align: center;">x</td><td style="text-align: center;">-</td></tr> <tr><td style="text-align: center;">-</td><td style="text-align: center;">1</td><td style="text-align: center;">-</td></tr> </table>	-	-	-	-	x	-	-	1	-	<table border="1" style="border-collapse: collapse; width: 30px; height: 30px;"> <tr><td style="text-align: center;">-</td><td style="text-align: center;">-</td><td style="text-align: center;">-</td></tr> <tr><td style="text-align: center;">1</td><td style="text-align: center;">x</td><td style="text-align: center;">-</td></tr> <tr><td style="text-align: center;">-</td><td style="text-align: center;">-</td><td style="text-align: center;">-</td></tr> </table>	-	-	-	1	x	-	-	-	-
-	1	-																																					
-	x	-																																					
-	-	-																																					
-	-	-																																					
-	x	1																																					
-	-	-																																					
-	-	-																																					
-	x	-																																					
-	1	-																																					
-	-	-																																					
1	x	-																																					
-	-	-																																					
(a)	(b)	(c)	(d)																																				
<table border="1" style="border-collapse: collapse; width: 30px; height: 30px;"> <tr><td style="text-align: center;">1</td><td style="text-align: center;">0</td><td style="text-align: center;">-</td></tr> <tr><td style="text-align: center;">0</td><td style="text-align: center;">x</td><td style="text-align: center;">-</td></tr> <tr><td style="text-align: center;">-</td><td style="text-align: center;">-</td><td style="text-align: center;">-</td></tr> </table>	1	0	-	0	x	-	-	-	-	<table border="1" style="border-collapse: collapse; width: 30px; height: 30px;"> <tr><td style="text-align: center;">-</td><td style="text-align: center;">0</td><td style="text-align: center;">1</td></tr> <tr><td style="text-align: center;">-</td><td style="text-align: center;">x</td><td style="text-align: center;">0</td></tr> <tr><td style="text-align: center;">-</td><td style="text-align: center;">-</td><td style="text-align: center;">-</td></tr> </table>	-	0	1	-	x	0	-	-	-	<table border="1" style="border-collapse: collapse; width: 30px; height: 30px;"> <tr><td style="text-align: center;">-</td><td style="text-align: center;">-</td><td style="text-align: center;">-</td></tr> <tr><td style="text-align: center;">-</td><td style="text-align: center;">x</td><td style="text-align: center;">0</td></tr> <tr><td style="text-align: center;">-</td><td style="text-align: center;">0</td><td style="text-align: center;">1</td></tr> </table>	-	-	-	-	x	0	-	0	1	<table border="1" style="border-collapse: collapse; width: 30px; height: 30px;"> <tr><td style="text-align: center;">-</td><td style="text-align: center;">-</td><td style="text-align: center;">-</td></tr> <tr><td style="text-align: center;">0</td><td style="text-align: center;">x</td><td style="text-align: center;">-</td></tr> <tr><td style="text-align: center;">1</td><td style="text-align: center;">0</td><td style="text-align: center;">-</td></tr> </table>	-	-	-	0	x	-	1	0	-
1	0	-																																					
0	x	-																																					
-	-	-																																					
-	0	1																																					
-	x	0																																					
-	-	-																																					
-	-	-																																					
-	x	0																																					
-	0	1																																					
-	-	-																																					
0	x	-																																					
1	0	-																																					
(e)	(f)	(g)	(h)																																				

Abbildung 5.4: Richtungsmasken. Die Bedeutung der Matrixeinträge ist folgendermaßen: '0' → Bildpunkt darf kein Element des Pfades sein, '1' → Bildpunkt muss Element des Pfades sein und '-' → Bildpunkt kann Element des Pfades sein.

Die folgenden Parameter bzw. Referenzen:

- vekPfade
- vekKreuzungspunkte
- aktPosition

- vorRichtung
- vorRichtung
- aktPfadnummer
- abbruchBedingung
- anzSchritte

müssen dem Algorithmus 5 übergeben werden.

Um eine Verdeckungsaufösung der Arme durchführen zu können, müssen vereinfachende Annahmen getroffen werden. Dazu folgende Konfigurationen:

1. Es existieren zwei Armsegmente (siehe Abbildung 5.5(a)).
2. Es existiert ein Armsegment ohne Kreuzungspunkte (siehe Abbildung 5.5(b)).
3. Es existiert ein Armsegment mit mindestens einem Kreuzungspunkt (siehe Abbildung 5.5(c))

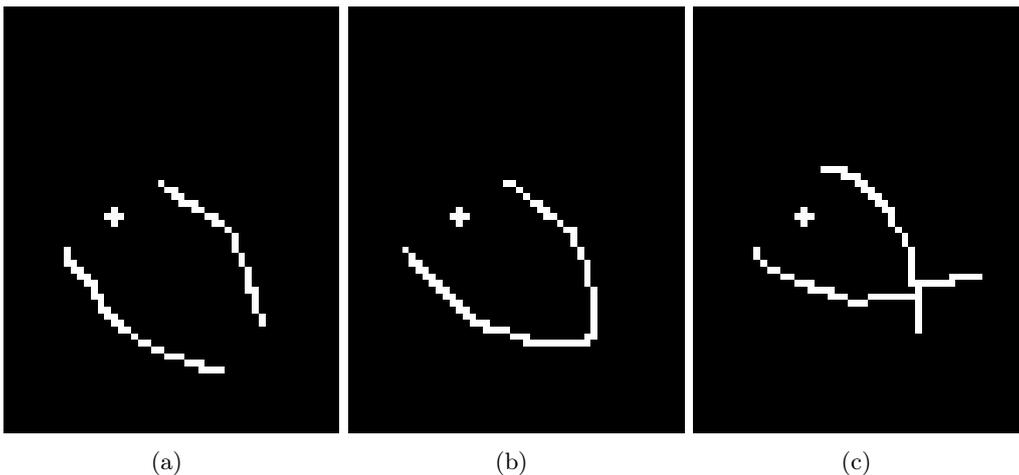


Abbildung 5.5: Drei mögliche Konfigurationen von Armstellungen nach der Segmentierung.

Unter einem Armsegment versteht man eine zusammenhängende Menge von Bildpunkten, welche den Armen zugeordnet werden. Ein Kreuzungspunkt existiert dann, wenn es in einem Punkt mehr als zwei mögliche Wege gibt.

Der Algorithmus macht die naheliegende Annahme, dass eine Verdeckung bzw. Überkreuzung der Arme nur für die Konfigurationen (b) und (c) in Abbildung 5.5 existiert. Für die Konfiguration (a) in Abbildung 5.5 werden zwei Arme angenommen und der Algorithmus zur Behandlung von Verdeckungen wird nicht angewandt. Enthalten die Grunddaten eine der beiden Konfigurationen (b) oder (c) in Abbildung 5.5, testet der Algorithmus, ob das Armsegment einem oder zwei Armen zuzuordnen ist. Dazu werden alle Endpunktpaare auf die beiden folgenden Kriterien getestet (Abbildung 5.6):

1. Liegen beide Punkte in einem bestimmten Bereich um das Körperzentrum bzw. der Mitte des Kopfes.

Algorithmus 5: rekursivePfadsuche()

Require: Initialisierung()

Ensure: Liste von Teilpfaden.

```
    aktPfad = vekPfade[ aktPfadnummer ]
2: aktPfade.push( aktPosition )
   while true do
4:     anzSchritte++
       Abbruch, falls anzSchritte > abbruchBedingung
6:     Nachbarschaftsmaske um letztes Element von aktPfade erstellen
       anzRichtungen = 0
8:     for alle Richtungen do
           Teste auf mögliche weiterführende Richtung mit Hilfe der Nachbarschaftsmaske
           und den Richtungsmasken
10:    end for
       for alle Richtungen  $\neq$  vorRichtung do
12:        if Richtung ist möglich then
           anzRichtungen++
           // Zuerst gefundene Richtungen merken.
14:         if anzRichtungen = 1 then
             ersteRichtung = Richtung
16:         end if
           // Kreuzungspunkt speichern.
           if anzRichtungen = 2 then
18:             vekKreuzungspunkte.push( Position )
           end if
           // Wenn Kreuzung existiert, dann rekursiver Aufruf.
20:         if anzRichtungen > 1 then
             vekPfade.push( neuer Pfad )
22:             Kopiere aktuellen Pfad in den neu erstellten
             neuePosition = letztes Element von aktueller Pfadrichtung + Richtung
24:             vorRichtung = inverse(Richtung)
             neuePfadnummer = aktPfadnummer++
26:             rekursivePfadsuche( vekPfade, vekKreuzungspunkte, neuePosition,
             vorPosition, neuePfadnummer, abbruchBedingung, anzSchritte )
           end if
28:         end if
       end for
       // Test auf Endpunkt.
30:       if anzRichtungen = 0 then
           Abbruch
32:       end if
       // Pfad entlang der gefundenen Richtung weiterverfolgen.
       neuePosition = letztes Element von aktPfade + ersteRichtung
34:       aktPfadvektor.push( neuePosition )
   end while
```

2. Liegen beide Punkte auf jeweils verschiedenen Seiten der Hyperebene (hier eine Gerade, siehe bspw. Abbildung 5.6).

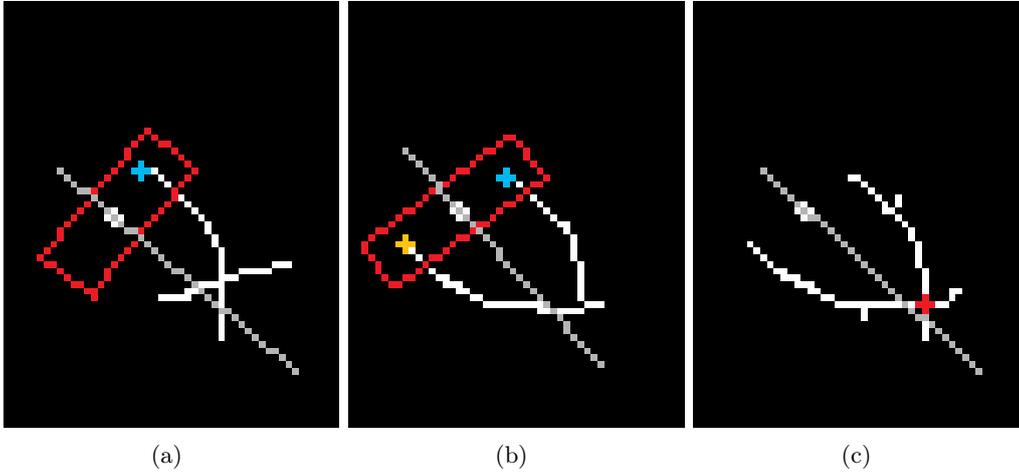


Abbildung 5.6: (a) Annahme, dass keine Verdeckung vorliegt. (b) Annahme, dass eine Verdeckung vorliegt. (c) Schätzung des Armkreuzungspunktes.

Erfüllt kein Paar beide Kriterien, stoppt der Algorithmus und die Berechnung wird nur für einen Arm weiter fortgeführt. Existiert mindestens ein Paar, für welches beide Kriterien erfüllt sind, wird von einer Überdeckung ausgegangen. In diesem Fall wird zwischen den Konfigurationen (b) und (c) in Abbildung 5.6 unterschieden.

Liegt die Konfiguration aus Abbildung 5.6(b) vor, existiert nur ein Pfad, welcher in der Mitte geteilt wird. Die beiden Pfadhälften werden dann je einem Arm zugeordnet. Ein solches Ergebnis entsteht bspw. dann, wenn sich beide Hände berühren.

Für die Konfiguration aus Abbildung 5.6(c) müssen weitere Schritte unternommen werden, um die korrespondierenden Endpunkte zu finden. Dafür wird zuerst der Kreuzungspunkt der Arme geschätzt. Als Kreuzungspunkt wird derjenige angenommen, welcher am nächsten zur Sichtlinie liegt (siehe Abbildung 5.6(c)). Bis dahin gibt es lediglich eine Schätzung für die Schulterpunkte beider Arme und den Kreuzungspunkt. Es fehlen zur vollständigen Beschreibung der Arme noch die Handpunkte und die Ellbogenpunkte. Die Ellbogenpunkte werden unabhängig vom Algorithmus, aus den festgestellten Schulterpunkten, den Handpunkten und den Pfaden der Arme bestimmt.

Um die Handpunkte zu schätzen, werden in einem weiteren Schritt zwei weitere Hyperebenen für die Klassifikation verwendet.

Die erste Hyperebene geht durch das Körperzentrum und den geschätzten Kreuzungspunkt. Die andere Hyperebene verläuft durch den Kreuzungspunkt und steht orthogonal auf der ersten Hyperebene. Die korrespondierenden Hand- und Schulterpunkte liegen jeweils auf den gegenüberliegenden Seiten der Hyperebenen (Abbildung 5.7(a)). Wird einem Schulterpunkt kein Handpunkt zugeordnet, wird für den Handpunkt der bisher angenommene Kreuzungspunkt verwendet (Abbildung 5.7(b)). Ein solches Ergebnis entsteht beispielsweise durch das Auflegen der Hand auf dem anderen Arm.

Während der Ausführung des Algorithmus, sind stets alle Pfade und Kreuzungspunkte innerhalb des Armsegments bekannt. Diese Informationen werden in einem rekursiven Ansatz aus den Grunddaten bestimmt (Abbildung 5.8).

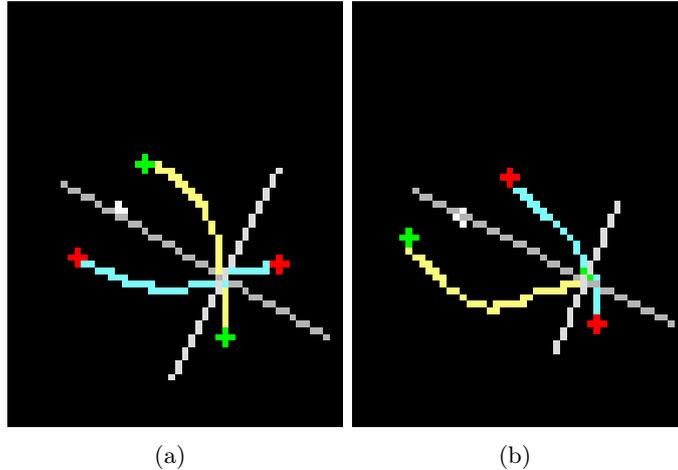


Abbildung 5.7: (a) Resultat einer geschätzten Kreuzung der Arme. (b) Resultat einer Armverdeckung.

Die Ellbogen werden gefunden, indem zwischen den Punkten der Schulter und Hand eine Gerade gezogen wird. Der Punkt der maximalen Krümmung wird als das Ellbogengelenk definiert. Gefunden wird dieser Punkt, indem zwischen dem Graphen und der Schulter-Hand-Geraden ein Lot gefällt wird. Die Position mit dem maximalen Abstand zwischen dem Graph und der Gerade definiert den gesuchten Punkt. Sind die Arme gestreckt, wird kein solcher Punkt gefunden. Der sich aus der Schätzung ergebende Winkel für die Beugung des Ellbogens ist gerade dann gleich dem Winkel des Oberarms bzw. Unterarms.

5.2 Robuste Merkmale

Die bisher beschriebene Vorgehensweise zur Berechnung von Positionsdaten und Winkelstellungen lässt sich effizient umsetzen. Das Problem sind allerdings Störeinflüsse, welche die Schätzung der Merkmale negativ beeinflussen. Die Störungen stammen, neben den systematischen Störungen der Kamera, auch von den Quantisierungseffekten bei der Digitalisierung. Die hier vorgestellten Methoden der Bildverarbeitung reagieren empfindlich auf derartige Effekte. Deshalb wird in [Graf und Wörn 2009a] vorgeschlagen, die Schätzungen der Merkmale zu robuster zu machen, indem ein modifizierter Kalman-Filter für jeden zu schätzenden Parameter eingesetzt und geeignet parametrisiert wird.

Der Kalman-Filter wird zur Glättung, Filterung und Prädiktion eingesetzt. Die drei Begriffe unterscheiden sich bezüglich der verwendeten Messdaten und des Zeitpunktes zur Schätzung des Zustands eines Systems.

Bei der Glättung wird versucht, einen zeitlich zurückliegenden Zustand zu interpolieren. Bei der Prädiktion wird versucht, einen in der Zukunft liegenden Zustandsvektor zu extrapolieren. Bei der Filterung ist man bestrebt, den gegenwärtigen Zustandsvektor mit vergangenen Messwerten zu schätzen.

Der Kalman-Filter ist ein optimaler linearer Filter. Optimale Schätzer erfüllen die drei Kriterien:

1. Erwartungstreue (minimaler Bias).

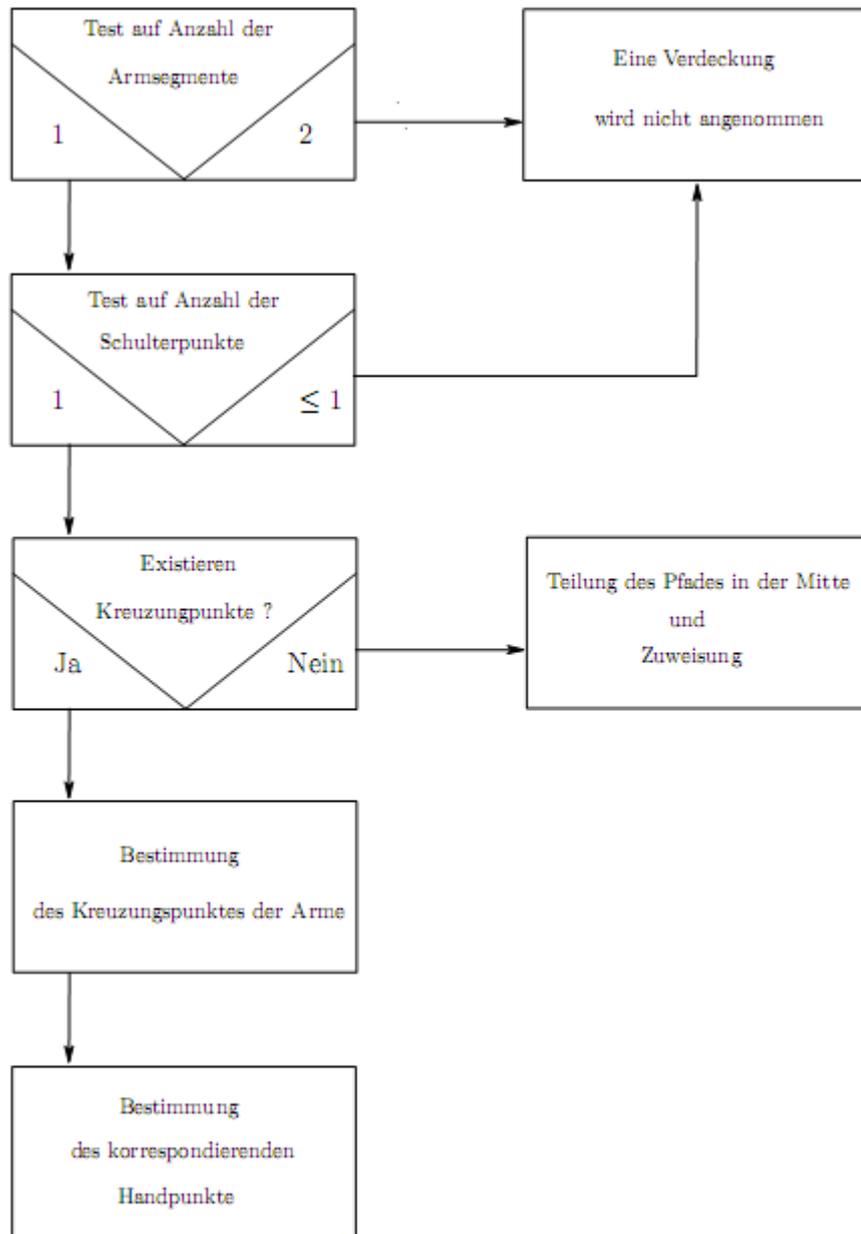


Abbildung 5.8: Analyse der Armstellungen.

2. Minimale (erwartungstreue) Varianz.

3. Konsistenz.

Schreibt man den mittleren quadratischen Fehler um, wird die Wichtigkeit der ersten beiden Kriterien klar. Es gilt, dass:

$$E\{(\hat{\mathbf{x}} - \mathbf{x})^2\} = \text{Var}(\hat{\mathbf{x}}) + \text{bias}^2(\hat{\mathbf{x}}) \quad .$$

Dabei ist $\hat{\mathbf{x}}$ der geschätzte Zustandsvektor und \mathbf{x} der tatsächliche Zustandsvektor. Die Konsistenz fordert:

$$\lim_{i \rightarrow \infty} P\{\mathbf{x}_i - \hat{\mathbf{x}}_i > \epsilon: \epsilon > 0 \text{ beliebig}\} = 0 \quad .$$

Um den Kalman-Filter verwenden zu können, sind die in der Tabelle 5.1 angegebenen Formeln implementiert worden.

Modellgleichungen	
System-Modell	$\mathbf{x}_k = F_{-1}\mathbf{x}_{k-1} + \mathbf{w}_{k-1}$, mit $\mathbf{w}_{k-1} \sim N(\mathbf{0}, Q_{k-1})$
Mess-Modell	$\mathbf{z}_k = H_k\mathbf{x}_k + \mathbf{v}_k$, mit $\mathbf{v}_k \sim N(\mathbf{0}, R_k)$
Voraussetzungen	
Initiale Belegung	$E[\mathbf{x}_0] = \hat{\mathbf{x}}_0$, $E[(\mathbf{x}_0 - \hat{\mathbf{x}}_0)(\mathbf{x}_0 - \hat{\mathbf{x}}_0)^T] = P_0$
Annahmen	$E[\mathbf{w}_i\mathbf{v}_j^T] = 0$, für beliebige i, j
Extrapolation	
Systemzustandsvektor (Extrapol.)	$\hat{\mathbf{x}}_k^- = F_{k-1}\hat{\mathbf{x}}_{k-1}^+$
Fehlerkovarianzmatrix (Extrapol.)	$P_k^- = F_{k-1}P_{k-1}^+F_{k-1}^T + Q_{k-1}$
Aktualisierungsschritt	
Systemzustandsvektors (Aktual.)	$\hat{\mathbf{x}}_k^+ = \hat{\mathbf{x}}_k^- + K_k[\mathbf{z}_k - H_k\hat{\mathbf{x}}_k^-]$
Fehlerkovarianzmatrix (Aktual.)	$P_k^+ = [I - K_kH_k]P_k^-$
Kalman-Gewichtungsmatrix	$K_k = P_k^-H_k^T[H_kP_k^-H_k^T + R_k]^{-1}$

Tabelle 5.1: Die diskreten Kalman-Formeln, modifiziert nach [Gelb 1974].

Die Vorgehensweise des Kalman-Filters sieht vor, dass man in einem Zeitpunkt k den Systemzustandsvektor $\hat{\mathbf{x}}_k^+$ bestimmen möchte. Dazu benötigt man die noch zu bestimmende Kalman-Gewichtungsmatrix K_k . Um K_k zu bestimmen, muss die Spur der Fehlerkovarianzmatrix P_k^+ minimiert werden. Dies geschieht, indem man $\frac{\partial}{\partial K_k} \text{Spur}\{P_k^+\}$ berechnet. Aufgrund der Tatsache, dass in P_k^+ die Kalman-Gewichtungsmatrix in quadratischer Form auftritt, lässt sich die

abgeleitete Form gleich Null setzen und anschließend nach K_k auflösen. Anschließend wird K_k in P_k^+ eingesetzt und damit eine aktualisierte Fehlerkovarianzmatrix erzeugt. Der letzte Schritt besteht in der Extrapolation der relevanten Größen zum nächstliegenden Zeitpunkt. Damit erhält man $\hat{\mathbf{x}}_{k+1}^-$, P_{k+1}^- . Zum neuen Zeitpunkt werden die eben aufgeführten Arbeitsschritte wiederholt.

Diese qualitative Vorgehensweise soll im Folgenden durch Herleitungen gezeigt werden. Die hier gewählten Herleitungen folgen den Anweisungen aus [Gelb 1974]. Dort wird gefordert, dass

$$\hat{\mathbf{x}}_k^+ = K_k' \tilde{\mathbf{x}}_k^- + K_k \mathbf{z}_k \quad (5.6)$$

gilt. Um die beiden zeitvarianten Matrizen K_k' und K_k zu bestimmen, benötigt man zuerst folgende Definitionen:

$$\hat{\mathbf{x}}_k^+ := \mathbf{x}_k + \tilde{\mathbf{x}}_k^+ \quad , \quad (5.7)$$

$$\hat{\mathbf{x}}_k^- := \mathbf{x}_k + \tilde{\mathbf{x}}_k^- \quad . \quad (5.8)$$

Setzt man die Gleichung (5.8) in die Gleichung (5.6) ein, so erhält man nach einigen Umformungen:

$$\tilde{\mathbf{x}}_k^+ = [K_k' + K_k H_k - I] \mathbf{x}_k + K_k' \tilde{\mathbf{x}}_k^- + K_k \mathbf{v}_k \quad . \quad (5.9)$$

Per Definition gilt $E\{\mathbf{v}_k\} = 0$. Dann muss für erwartungstreuere $\tilde{\mathbf{x}}_k^+$, d.h. $E\{\tilde{\mathbf{x}}_k^+\} = 0$ gelten, dass $[K_k' + K_k H_k - I] = 0$ ist. Daraus folgt:

$$K_k' = I - K_k H_k \quad (5.10)$$

$$\Rightarrow \hat{\mathbf{x}}_k^+ = (I - K_k H_k) \tilde{\mathbf{x}}_k^- + K_k \mathbf{z}_k \quad (5.11)$$

$$\Rightarrow \tilde{\mathbf{x}}_k^+ = \tilde{\mathbf{x}}_k^- + K_k [\mathbf{z}_k - H_k \tilde{\mathbf{x}}_k^-] \quad (5.12)$$

$$\Rightarrow \tilde{\mathbf{x}}_k^+ = (I - K_k H_k) \tilde{\mathbf{x}}_k^- + K_k \mathbf{v}_k \quad . \quad (5.13)$$

Im vorletzten Schritt wird die noch unbekannte Kalman-Gewichtungsmatrix K_k benötigt. Die Bestimmung gelingt mit Hilfe der Fehlerkovarianzmatrix P_k^+ . Es ist:

$$P_k^+ = E\{\tilde{\mathbf{x}}_k^+ \tilde{\mathbf{x}}_k^{+T}\} \quad (5.14)$$

$$\Leftrightarrow P_k^+ = (I - K_k H_k) P_k^- (I - K_k H_k)^T + K_k R_k K_k^T \quad .$$

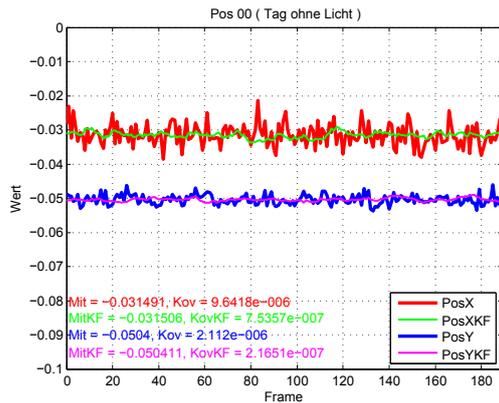
Durch $\frac{\partial}{\partial K_k} P_k^+ = 0$ und anschließender Umformung erhält man:

$$K_k = P_k^- H_k^T [H_k P_k^- H_k^T + R_k]^{-1} \quad . \quad (5.15)$$

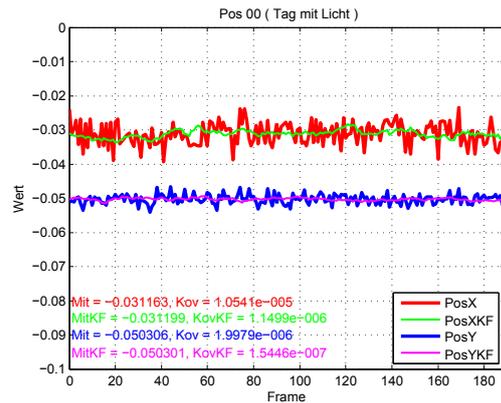
Die so gewonnene Darstellung der Kalman-Gewichtungsmatrix erlaubt eine qualitative Diskussion über das Verhalten des Kalman-Filters. Die in der Kalman-Gewichtungsmatrix auftauchende Messkovarianz R wird im Rahmen dieser Arbeit verwendet, um in Abhängigkeit von gemessenen Veränderungen die Kalman-Gewichtung neu zu bewerten.

5.2.1 Ein Kalman-Filter für Kinematikkonfigurationen

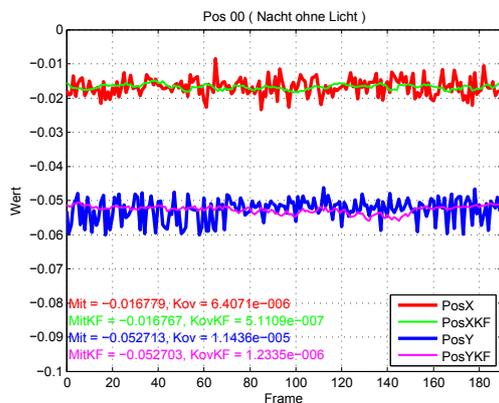
Der klassische Kalman-Filter, eingesetzt zur Glättung von Merkmalsdaten, erzielt bereits eine signifikante Verbesserung gegenüber den originalen Schätzwerten (siehe Abbildungen 5.9) und kann weiter gesteigert werden. Dazu wird ein adaptierter Kalman-Filter eingesetzt



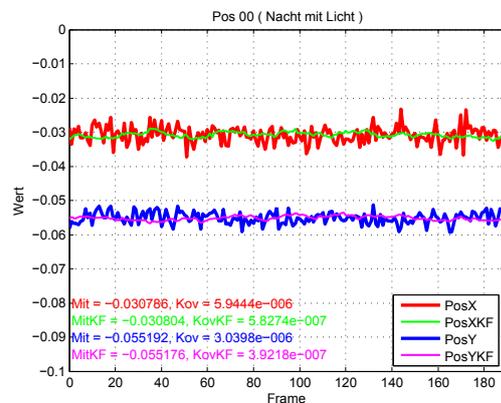
(a) Ohne künstliches Licht.



(b) Mit künstlichem Licht.



(c) Ohne künstliches Licht.



(d) Mit künstlichem Licht.

Abbildung 5.9: Schätzung der Position im Zentrum der Bildfläche bei Tageslicht und bei Dunkelheit.

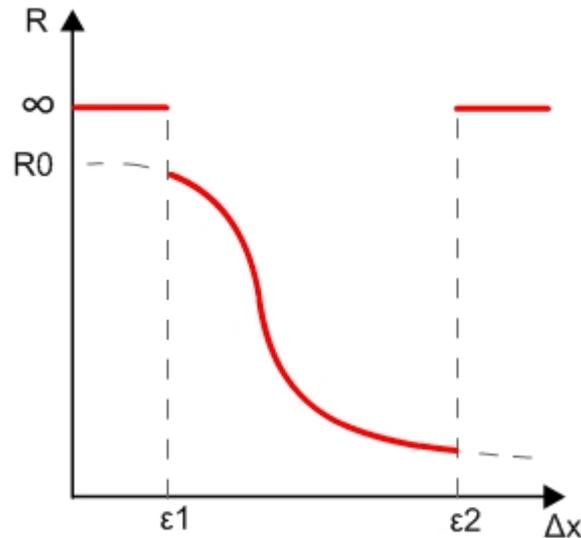


Abbildung 5.10: Regulierende Funktion für den Ausgleich kleiner und größerer Messstörungen.

[Graf und Wörn 2009a]. Zusätzliche experimentelle Untersuchungen wurden in [Zhou 2009] durchgeführt.

Die Idee für den adaptierten Kalman–Filter (AKF) liegt darin, die Kovarianz R zu benutzen, um den Kalman–Verstärkungsfaktor zu regulieren. Dazu folgende Gleichung:

$$R = \begin{cases} \infty & , \Delta x < \epsilon_1 \text{ oder } \epsilon_2 < \Delta x \\ R_0 \exp\{-\frac{\Delta x^2}{2\sigma^2}\} & , \text{sonst.} \end{cases} \quad (5.16)$$

Eine Skizze, mit der sich die Funktionsweise diskutieren lässt, ist in Abbildung 5.10 dargestellt.

Es wird angenommen, dass die auszuwertenden Bildfolgen aus einer kohärenten Bildfolge stammen. Deshalb können kleine Zustandsänderungen auf das zugrundeliegende Rauschen in den Kamerabildern zurückgeführt werden. Zustandsdifferenzen zwischen den beiden Schranken ϵ_1 und ϵ_2 werden als normale Veränderungen angesehen und deshalb entsprechend gewichtet. Für Werte nahe ϵ_1 wird der aktuelle Zustand stark von der alten Zustandsschätzung beeinflusst. Für Werte nahe bei ϵ_2 , wird R klein gewählt. In diesem Fall ist die aktuelle Zustandsschätzung stark vom aktuellen Messwert abhängig. Für große Differenzwerte Δx , d.h. Werte größer ϵ_2 wird angenommen, dass es sich um Störungen in Form von Ausreißern handelt. Deshalb wird die Kovarianz R wieder mit großen Werten ausgestattet, damit die vergangenen Schätzungen über die aktuellen Messungen dominieren.

Die Werte der Grenzen wurden in [Zhou 2009] experimentell ermittelt und sind in Tabelle 5.2 zusammengefasst.

In [Gelb 1974] werden die Grundlagen des Kalman–Filters (KF) hervorragend dargestellt. Mit Hilfe des KF kann der Zustand eines Systems geschätzt werden, wobei der mittlere qua-

Merkmal	Untere Schranke	Obere Schranke
Position x	0.05	-
Position y	0.05	-
Körpergröße	0.8	-
Orientierung Oberkörper	2.5	80.0
Orientierung Kopf	8.5	60.0
Oberarm (li. und re., $\{\alpha, \beta, \gamma\}$)	2.5	25.0
Unterarm (li. und re.)	5.0	40.0

Tabelle 5.2: Empirisch ermittelte Werte für ϵ_1 und ϵ_2 .

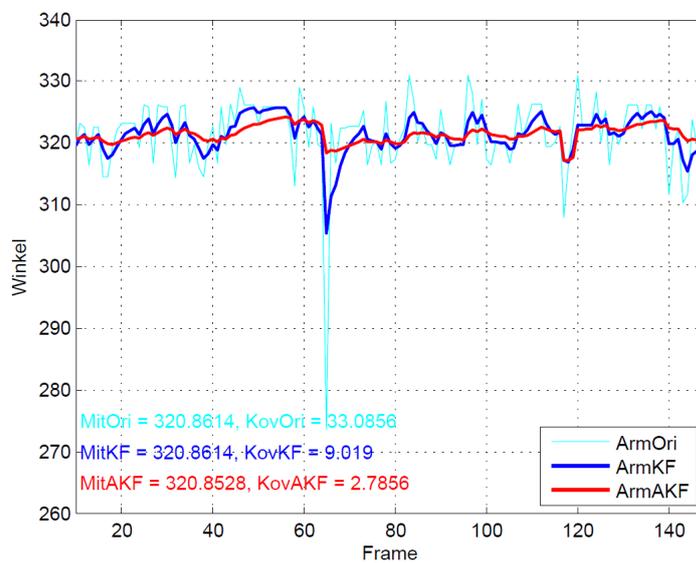
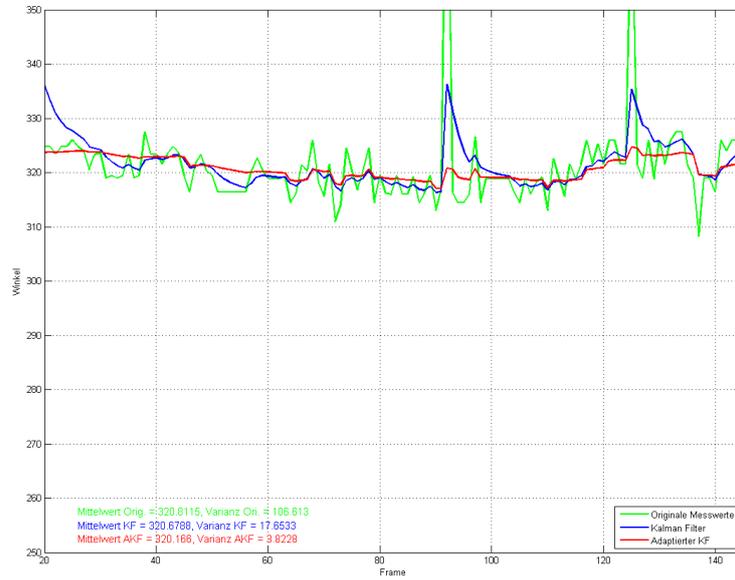
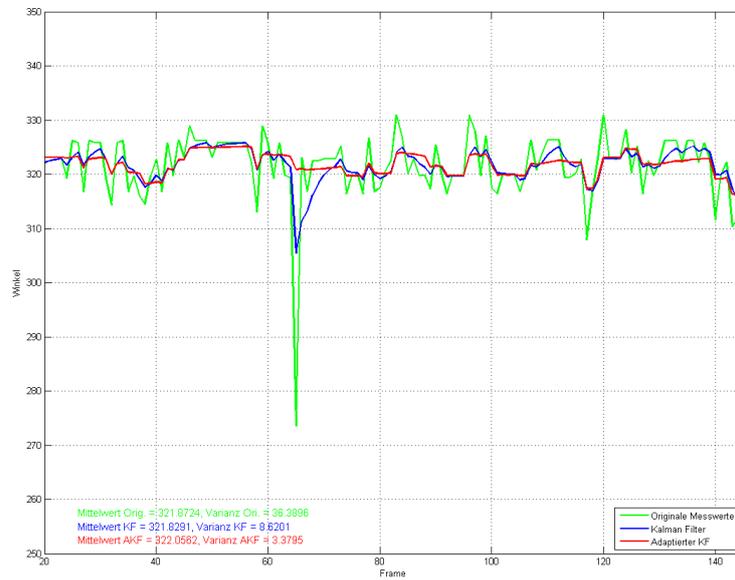


Abbildung 5.11: Experiment mit fixiertem Oberarmwinkel (ca. 320°).

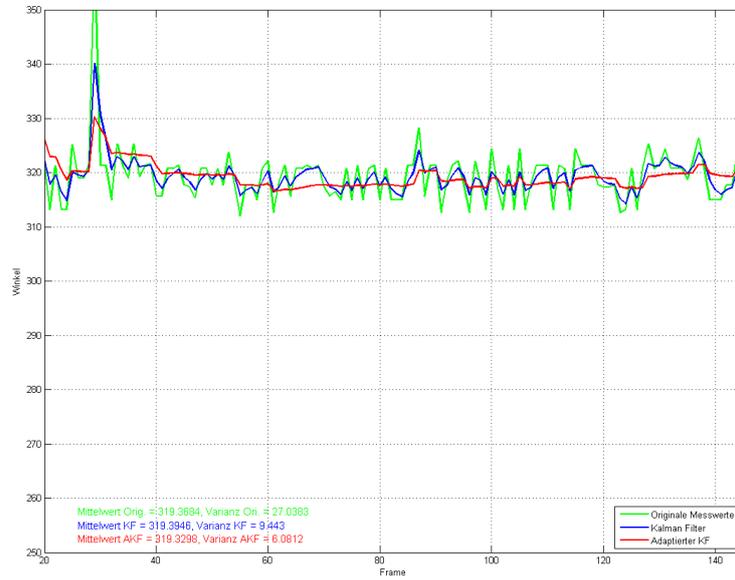


(a) Ohne künstliches Licht.

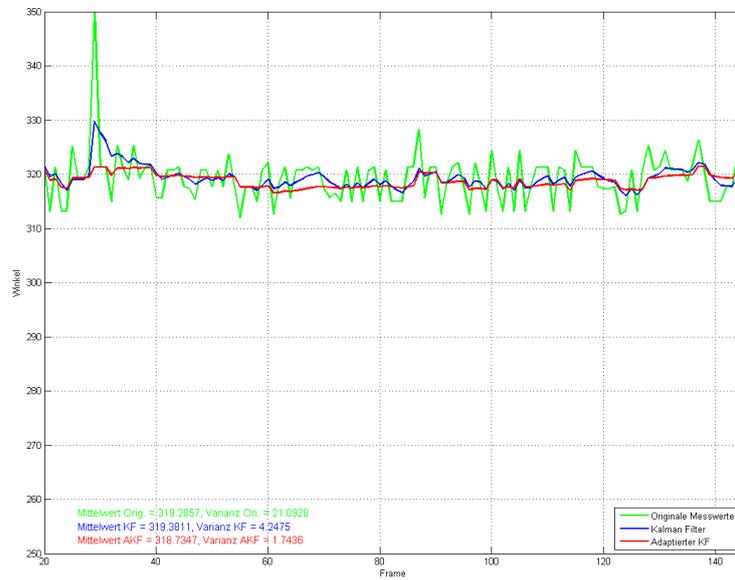


(b) Mit künstlichem Licht.

Abbildung 5.12: Schätzung der Position im Zentrum der Bildfläche bei Tageslicht.



(a) Ohne künstliches Licht.



(b) Mit künstlichem Licht.

Abbildung 5.13: Schätzung der Position im Zentrum der Bildfläche bei Dunkelheit.

drastische Fehler minimiert wird. Außerdem sind Aussagen über Zustände der Vergangenheit, Gegenwart und Zukunft möglich. Die rekursive Struktur ermöglicht eine effiziente Berechnung. Der KF wird in vielen Bereichen eingesetzt, bspw. bei der Verfolgung von Personen in Bildfolgen, oder generell zur Zustandsschätzung bei verrauschten Messungen.

Hier wird der KF zur Parameterschätzung eines durch MAROCO bereitgestellten Menschmodells verwendet.

System Modellierung

Der KF wird eingesetzt, um den Zustand $\mathbf{x}_k \in \mathbb{R}^n$ eines zeit-diskreten Systems zu schätzen. Das System-Modell (Abbildung 5.14 (a)) wird durch lineare Differenzen-Gleichungen beschrieben:

$$\mathbf{x}_k = \mathbf{A}\mathbf{x}_{k-1} + \mathbf{B}\mathbf{u}_{k-1} + \mathbf{w}_{k-1} \quad . \quad (5.17)$$

Der Vektor $\mathbf{u}_k \in \mathbb{R}^n$ beschreibt den Systemeingang. Das Systemrauschen und mögliche Modellierungsfehler werden durch die normalverteilte Zufallsvariable $\mathbf{w}_k \sim N(0, \mathbf{Q}_k)$ berücksichtigt.

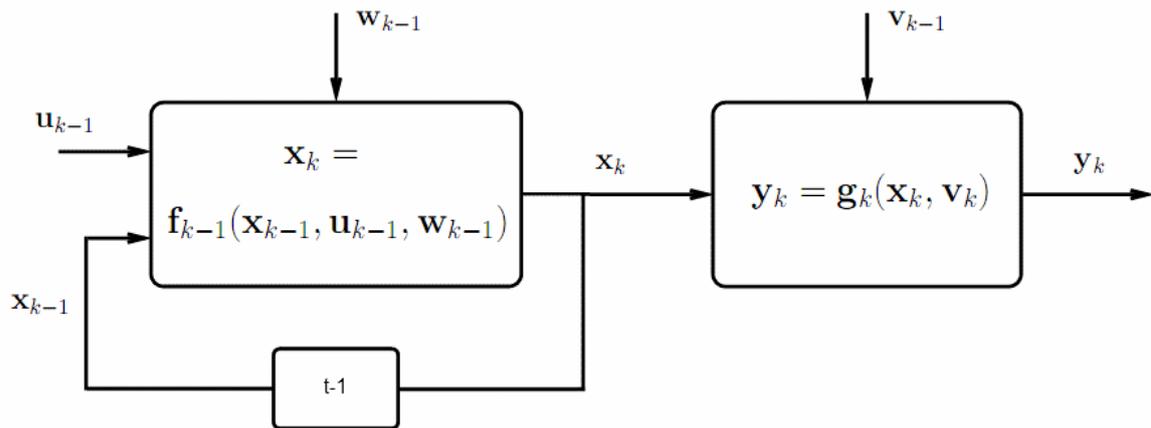


Abbildung 5.14: Zeit-diskretes System. Linker Block: Prozess-Modell. Rechter Block: Mess-Modell

Für das Mess-Modell (Abbildung 5.14 (b)) wird der Zustand \mathbf{x}_k und die Beobachtung bzw. Messung $\mathbf{z}_k \in \mathbb{R}^m$ in einen linearen Zusammenhang gebracht:

$$\mathbf{z}_k = H\mathbf{x}_k + \mathbf{v}_k \quad . \quad (5.18)$$

Durch die Addition der normalverteilten Zufallsvariable $\mathbf{v}_k \sim N(0, R_k)$, wird das Mess-Rauschen berücksichtigt. Die beiden Zufallsvariablen \mathbf{w}_k und \mathbf{v}_k werden als unabhängig angenommen.

System des Menschmodells

In Abbildung 5.14 ist ein lineares zeit-diskretes-System abgebildet, welches in MAROCO integriert wurde um kinematische Parameter zu schätzen. Der Zustand zum Zeitpunkt k , wird durch den Zustandsvektor

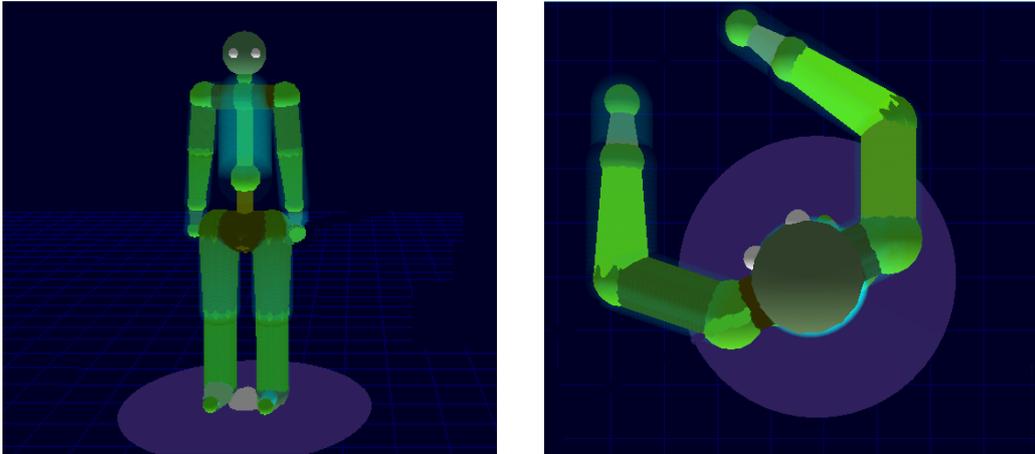


Abbildung 5.15: Visualisierung des Mensch-Modells.

$$\mathbf{x}_k = \begin{pmatrix} x \\ y \\ z \\ \theta_m \\ \theta_k \\ \alpha_{l1} \\ \beta_l \\ \alpha_{r1} \\ \beta_r \\ \alpha_{l2} \\ \alpha_{l3} \\ \alpha_{r2} \\ \alpha_{r3} \end{pmatrix} \quad \begin{array}{l} \text{Körper Postion } x \\ \text{Körper Postion } y \\ \text{Körper Postion } z \\ \text{Körper Orientierung} \\ \text{Kopf Orientierung} \\ \text{Winkel Schulter links 1} \\ \text{Winkel Ellbogen links} \\ \text{Winkel Schulter rechts 1} \\ \text{Winkel Ellbogen rechts} \\ \text{Winkel Schulter links 2} \\ \text{Winkel Schulter links 3} \\ \text{Winkel Schulter rechts 2} \\ \text{Winkel Schulter rechts 3} \end{array} \quad (5.19)$$

bestimmt.

Der Systemeingang \mathbf{u} wird in dem System-Modell nicht beachtet. Die Übergangsmatrix \mathbf{A} und die Kovarianzmatrix \mathbf{Q} des Systemrauschens haben folgende Form:

$$\mathbf{A} = \begin{pmatrix} \mathbf{M}_{A_1} & 0 & \dots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \mathbf{M}_{A_{13}} \end{pmatrix}, \quad \mathbf{Q} = \begin{pmatrix} \mathbf{M}_{Q_1} & 0 & \dots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \mathbf{M}_{Q_{13}} \end{pmatrix}. \quad (5.20)$$

Es können drei Varianten sinnvoll eingesetzt werden:

1. $M_{A_i} = 1$ und $M_{Q_i} = \tilde{q}_i$.

2. Kontinuierliches weißes Rauschen Prozess-Beschleunigungsmodell (engl. CWNA):

$$\mathbf{M}_{A_i} = \begin{pmatrix} 1 & \Delta t \\ 0 & 1 \end{pmatrix}, \quad M_{Q_i} = \begin{pmatrix} \frac{1}{3}\Delta t^3 & \frac{1}{2}\Delta t^2 \\ \frac{1}{2}\Delta t^2 & \Delta t \end{pmatrix} \tilde{q}_i. \quad (5.21)$$

Jede Komponente wird erweitert auf (x_i, \dot{x}_i)

3. Kontinuierliches Wiener-Prozess Beschleunigungsmodell (engl. CWPA):

$$M_{A_i} = \begin{pmatrix} 1 & \Delta t & \frac{1}{2}\Delta t^2 \\ 0 & 1 & \Delta t \\ 0 & 0 & 1 \end{pmatrix}, \quad M_{Q_i} = \begin{pmatrix} \frac{1}{20}\Delta t^5 & \frac{1}{8}\Delta t^4 & \frac{1}{6}\Delta t^3 \\ \frac{1}{8}\Delta t^4 & \frac{1}{3}\Delta t^3 & \frac{1}{2}\Delta t^2 \\ \frac{1}{6}\Delta t^3 & \frac{1}{2}\Delta t^2 & \Delta t \end{pmatrix} \tilde{q}_i. \quad (5.22)$$

Jede Komponente wird erweitert auf $(x_i, \dot{x}_i, \ddot{x}_i)$

Wobei hier die zweite und dritte Variante verwendet wurden. Der Faktor \tilde{q}_i bezeichnet jeweils die Varianz des Systemrauschens der i -ten Komponente.

Der Messvektor \mathbf{z}_k entspricht dem Zustandsvektor \mathbf{x}_k . Die Matrix H der Messabbildung (5.18) ist unabhängig von k . Sie wird an die jeweilige Variante angepasst.

Die einzelnen Komponenten des Messvektors \mathbf{z}_k werden aus den Tiefenmessungen der 3D-Kamera bestimmt. Auf die dazu angewendeten Verfahren wird in dieser Arbeit nicht eingegangen.

Die Unsicherheiten der Messung bzw. der Verfahren, wird wie bereits erwähnt, durch das additive weiße Rauschen \mathbf{v}_k modelliert. Die Kovarianzmatrix R_k hat folgende Form:

$$R_k = \begin{pmatrix} r_1 & 0 & \dots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & r_{13} \end{pmatrix},$$

$$r_i = \begin{cases} \exp\left(-\frac{1}{2}\frac{(z_{i_k} - \hat{x}_{i_{k-1}})^2}{\tilde{r}_i^2}\right) & , l_i \leq |(z_{i_k} - \hat{x}_{i_{k-1}})| \leq u_i, \\ L & , \text{sonst.} \end{cases} \quad (5.23)$$

Der Faktor \tilde{r}_i bezeichnet die Varianz des Messrauschens, der i -ten Komponente. Der Faktor L steht für einen großen Wert. Ist die aktuelle Messung sehr nahe an der letzten Schätzung, wird ein Stillstand angenommen. Durch das Einsetzen von L als Varianz soll verhindert werden, dass durch kleine Ungenauigkeiten der Messungen eine Bewegung geschätzt wird. Ist die Differenz der aktuellen Messung und der a posteriori Schätzung des letzten Zeitschritts sehr groß, wird von einer fehlerhaften Messung ausgegangen. Der Faktor L bewirkt, dass die Messung bei der aktuellen a posteriori Schätzung nicht beachtet wird.

Einbettung der OF-Schätzung in den Kalman-Filter

Der KF eignet sich auch dazu, Informationen aus verschiedene Quellen zu fusionieren. Diese Informationen sind bspw. Messungen von unterschiedlichen Sensoren oder stammen wie in diesem Fall aus der unterschiedlichen Verarbeitung der gleichen Messung.

In dem bestehenden System wird aus den Tiefendaten die Messung \mathbf{z}_k des Zustandes \mathbf{x}_k ermittelt. Die OF-Berechnung aus diesen Daten liefert weitere, von der bisherigen Verarbeitung unabhängige Informationen. Dabei handelt sich um die Schätzung der Verschiebungen einzelner Punkte. Daraus können bspw. die Geschwindigkeit des Oberkörpers oder die Winkelgeschwindigkeiten der Armsegmente ermittelt werden.

Um die Informationen über die Geschwindigkeit einzelner Komponenten in die Zustandschätzung aufzunehmen, wird der Zustandsvektor \mathbf{x}_k für die betreffenden Komponenten erweitert:

$$x_i \longrightarrow \mathbf{k}_i = \begin{pmatrix} x_i \\ \dot{x}_i \end{pmatrix} . \quad (5.24)$$

Die zugehörige Übergangsmatrix \mathbf{A} und die Kovarianz \mathbf{Q} (5.20) des System-Modells haben die Form:

$$\mathbf{M}_{A_i} = \begin{pmatrix} 1 & \Delta t \\ 0 & 1 \end{pmatrix} , \quad \mathbf{M}_{Q_i} = \begin{pmatrix} \frac{1}{3}\Delta t^3 & \frac{1}{2}\Delta t^2 \\ \frac{1}{2}\Delta t^2 & \Delta t \end{pmatrix} \tilde{q}_i . \quad (5.25)$$

Der Messvektor \mathbf{z}_k wird dem Zustand \mathbf{x}_k angepasst:

$$z_i \longrightarrow \mathbf{l}_i = \begin{pmatrix} z_i \\ \dot{z}_i \end{pmatrix} . \quad (5.26)$$

Die Messabbildung \mathbf{H} entspricht der Einheitsmatrix \mathbf{I} . Die Kovarianz \mathbf{R} des Messrauschens hat die folgende Form:

$$\mathbf{R}_k = \begin{pmatrix} r_1 & 0 & \dots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & r_n \end{pmatrix} . \quad (5.27)$$

Für Komponenten, welche nicht eingebettet werden, bleiben die Diagonalelemente wie in Gleichung (5.23) beschrieben. Komponenten, die von der Einbettung betroffen sind, haben folgende Einträge:

$$z_i : r_i = \begin{cases} \exp\left(-\frac{1}{2} \frac{(z_i - \hat{x}_{i_{k-1}})^2}{\tilde{r}_i^2}\right) & , s_i \leq \dot{x}_{i_{k-1}} \wedge |(z_{i_k} - \hat{x}_{i_{k-1}})| \leq u_i \\ L & , \text{sonst} \end{cases} \quad (5.28)$$

$$\dot{z}_i : r_i = 0 .$$

Der Faktor L steht für einen großen Wert. Wird nahezu keine Veränderung der Komponente geschätzt, wird die Varianz sehr groß gesetzt, so dass durch ungenaue Messungen keine Bewegung berechnet wird. Dazu wird $\dot{x}_{i_{k-1}}$ betrachtet, welches als konstant über einen Zeitschritt hinweg angenommen wird.

Ist die Differenz der Messung und Schätzung sehr groß, wird von einer fehlerhaften Messung ausgegangen, und die Varianz nimmt ebenfalls den Wert L an. Das hat zur Folge, dass die Messung bei der a posteriori Schätzung nicht beachtet wird.

Für \dot{z}_i wird die Varianz als 0 angenommen. Das bewirkt, dass allein die Messung des OF den a posteriori Zustand festlegt. Eine Begründung für diese Definition wird im Folgenden angegeben.

Durch die vorgenommenen Änderungen des Systems, wird in dem Zeit-Update (Abbildung 5.14) die Informationen der OF-Schätzung zur Prädiktion des *a priori* Zustands \mathbf{x}_k^- verwendet. Bei der Aktualisierung der Messung (Abbildung 5.14) wird die Messung, welche aus dem bisherigen Verfahren stammt, als Rückkopplung zur Korrektur des *a priori* Zustands verwendet. Der resultierende *a posteriori* Zustand wird mithilfe beider Informationsquellen geschätzt, wobei der Kalman-Verstärkungsfaktor diese gewichtet.

Dies betrifft jedoch nur Komponenten, für welche die Einbettung angewandt wird, da die Komponenten untereinander unabhängig sind. Das bedeutet ebenfalls, dass die Berechnungen der Filterung für jede Komponente einzeln durchgeführt werden können.

Einbettung von 3D OF Schätzungen

Für diesen Abschnitt wird angenommen, dass der OF in 3D geschätzt wurde. Außerdem sind alle relevanten Armpunkte als 3D Daten vorhanden, wie in Kapitel 3D Optische Fluss beschrieben. Unter dieser Voraussetzung kann die Fusion auf fast alle Komponenten des Zustandsvektors \mathbf{x}_k angewandt werden. Die Ausnahmen sind die Kopf- und die Körperorientierung.

3D-OF Schätzung eines Punktes

Zur Prädiktion muss der 3D-OF an einer bestimmten Positionen ermittelt werden. Dazu wird die Faltung der 2D-OF-Funktionen u_1 und u_2 mit einer lokalen Gewichtungsfunktion G , für diese Position ausgewertet:

$$\begin{aligned}\bar{u}_1 &= (G * u_1)(\hat{x}_{k-1}, \hat{y}_{k-1}) \\ \bar{u}_2 &= (G * u_2)(\hat{x}_{k-1}, \hat{y}_{k-1})\end{aligned}\quad (5.29)$$

Als Faltungsfunktion kann bspw. eine isotrope Gaußfunktion gewählt werden. Eine einfachere und weniger rechenintensive Funktion wäre bspw. die Rechteckfunktion.

Das Ergebnis wird verwendet um den 3D OF $\bar{\mathbf{u}} = (\bar{u}_x, \bar{u}_y, \bar{u}_z)^T$ in diesem Punkt zu bestimmen.

Körperposition

Da der OF in 3D gegeben ist, können die Körpergeschwindigkeiten direkt berechnet werden. Dazu wird, wie oben beschrieben, der 3D-OF an der *a posteriori* Positionsschätzung im letzten Zeitschritt $t - 1$ ermittelt. Die komponentenweise Teilung durch das Zeitintervall Δt liefert das Ergebnis:

$$\begin{aligned}\dot{x} &= \frac{\bar{u}_x}{\Delta t} \\ \dot{y} &= \frac{\bar{u}_y}{\Delta t} \\ \dot{z} &= \frac{\bar{u}_z}{\Delta t}\end{aligned}\quad (5.30)$$

Die Körperposition z bezeichnet die Höhe des Körpers.

Im Folgenden werden Berechnungen bzgl. Armkomponenten beispielhaft für einen Arm durchgeführt. Daher werden die Indizes l und r weggelassen.

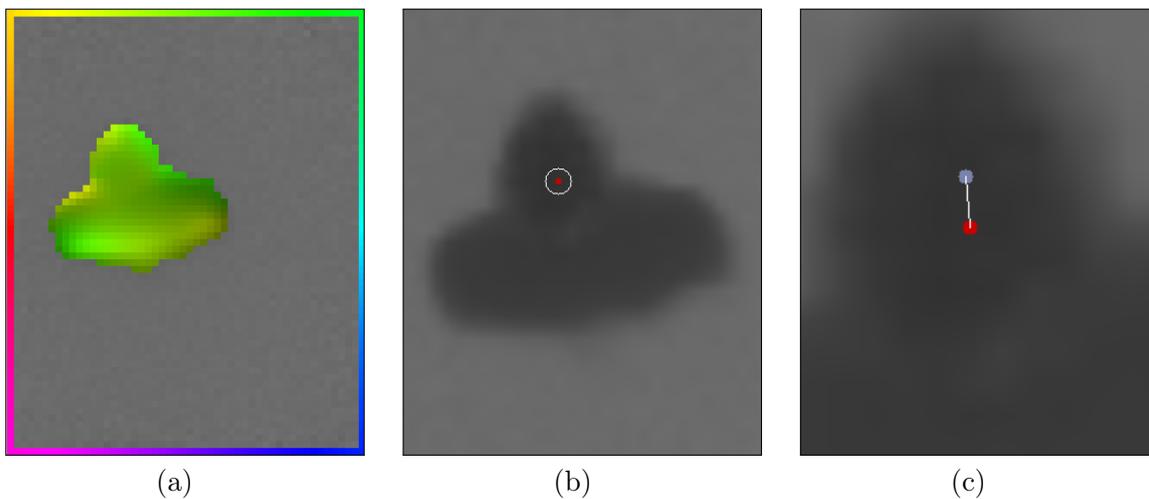


Abbildung 5.16: (a) OF-Schätzung zum Zeitpunkt $t - 1$. (b) Umgebung zur OF-Schätzung des Körpermittelpunktes. (c) Prädiktion (blau) des Körpermittelpunktes zum Zeitpunkt t .

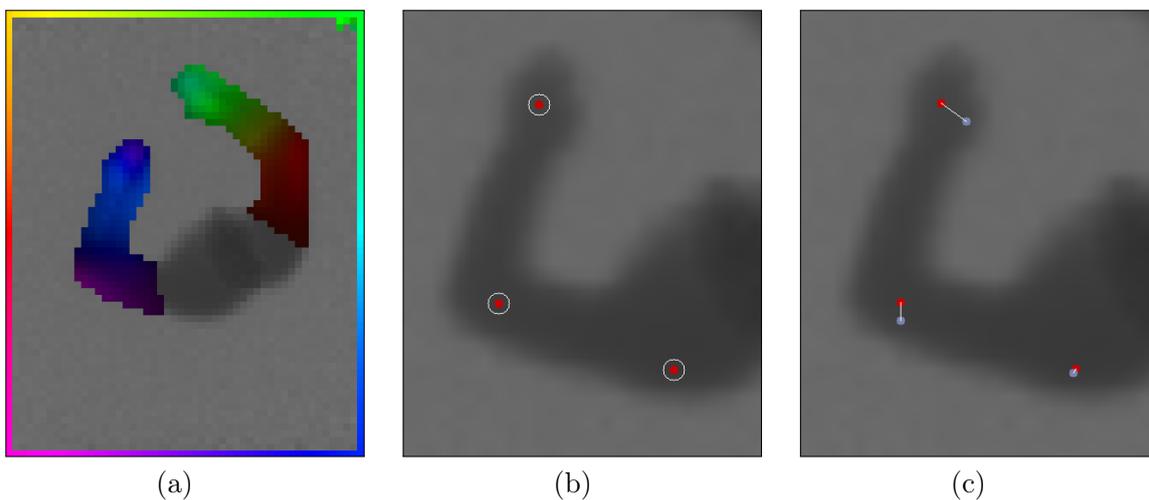


Abbildung 5.17: (a) OF-Schätzung zum Zeitpunkt $t - 1$. (b) Umgebungen zur OF-Schätzung der Hand, Ellbogen und Schulter. (c) Prädiktionen (blau) der Hand, Ellbogen und Schulter zum Zeitpunkt t .

Schulterwinkel

Der Oberarm besitzt bzgl. der Schulter drei Freiheitsgrade. Er kann sich in horizontaler und vertikaler Richtung bewegen und um seine eigene Längsachse drehen. Diese Bewegungen werden entsprechend durch die Schulterwinkel α_1 , α_2 und α_3 beschrieben. Um die Geschwindigkeiten der Winkel zu bestimmen, werden die Differenzen der Winkel vor und nach der Prädiktion durch den OF berechnet. Als Wert vor der Prädiktion, wird der *a posteriori* Zustand \hat{x}_{k-1} des vorherigen Zeitschritts $k - 1$ verwendet. Die Schulterposition wird mit $\mathbf{s} = (s_x, s_y, s_z)^T$, die Ellbogenposition mit $\mathbf{e} = (e_x, e_y, e_z)^T$ und die Handposition mit $\mathbf{h} = (h_x, h_y, h_z)^T$ beschrieben. Positionen werden mithilfe des 3D-OF auf folgende Weise prädiziert:

$$\mathbf{p}_k^- = \hat{\mathbf{p}}_{k-1} + \bar{\mathbf{u}} \quad . \quad (5.31)$$

Zur Berechnung der Geschwindigkeiten werden zuerst die Prädiktionen ermittelt. Der prädizierte Winkel $\alpha_{1_k}^-$ wird folgendermaßen berechnet:

$$\begin{aligned} d_x &= e_{x_k}^- - s_{x_k}^- \quad , \\ d_y &= e_{y_k}^- - s_{y_k}^- \quad , \\ \alpha_{1_k}^- &= \tan\left(\frac{d_y}{d_x}\right) \quad . \end{aligned} \quad (5.32)$$

Die Berechnung des prädizierten Winkels $\alpha_{2_k}^-$ ergibt sich zu:

$$\begin{aligned} d_z &= e_{z_k}^- - s_{z_k}^- \quad , \\ d_l &= \sqrt{d_x^2 + d_y^2} \quad , \\ \alpha_{2_k}^- &= \tan\left(\frac{d_z}{d_l}\right) \quad . \end{aligned} \quad (5.33)$$

Des Berechnung des prädizierten Winkels $\alpha_{3_k}^-$ ist eine Approximation. Der Ellbogenwinkel wird dabei als $\approx \frac{\pi}{2}$ angenommen:

$$\begin{aligned} d_x &= h_{x_k}^- - e_{x_k}^- \quad , \\ d_y &= h_{y_k}^- - e_{y_k}^- \quad , \\ d_z &= h_{z_k}^- - e_{z_k}^- \quad , \\ d_l &= \sqrt{d_x^2 + d_y^2} \quad , \\ \alpha_{3_k}^- &= \tan\left(\frac{d_z}{d_l}\right) \quad . \end{aligned} \quad (5.34)$$

Die resultierenden Winkelgeschwindigkeiten sind:

$$\begin{aligned}
\dot{\alpha}_1 &= \frac{\alpha_{1k}^- - \hat{\alpha}_{1k-1}}{\Delta t} , \\
\dot{\alpha}_2 &= \frac{\alpha_{2k}^- - \hat{\alpha}_{2k-1}}{\Delta t} , \\
\dot{\alpha}_3 &= \frac{\alpha_{3k}^- - \hat{\alpha}_{3k-1}}{\Delta t} ,
\end{aligned} \tag{5.35}$$

Ellbogenwinkel

Zur Berechnung der Winkelgeschwindigkeit des Ellbogens wird ebenfalls zuerst die Prädiktion des Winkels berechnet:

$$\begin{aligned}
d_x &= h_{x_k}^- - e_{x_k}^- , \\
d_y &= h_{y_k}^- - e_{y_k}^- , \\
\beta_{1k}^- &= \pi - \alpha_{1k}^- - \tan\left(\frac{d_y}{d_x}\right) .
\end{aligned} \tag{5.36}$$

Die resultierende Winkelgeschwindigkeit ist:

$$\dot{\beta} = \frac{\beta_k^- - \hat{\beta}_{k-1}}{\Delta t} . \tag{5.37}$$

Einbettung von 2D OF Schätzungen

Ist der OF in 2D gegeben, wird die Einbettung für die gleichen Komponenten wie im vorherigen Abschnitt angewandt. Es gelten ebenfalls die selben Berechnungsvorschriften. Die Berechnung der Winkelgeschwindigkeiten unterscheidet sich in folgenden Punkten:

1. Die Punkte, für welche die Berechnungen ausgeführt werden, stammen direkt aus den Tiefenmessungen. Es findet keine Transformation in den Darstellungsraum (vgl. Abschnitt 4.4.4) statt.
2. Der OF in einem bestimmten Punkt, wird wie im vorherigen Abschnitt berechnet, aber es findet keine Umrechnung in 3D statt.
3. Die Prädiktion von Positionen durch den OF findet in 2D statt. Die Positionen eines Punktes bzgl. der z -Achse werden aus den Tiefenmessungen ermittelt.

5.2.2 Projektion von Kugelmodellen auf die Bildebene

Um effizient die Abstände zwischen den Kinematiken zu berechnen und die Verdeckungsanalyse durchzuführen, werden die Achsen der kinematischen Modelle durch aneinandergereihte Kugeln repräsentiert (siehe Abbildung 5.18). Um eine Segmentierung des Roboters im Kamerabild möglich zu machen, wird zu jeder Roboterkugel in 3D-Weltkoordinaten das 2D-Bild in der Kameraebene berechnet.

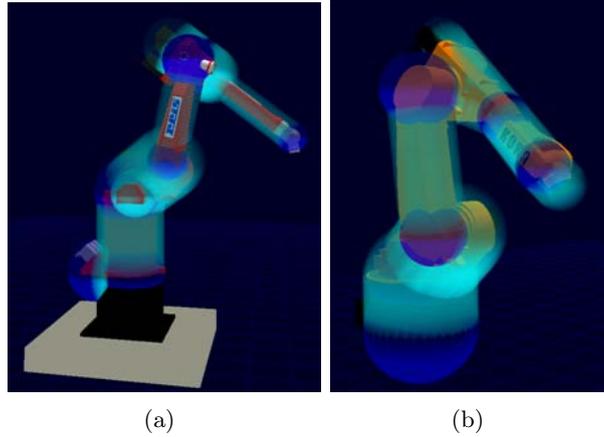


Abbildung 5.18: Repräsentation der kinematischen Strukturen mit Kugelmodellen.

Kameramodell

Zur Abbildung des 3D-Szene auf die 2D-Kamerabildebene wird ein Lochkameramodell verwendet (siehe bspw. [Schreer 2005]). Die 3D-Szenekoordinaten werden dabei durch eine perspektivische Transformation auf die 2D-Bildebene projiziert. Anschließend werden die 2D Bildkoordinaten in 2D-Kamerakoordinaten transformiert.

Verwendet man euklidischen Koordinaten anstelle von homogene Koordinaten, wird ein Punkt $(x, y, z, 1)$ im 3D-Weltkoordinatensystem durch die Matrix

$$P = \begin{pmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \quad (5.38)$$

in einen Punkt (x', y', s) im 2D Weltkoordinatensystem perspektivisch transformiert. Die euklidischen Koordinaten sind dann $(x/s, y/s)$ für $s \neq 0$. Dabei bezeichnet die Brennweite f den Abstand der Bildebene zum Brennpunkt.

Bei dieser Transformation liegt das Projektionszentrum C im Ursprung des 3D Weltkoordinatensystems. Befindet sich das Projektionszentrum an einer anderen Position, kann das Weltkoordinatensystem mit dem Ursprung O durch eine Rotation R und anschließende Translation t in das Kamerakoordinatensystem mit dem Ursprung C überführt werden. In homogenen Koordinaten geschieht das durch die Transformationsmatrix

$$D = \begin{pmatrix} R & t \\ 0^T & 1 \end{pmatrix} \quad (5.39)$$

Damit ergibt sich für die Transformation von 3D Weltkoordinaten in 2D Kamerakoordinaten:

$$\mathbf{x}' = P' \mathbf{x}, \text{ mit } P' = PD \quad (5.40)$$

Die Transformation der 2D-Kamerakoordinaten in Bildkoordinaten geschieht durch die horizontale und vertikale Skalierung k_u und k_v . Mit der Translation $(u_0, v_0)^T$ erhält man die Transformationsmatrix:

$$H = \begin{pmatrix} k_u & 0 & u_0 \\ 0 & k_v & v_0 \\ 0 & 0 & 1 \end{pmatrix} \quad (5.41)$$

Daraus erzeugt man die Kameramatrix K mit

$$K = HPD = HP' \quad . \quad (5.42)$$

Ein Punkt \mathbf{x} in 3D Weltkoordinaten transformiert sich in 2D Bildkoordinaten \mathbf{x}' durch

$$\mathbf{x}' = K\mathbf{x} \quad (5.43)$$

Die nichtlinearen Verzerrungen δ_u, δ_v und verzerrten Koordinaten u_d und v_d erfüllen die folgende Gleichung:

$$u = u_d + \delta_u, v = v_d + \delta_v \quad . \quad (5.44)$$

Werden die radialen Verzerrungen κ_i und tangentialen Verzerrungen η_i mit berücksichtigt gilt:

$$\delta_u = u_d(\kappa_1 r_d^2 + \kappa_2 r_d^4 + \dots) + [\eta_1(r_d^2 + 2u_d^2) + 2\eta_2 u_d v_d](1 + \eta_3 r_d^2 + \dots) \quad (5.45)$$

$$\delta_v = v_d(\kappa_1 r_d^2 + \kappa_2 r_d^4 + \dots) + [2\eta_1 x_d v_d + \eta_2(r_d^2 + 2v_d^2)](1 + \eta_3 r_d^2 + \dots) \quad (5.46)$$

mit

$$r_d = \sqrt{u_d^2 + v_d^2} \quad . \quad (5.47)$$

Berücksichtigt man nur die radialen Verzerrungen, gilt:

$$u = u_d(1 + \kappa_1 r_d^2) \quad , \quad (5.48)$$

$$v = v_d(1 + \kappa_1 r_d^2) \quad . \quad (5.49)$$

Zur perspektivischen Transformation einer Kugel

Die Gleichung einer Kugel im euklidischen Koordinatensystem lautet

$$x^2 + y^2 + z^2 = r^2 \quad . \quad (5.50)$$

Im projektiven Raum P^3 wird dies zu einer Quadrik mit der Matrix Q :

$$\mathbf{x}^T Q \mathbf{x} = 0 \quad , \quad Q = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -r^2 \end{pmatrix} \quad , \quad \mathbf{x} \in P^3 \quad (5.51)$$

Nach [Calion 2009] ist die Kontur einer Quadrik unter der Kameramatrix P eine Konik und es gilt für die Matrix der dualen Quadrik Q^* und der duale Konik C^* :

$$C^* = P Q^* P^T \quad . \quad (5.52)$$

Die duale Konik Q^* erhält man durch Bildung der adjunkten Matrix von Q . Die Einträge von Q^* sind

$$Q_{ij}^* = (-1)^{i+j} \det(Q'_{ij}) \quad , \quad (5.53)$$

wobei Q'_{ij} die Matrix ist, welche man durch Streichung der i -ten Zeile und j -ten Spalte aus Q erhält. Falls die Konik C invertierbar ist, gilt, da C symmetrisch ist:

$$C^{-1} = \frac{1}{\det(C)} \cdot (C^*)^T = \frac{1}{\det(C)} \cdot C^* \quad . \quad (5.54)$$

Die inverse Matrix C unterscheidet sich also von der adjunkten Matrix C^* nur durch einen Faktor. Deshalb entspricht die duale Konik C^* der Konik C^{-1} .

Im Falle einer Ellipse erhält man durch Rotation und Translation der Ellipsenmatrix C eine Ellipsenmatrix C' deren Mittelpunkt im Ursprung liegt und deren Hauptachsen auf den Koordinatenachsen liegen. Hat man

$$C = \begin{pmatrix} C_R & C_T \\ C_T^T & f \end{pmatrix} \quad (5.55)$$

und hat C_R die Eigenwerte λ_1 und λ_2 mit den korrespondierenden Eigenvektoren \mathbf{v}_1 und \mathbf{v}_2 , erzwingt man durch die Hauptachsentransformation eine Diagonalisierung

$$C_R = R \cdot \text{diag}(\lambda_1, \lambda_2) \cdot R^T \quad , \text{ mit} \quad (5.56)$$

$$R = \begin{pmatrix} \mathbf{v}_1 & \mathbf{v}_2 \end{pmatrix} \quad . \quad (5.57)$$

Zusammen mit der Translationsmatrix

$$T = -R \cdot \text{diag}(\lambda_1, \lambda_2)^{-1} \cdot R^T C_T \quad . \quad (5.58)$$

ergibt sich nach [Hartley und Zisserman 2004] eine homogene Transformationsmatrix

$$H = \begin{pmatrix} R & T \\ \mathbf{0}^T & 1 \end{pmatrix} \quad , \quad (5.59)$$

welche die Matrix zur Beschreibung der Ellipse C in die Normalform mit der Ellipsenmatrix C' transformiert:

$$C' = H^T C H \quad . \quad (5.60)$$

Die Ellipsengleichung liegt dann in parametrischer Normalform vor:

$$\frac{\lambda_1}{-C'_{33}} x^* + \frac{\lambda_2}{-C'_{33}} y^* = 1 \quad . \quad (5.61)$$

Durch die Normalform lässt sich das Bild der Kugel in der 2D-Bildebene abbilden und kann dort die gewünschte Segmentierung durchführen.

5.2.3 Experimente und Bewertung

In der Abbildung 5.19 sind verschiedene Bewegungsformen aufgezeichnet und ausgewertet worden:

1. Bewegung des Körpers im Gesichtsfeld der Kamera (Bilder 0-618).
2. Veränderung der Armstellung bei stillstehendem Körper (Bilder 619-1200).
3. Drehung des Oberkörpers (Bilder 1201-1700).
4. Stillstand (Bilder 1701-2196).
5. Bewegung der Arme bei bewegtem Körper (Bilder 2197-3000).

Die Tabelle 5.3 quantifiziert die Überdeckung durch die beiden statistischen Momente - Mittelwert und Varianz. In Abbildung 5.20 sind ausgewählte Überdeckungen der Bildfolge mit den rückprojizierten Kugelmodellen.

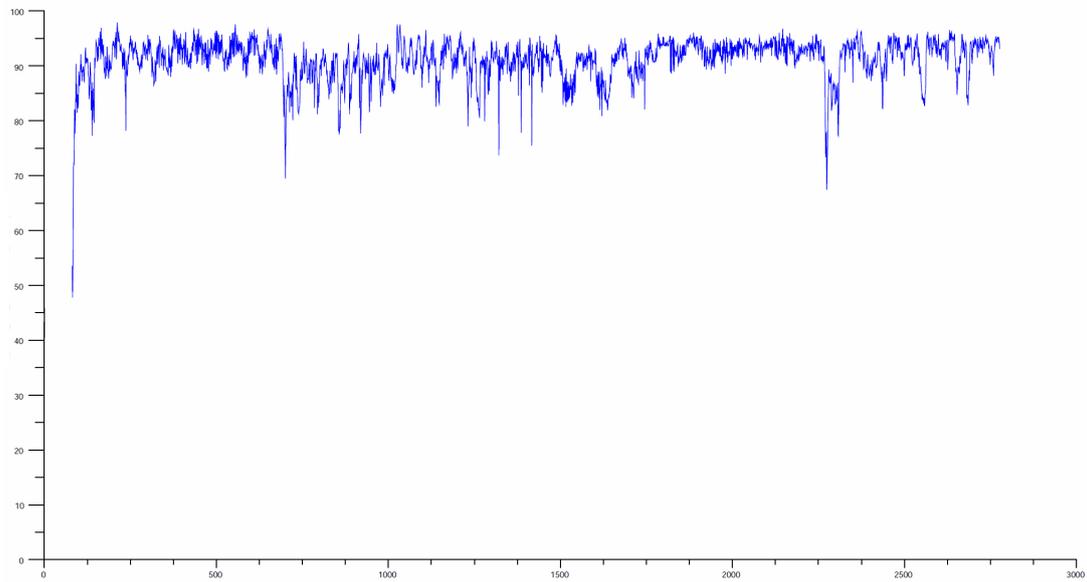


Abbildung 5.19: Experimentelle Auswertung des Verdeckungsgrades. Auf der Abszisse sind die einzelnen Bilder abgebildet, die Ordinate gibt den Verdeckungsgrad wieder.

Gruppierung	Mittelwert	Varianz
1	91.60	19.82
2	89.46	14.71
3	90.07	9.67
4	93.13	2.33
5	91.60	14.53

Tabelle 5.3: Quantifizierung der Überdeckungsgrade durch die beiden ersten statistischen Momente.

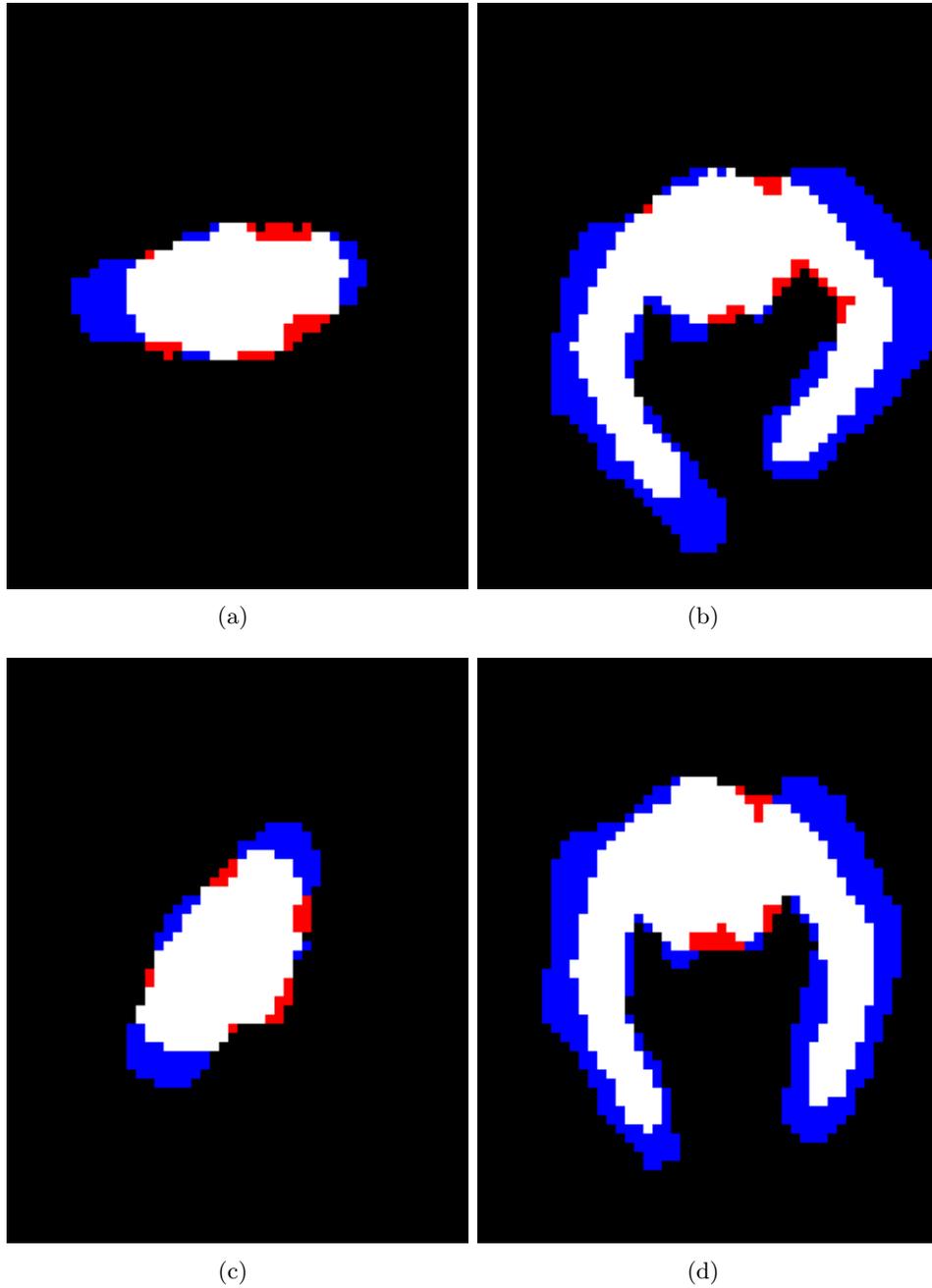


Abbildung 5.20: Ausgewählte Bilder zur Visualisierung des Verdeckungsgrades.

Kapitel 6

Risikoschätzung

Die Risikoschätzung soll im Rahmen dieser Arbeit durch ausgewählte Merkmale bestimmt werden. Der Vorteil liegt darin, dass bei einer auf Merkmalen basierenden Verarbeitung ein Fülle von geschätzten Werten Einfluss auf die Berechnungen haben. Dadurch lassen sich komplexe Sachverhalte differenzierter und flexibler behandeln.

Die Merkmale für die Risikoschätzung stammen zum großen Teil aus den Resultaten der Bildfolgenauswertung oder sind a priori bekannt, wie bspw. die Kinematikkonfiguration der Roboter.

Durch die Bereitstellung von Merkmalen lassen sich auch Methoden des maschinellen Lernens dazu verwenden, das Risiko zu bestimmen. Im Rahmen dieser Arbeit wird darauf jedoch nicht weiter eingegangen, der interessierte Leser sei auf [Graf u.a. 2010b] verwiesen. Dort findet ein erster Vergleich zwischen den in MAROCO verwendeten zweisträngigen unscharfen Logiken und den Unterstützungsvektormethoden¹ statt.

Eine besonders interessante Möglichkeit bei der Kenntnis von kinematischen Parametern besteht darin, dass man das Risiko, welches während der Mensch–Roboter–Kooperation entsteht, differenzierter schätzen kann, als bspw. Verfahren, die ohne dieses Zusatzwissen arbeiten.

In MAROCO stehen mehrere Möglichkeiten zur Verfügung, das Risiko für den Werker zu quantifizieren. Eine vielversprechende Möglichkeit sind dabei die zweisträngigen unscharfe Logiken [Kiendl 1997], die im Rahmen dieser Arbeit an die Bedürfnisse der MRK angepasst und weiterentwickelt wurden (siehe dazu [Graf u.a. 2010b]).

6.1 Unscharfe Logiken für die Risikoquantifizierung

Eine grundsätzliche Besonderheit der unscharfe Logiken im Allgemeinen besteht darin, dass Experten–Wissen explizit in Form von allgemein verständlichen Regeln formuliert werden kann. Dagegen haben Methoden des maschinellen Lernens in den letzten Jahren große Fortschritte gemacht und sind deshalb ernstzunehmende Konkurrenten. In MAROCO wird versucht, beide Theorien in einem aktuellen Forschungsgebiet anzuwenden und gegeneinander zu evaluieren.

Diese Vorgehensweise führt zu einer impliziten multikriteriellen Optimierung. Es muss auf der einen Seite das Risiko für den Menschen minimiert werden und auf der anderen Seite die Laufzeit des Roboters maximiert werden.

¹engl. support vector machines.

Das Risiko für den Menschen wird in MAROCO minimiert, indem die maximale Arbeitsgeschwindigkeit durch MAROCO begrenzt werden kann, aber insbesondere eine proaktive Bahnplanungsmethode [Graf u.a. 2009b]. Die Risikoquantifizierung, erzielt durch die zweisträngige unscharfe Logik, spielt dabei eine entscheidende Rolle.

6.2 Grundlagen

Klassische (einsträngige) unscharfe Logiken sind im Allgemeinen folgendermaßen aufgebaut: Zuerst müssen reelle Eingangsgrößen, die in der Regel aus einem Prozess zur Merkmalsbildung stammen, dem Fuzzy-System eingegeben werden. Es kommt zur sogenannten Fuzzifizierung der Eingangsgrößen. Die Ausgabe der Fuzzifizierung sind linguistische Terme, wobei man zusätzlich zu den linguistischen Termen einen Glaubensgrad über die Erfüllung bestimmter linguistischer Terme erhält. Dieser Glaubensgrad ist in der Regel ein Skalar zwischen $[0, 1]$. Um die entstehenden Informationen weiterzuverarbeiten, müssen a priori unscharfe Operatoren festgelegt werden. Anschließend werden die Prämisse und die Konklusion berechnet. Die Prämisse stellt die unscharfe Verknüpfung von linguistischen Termen dar, wohingegen die Konklusion die Ausgabe nach der unscharfen Inferenz liefert. Die unscharfe Inferenz ist eine Entscheidungsfunktion in Form eines logischen Wenn-Dann-Operators und bestimmt die Konklusion. In der Regel werden mehrere Konklusionen parallel ausgegeben. Um später einen einzelnen Ausgabewert (das geschätzte Risiko) zu erzeugen, müssen die Konklusionen der Regeln akkumuliert werden. Eine mögliche Realisierung ist die Implementierung des Maximum-Operators. Dabei werden nur die maximalen Werte zugehörig zu den Risikowerten akkumuliert. Der letzte Schritt der Verarbeitungskette ist die Defuzzifizierung der akkumulierten Ausgabefunktion. Dies wird durch bspw. durch die Schwerpunktmethode erreicht. Weitere Details zu den einsträngigen unscharfen Logiken können bspw. in [Böhme 1993] nachgelesen werden.

Die zweisträngigen unscharfen Logiken sind eine natürliche Erweiterung der einsträngigen unscharfen Logiken. Es wird die Möglichkeit geboten, neben Empfehlungen von Regeln, explizit und unabhängig, Warnungen zu formulieren. Besonderheiten der zweisträngigen unscharfen Logiken sind bspw. eine vergleichsweise geringere Regelbasis, insbesondere für komplexe Aufgabenstellungen. Die Warnungen können explizit zu Verboten gemacht werden, was bei der Formulierung von Regeln für einen Experten zu einer natürlicheren Formulierung der Regelbasis führt. Das bedeutet, dass einsträngige Logiken gleich mächtig sind, aber in der Regel komplizierte Regeln benötigt werden um bspw. einfache Verbote zu realisieren. Ein entscheidender Unterschied liegt darin, dass die zweisträngigen Logiken eine zusätzliche Verarbeitung in Form einer sogenannten Hyperinferenzstrategie benötigen. Das bedeutet, dass man auf natürliche Weise Regeln in Form von Expertenwissen (Empfehlungen und Warnungen) integrieren kann, allerdings für den Preis der Implementierung einer Hyperinferenzstrategie. Die Hyperinferenzstrategie liefert eine Ausgabefunktion unter Berücksichtigung eines Vetos. In der Literatur [Kiendl 1997] werden derzeit verschiedene Arten von Vetos dargestellt (z.B. schwaches- und starkes-Veto). In MAROCO wird ein neuartiges Veto verwendet. Der Vorteil dieses Vetos liegt in der Berücksichtigung jeder Warnung, sofern die Warnungen einen Schwellwert nicht überschreiten. Das schwache Veto beachtet nur dann Warnungen, falls diese größer sind als die Empfehlung. Das starke Veto reagiert nur auf Warnungen die als Verbote auftreten, das bedeutet, dass die Gewichtung der Warnungen irrelevant sind. In MAROCO wird die Hyperdefuzzifizierung analog zur klassischen Defuzzifizierung (Schwerpunktmethode)

realisiert. Die speziellen Eigenschaften der Hyperdefuzzifizierung müssen für die hier vorliegende Anwendung nicht explizit berücksichtigt werden, da durch die Regelbasis definiert ist, dass die Sicherheit für den Menschen stets gewährleistet ist. Nach der Defuzzifizierung erhält man einen scharfen Ausgabewert.

6.3 Eine adaptierte Realisierung für MAROCO

Unschärfe Logiksysteme kodieren Expertenwissen, dass in Form von natürlich-sprachlichen Regeln formulierbar ist. Ein für diese Arbeit ausgewählter Satz von positiven und negativen Regeln für die Gewährleistung der Sicherheit des Menschen, ist in Tabelle 6.1 zusammengefasst.

Positive Regeln	
Antezedens	Sukzedens
$A \leftrightarrow$ sehr groß	$R \leftrightarrow$ außer Gefahr
$A \leftrightarrow$ groß $\wedge G \leftrightarrow (PG \vee PK \vee N)$	$R \leftrightarrow$ außer Gefahr
$A \leftrightarrow$ groß $\wedge G \leftrightarrow NG$	$R \leftrightarrow$ sicher
$A \leftrightarrow$ mäßig $\wedge G \leftrightarrow (PG \vee PK \vee N)$	$R \leftrightarrow$ außer Gefahr
$A \leftrightarrow$ mäßig $\wedge G \leftrightarrow NK$	$R \leftrightarrow$ sicher
$A \leftrightarrow$ mäßig $\wedge G \leftrightarrow NG$	$R \leftrightarrow$ mittelmäßig
$A \leftrightarrow$ gering $\wedge G \leftrightarrow PG$	$R \leftrightarrow$ sicher
$A \leftrightarrow$ gering $\wedge G \leftrightarrow PK$	$R \leftrightarrow$ mittelmäßig
$A \leftrightarrow$ gering $\wedge G \leftrightarrow N$	$R \leftrightarrow$ unsicher
$A \leftrightarrow$ gering $\wedge G \leftrightarrow NK$	$R \leftrightarrow$ riskant
$A \leftrightarrow$ gering $\wedge G \leftrightarrow NG$	$R \leftrightarrow$ gefährlich
$A \leftrightarrow$ gefährdet $\wedge G \leftrightarrow PG$	$R \leftrightarrow$ unsicher
$A \leftrightarrow$ gefährdet $\wedge G \leftrightarrow PK$	$R \leftrightarrow$ riskant
$A \leftrightarrow$ gefährdet $\wedge G \leftrightarrow (NG \vee NK \vee N)$	$R \leftrightarrow$ gefährlich
$OO \leftrightarrow$ abgewandt $\wedge KO \leftrightarrow$ abgewandt $\wedge G \leftrightarrow (NG \vee NK \vee N)$	$R \leftrightarrow$ gefährlich
Negative Regeln	
Antezedens	Sukzedens
$OO \leftrightarrow$ zugewandt $\wedge KO \leftrightarrow$ zugewandt	$R \leftrightarrow$ {gefährlich, riskant, unsicher} <i>unerwünscht</i>
$OO \leftrightarrow$ zugewandt $\wedge KO \leftrightarrow$ seitlich abgewandt	$R \leftrightarrow$ {gefährlich, riskant} <i>unerwünscht</i>
$OO \leftrightarrow$ abgewandt $\wedge KO \leftrightarrow$ abgewandt $\wedge G \leftrightarrow (PG \vee PK \vee N)$	$R \leftrightarrow$ {gefährlich, riskant} <i>unerwünscht</i>
$OO \leftrightarrow$ abgewandt $\wedge KO \leftrightarrow$ abgewandt $\wedge G \leftrightarrow (NG \vee NK \vee N)$	$R \leftrightarrow$ {außer Gefahr, sicher, mittelmäßig} <i>unerwünscht</i>

Tabelle 6.1: Positive und negative Regeln für ein in MAROCO realisiertes zwei-strängiges unscharfes Logik System.

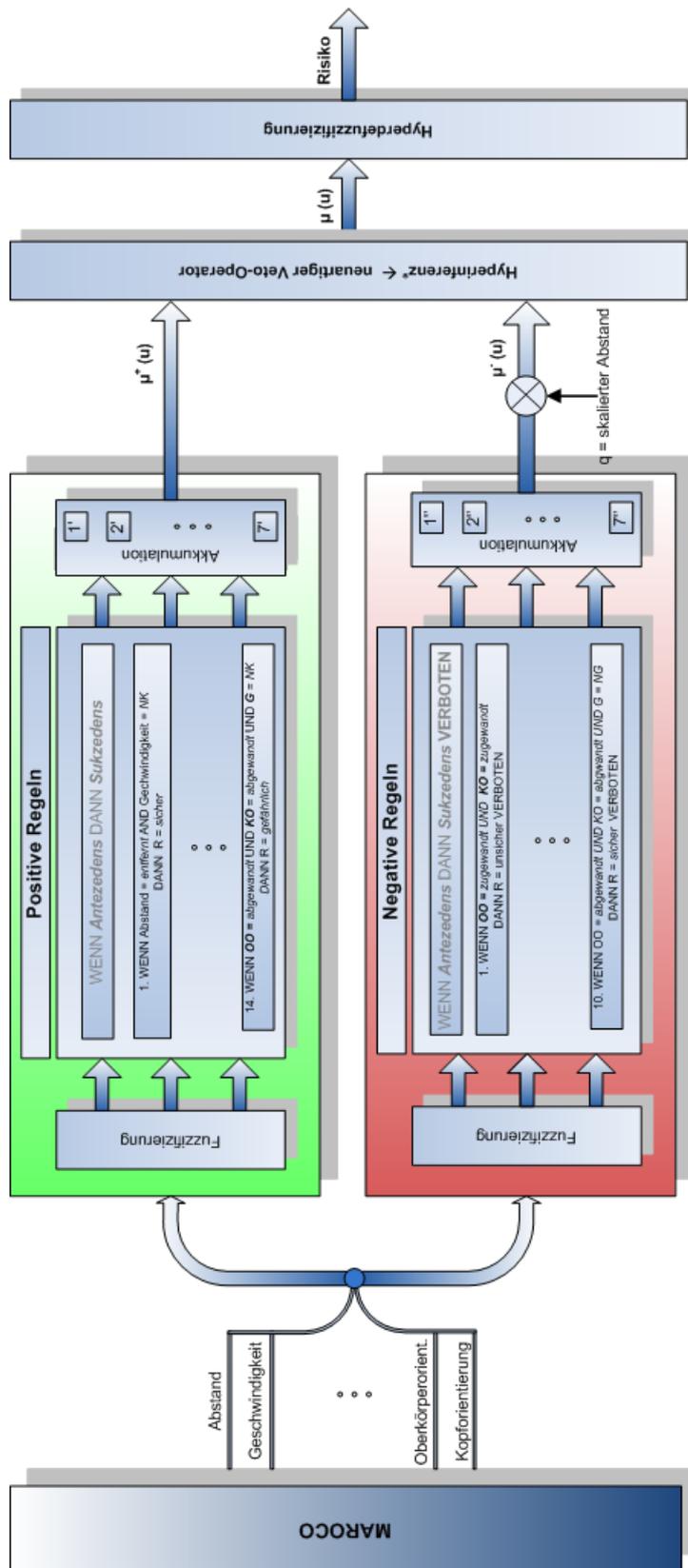


Abbildung 6.1: Adaptierte zweisträngige unscharfe Logik aus [Graf u.a. 2010b].

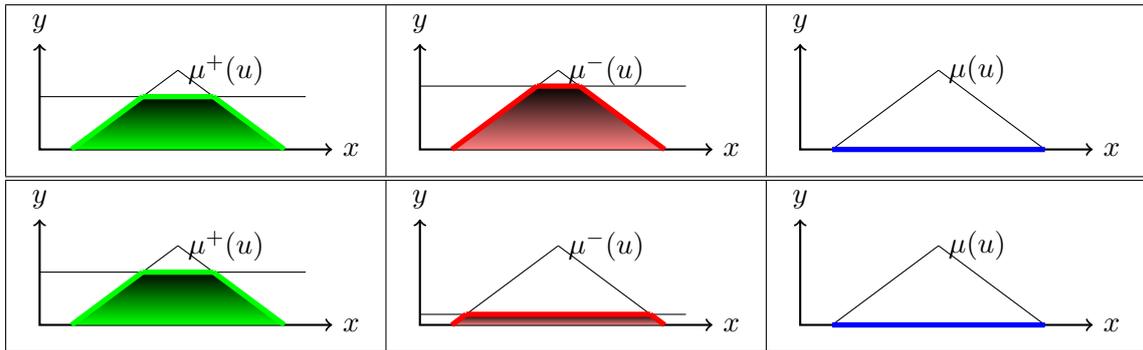


Abbildung 6.2: Starkes Veto.

Eine weitere Entwicklung, die im Rahmen dieser Arbeit geleistet wurde, ist die Bestimmung eines neuartigen Veto-Operators. In der Literatur (z.B. [Kiendl 1997]) findet man bspw. das schwache oder starke Veto. Beide Operatoren sind für das Vorhaben im Rahmen dieser Arbeit ungeeignet.

Die Veto-Operatoren kombinieren in unterschiedlicher Weise die aktivierten positiven und negativen Regeln. Die Operatoren sind der Kern der Hyperinferenz. Für die Bedürfnisse einer sicheren MRK hatten das starke und schwache Veto insbesondere folgende Nachteile:

1. Das starke Veto, definiert durch $\mu(u) = \begin{cases} \mu^+(u) & , \text{ falls } \mu^-(u) = 0 \\ 0 & , \text{ sonst} \end{cases}$, operiert sehr eingeschränkt (siehe Abbildung 6.2). Eine differenzierte Gewichtung in Abhängigkeit der Fläche unter der aktivierten positiven Regel wird nicht berücksichtigt. Das bedeutet, dass die Fläche unter der negativen Regel zu viel Gewicht bekommt, denn wann immer die Fläche nicht verschwindend gering ist, wird ein Verbot erzeugt. Damit wird auch die Flexibilität der zweisträngigen unscharfen Logik umgangen, denn der Gewichtungsfaktor wird außer Kraft gesetzt. Das bedeutet praktisch, dass die zweisträngige unscharfe Logik weniger effizient arbeitet als sie es eigentlich sollte.
2. Das schwache Veto, definiert durch $\mu(u) = \begin{cases} \mu^+(u) & , \text{ falls } \mu^+(u) \geq \mu^-(u) \\ 0 & , \text{ sonst} \end{cases}$, stellt prinzipiell ein duales Problem zum starken Veto dar. Der schwache Veto-Operator kann anhand der folgenden Beispiele und Abbildungen (siehe Abbildung 6.3) diskutiert werden:
 - (a) Die Fläche unter der positiven Regel ist kleiner als die eingeschlossene Fläche unter der negativen Regel. Es wird stets ein Verbot erzeugt. Dieser Fall ist erwünscht.
 - (b) Die positive Regel schließt eine größere Fläche ein als die negative Regel, unabhängig von der Differenz der beiden Flächen voneinander. In jedem dieser Fälle ist die Ausgabe äquivalent zur positiven Regel. Das bedeutet, dass die negative Regel keine Beachtung findet. Dieser Fall ist unerwünscht, da er zu einer undifferenzierten Bewertung führt.

Ein Kompromiss stellt das in [Graf u.a. 2010b] vorgestellte Veto dar. Der neuartige Veto-Operator wird gebildet, indem zuerst die Fläche μ^- wie folgt in drei bestimmte Teilflächen aufgeteilt wird. Dazu müssen die Schnittpunkte des α -Schnitts mit der Kurve bestimmt werden. Anschließend wird eine Senkrechte parallel MRK zur Ordinate gebildet (Lot). Dadurch

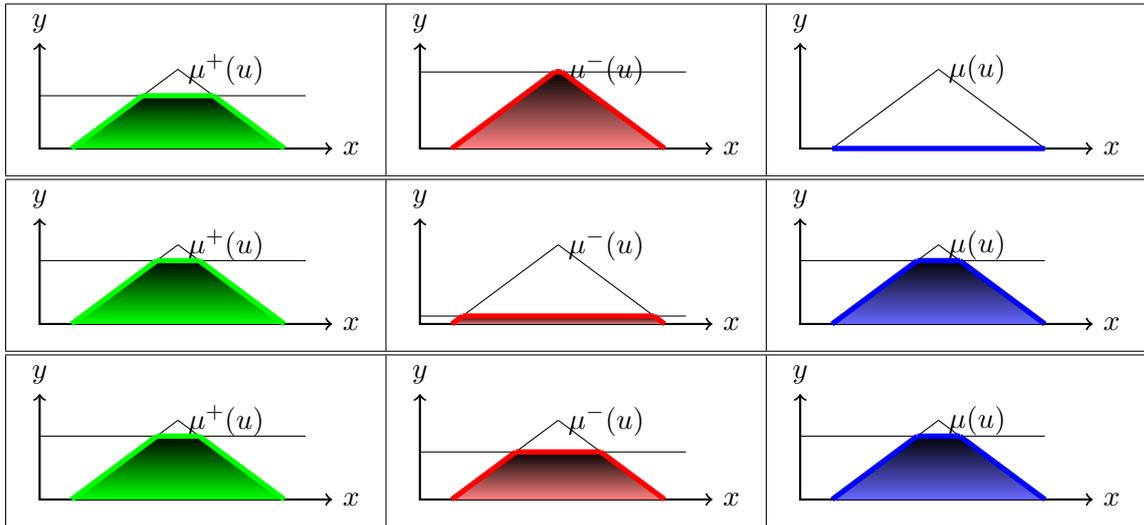


Abbildung 6.3: Schwaches Veto.

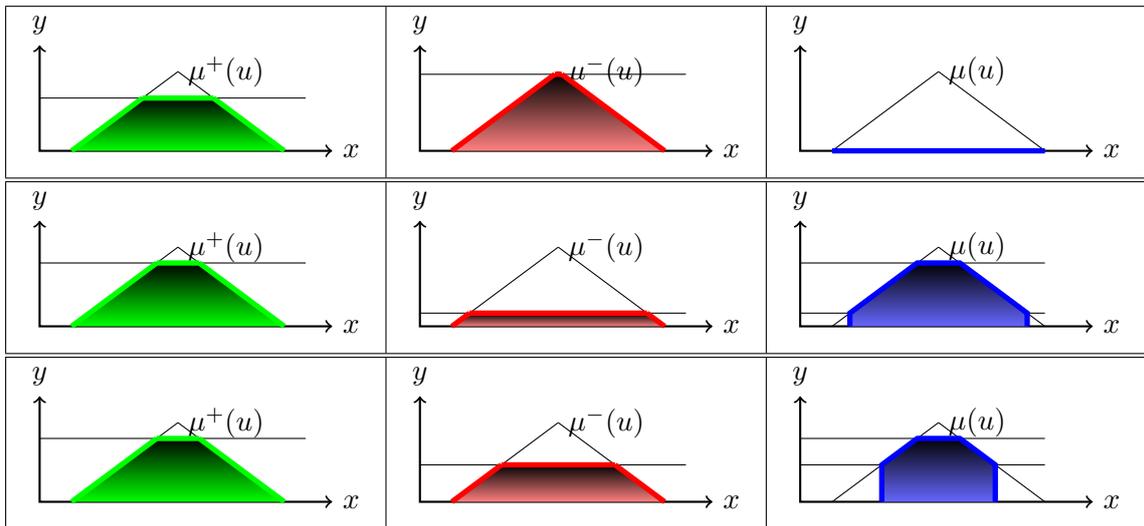
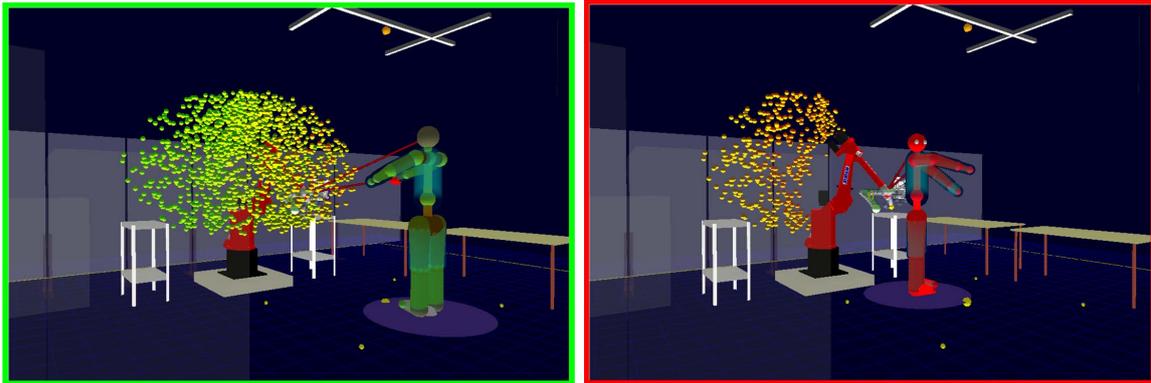


Abbildung 6.4: Eigenes Veto.

erhält man drei Flächenelemente. Die beiden äußeren Flächenelemente sind identisch, da es sich bei der Dreiecksfunktion um eine symmetrische Funktion handelt. Die äußeren Flächenelemente werden durch β bezeichnet. Um die gewünschte Fläche zu erhalten muss $\mu^+ \beta$ berechnet werden. Das daraus resultierende Flächenmaß lautet:

$$\mu(u) = \begin{cases} \mu^+(u) \beta(u) & , \text{ falls } \mu^+(u) > \mu^-(u) \\ 0 & , \text{ sonst.} \end{cases} \quad (6.1)$$



(a) Geringes Risiko.

(b) Erhöhtes Risiko.

Abbildung 6.5: Visualisierung des Risikos für einzelne Roboterkonfigurationen.

6.3.1 Diskussion

Eine Evaluierung verschiedenartiger Methoden zur Risikoquantifizierung für eine sichere MRK, basierend auf der Vorkenntnis kinematischer Parameter, wird in [Graf u.a. 2010b] vorgestellt. Die in der Abbildung 6.6 dargestellten Kurven stammen von Methoden, die im weiteren Verlauf durch M abgekürzt werden. Alle Methoden haben für ein sicheres Verfahren des Roboters gesorgt. Um die Experimente auszuwerten, wurde eine reale Sequenz mit einem Menschen in der Roboterzelle aufgezeichnet, wobei die Ausgaben des geschätzten Risikos durch die verschiedenen Methoden 1-5 mitaufgezeichnet wurden.

M1 basiert rein auf der kürzesten Abstandsberechnung zwischen dem Menschen und den Robotern. Der kürzeste Abstand schließt den Abstand der Arme zum Roboter mit ein. Dies ist besonders gut im ersten Teil des Experiments zu sehen (bis Bild 80). Da die Hände hier ausgestreckt, sich nahe am Roboter befinden, wird das Risiko durch M1 als hoch eingestuft. Die Methoden M2-M4 stellen linear gewichtete Gaußsche Verteilungen dar, die unterschiedlich komplex aufgebaut sind. Die Komplexität hängt dabei von der Anzahl der berücksichtigten Parameter ab.

Die in Abbildung 6.6 als unscharfe Methode gekennzeichnete Kennlinie repräsentiert die zwei-strängige unscharfe Logik.

Die in der Abbildung 6.6 unter den Methodennamen angegebenen Werte sind der Mittelwert und die Varianz der Kurvenverläufe. Durch die beiden Momente lassen sich die Effizienz und die Flexibilität der Methode quantifizieren. Man möchte gerne Methoden verwenden, die einen möglichst geringen Mittelwert haben und gleichzeitig flexibel auf gefährliche Situationen bei der Kooperation reagieren können.

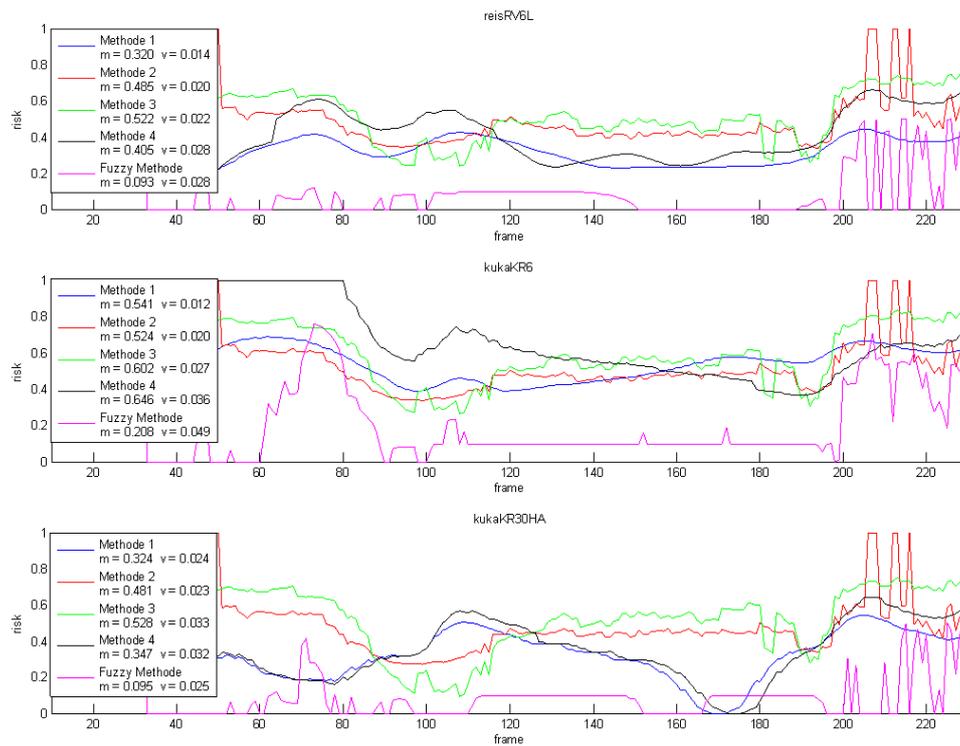


Abbildung 6.6: Ausgewählte Experimente aus [Graf u.a. 2010b].

Kapitel 7

Risikominimierung

Nachdem in den vorangehenden Kapitel beschrieben wurde, wie aus Merkmalen das Risiko quantifiziert wird, werden in den folgenden Abschnitten zwei Vorgehensweisen zur Minimierung des Risikos vorgestellt. Dabei wird nur kurz auf die erste Möglichkeit, die Geschwindigkeit der Roboters zu reduzieren, eingegangen. Wichtiger ist der zweite Abschnitt, der eine proaktive Bahnplanung für die sichere Mensch–Roboter–Kooperation vorstellt und an die Vorarbeiten der vergangenen Kapitel, insbesondere die Risikoquantifizierung, aufbaut.

7.1 Arbeitsgeschwindigkeit

Eine einfache Lösung zur Minimierung des Risikos besteht in der Begrenzung der maximalen Arbeitsgeschwindigkeit des Roboters. Die Begrenzung der maximalen Arbeitsgeschwindigkeit hat einen technischen Hintergrund bzgl. der Regelung der Roboter. Die Verwendung einer variablen oberen Grenze für die Arbeitsgeschwindigkeit des Roboters ist unproblematisch. Im Extremfall kann die maximale Arbeitsgeschwindigkeit zu Null gesetzt werden. Die Berücksichtigung dieser Vorgehensweise ist intuitiv, einleuchtend und wird gesetzlich gefordert (siehe 2.3 bzw. [DIN12453]).

Da die Begrenzung der maximalen Arbeitsgeschwindigkeit inzwischen zum Stand der Technik gehört, soll nicht weiter darauf eingegangen werden.

7.2 Eine Bahnplanung für die sichere MRK

Die Integration eines geschätzten Menschmodells in eine industrielle Robotikzelle ist neuartig. Die sich daraus ergebenden Möglichkeiten im Hinblick auf eine sichere Interaktion bzw. Kooperation zwischen Mensch und Roboter, sind umfangreicher gegenüber den Möglichkeiten bisheriger Systeme (siehe Abschnitt 2.2 zum Vergleich).

Deshalb wird in diesem Abschnitt eine Vorgehensweise vorgestellt, welche die Kinematik des Menschen (indirekt) bei der Bahnplanung berücksichtigt.

7.2.1 Grundlagen des Verfahrens

Klassische Ansätze zur Bahnplanung bilden die Problemstellung auf ein graphentheoretisches Suchproblem ab. Für diese Art von Problemen kennt man eine Vielzahl von Ansätzen, um kürzeste Wege zwischen Paaren von Punkten zu finden. Die Knoten des Graphen repräsentieren

in diesem Abschnitt die Winkelkonfigurationen des Roboters. Da im Rahmen dieser Arbeit nur die Konfiguration der ersten drei Achsen für den Menschen zu gefährlichen Konfigurationen führen können, lässt sich der Konfigurationsraum in 3D betrachten. Die Knoten, also die Verbindungspunkte zwischen den Kanten, sind gewichtet. Bei der Optimierung spielen die Knotengewichtungen eine entscheidende Rolle, da nach ihnen optimiert wird. In dieser Arbeit sind die Kantengewichte bestimmt in Abhängigkeit der Entfernung zwischen zwei Roboterkonfigurationen im Winkelraum, aber auch durch das quantifizierte Risiko. Deshalb findet eine multikriterielle Optimierung statt, denn auf der einen Seite möchte man einen effizient arbeitenden Roboter mit möglichst geringer Stillstandzeit, andererseits darf unter keinen Umständen die Gefährdung des Menschen riskiert werden.

Das im Rahmen dieser Arbeit entwickelte Bahnplanungsverfahren kann in zwei Phasen eingeteilt werden:

1. Eine Initialisierungsphase, welche am Anfang des Auswertezyklus durchgeführt wird.
2. Eine Laufzeitphase.

Bei der Initialisierung wird ein Graph im Konfigurationsraum des Roboters aufgebaut. Die Verteilung der Knoten geschieht unabhängig und gleichverteilt. Anschließend werden die Knoten durch ein k -nächste Nachbarn Verfahren mit Kanten verbunden. Anschließend wird ein initialer Pfad vom Start- zum Zielknoten berechnet. Die Berechnung des Pfades basiert auf dem heuristischen Suchverfahren A^* (siehe z.B. [Turau 1996], [Russell und Norvig 2003]).

In der Laufzeitphase findet die eigentliche proaktive Planung statt. In jedem Zeitschritt der aktuell bestimmten Bahn, wird eine Trajektorie (Interpolation und Überschleifen), für zwei auf der Bahn benachbarte Knoten berechnet. Diese lokale Optimierung macht die Anwendung einer geplanten Bahn möglich, da der Graph aus der Initialisierungsphase lediglich den möglichen Konfigurationsraum unterabtastet (digitalisiert).

Die errechneten Zwischenpunkte werden dann an die Robotersteuerung übergeben, welche anschließend die Kinematik des Roboters anpasst.

Wichtig bei der Anwendung der Bahnplanung ist die permanente Überprüfung auf erhöhte Kollisionsgefahr. Überschreitet diese Prüfung einen empirisch ermittelten Grenzwert, bzw. einen durch die DIN-Normen festgelegten Grenzwert, muss umgeplant werden. Dabei muss ein Kompromiss gefunden werden. Auf der einen Seite ist es unzureichend, nur den nächsten Knoten der Bahn zu betrachten, andererseits müssen nicht alle möglichen Knoten des Pfades überprüft werden. In dem hier vorliegenden Verfahren wird deshalb eine lokale Nachbarschaft von Winkelkonfigurationen überprüft.

Bei der Erzeugung eines Graphen muss berücksichtigt werden, dass eine Diskretisierung zu Freiräumen im Konfigurationsraum führt. In den Freiräumen können Objekte, deren Dimensionierung kleiner ist als die Abstände zwischen den Knoten, mit dem Roboter kollidieren (siehe Abbildung 7.2). Im Falle der hier vorliegenden Anwendung, kennt man eine Schätzung der Maße des sich in der Zelle befindenden Menschen und kann deshalb Grenzwerte festlegen, so dass eine sich in der Zelle bewegendes Person nicht in den Freiräumen des Konfigurationsraumes verschwindet.

Die Suche kann in einem Konfigurationsraum zu kritischen Rechenzeiten führen. Im Extremfall kann es zu einer Verletzung der Echtzeitfähigkeit kommen. Deshalb wird ein Verfahren eingesetzt, das neben anderen positiven Eigenschaften über die Möglichkeit verfügt, unterbrechbar zu sein. Das in [Graf u.a. 2009b] vorgestellte Verfahren ist von wichtiger praktischer Bedeutung für die sichere Mensch-Roboter-Kooperation.

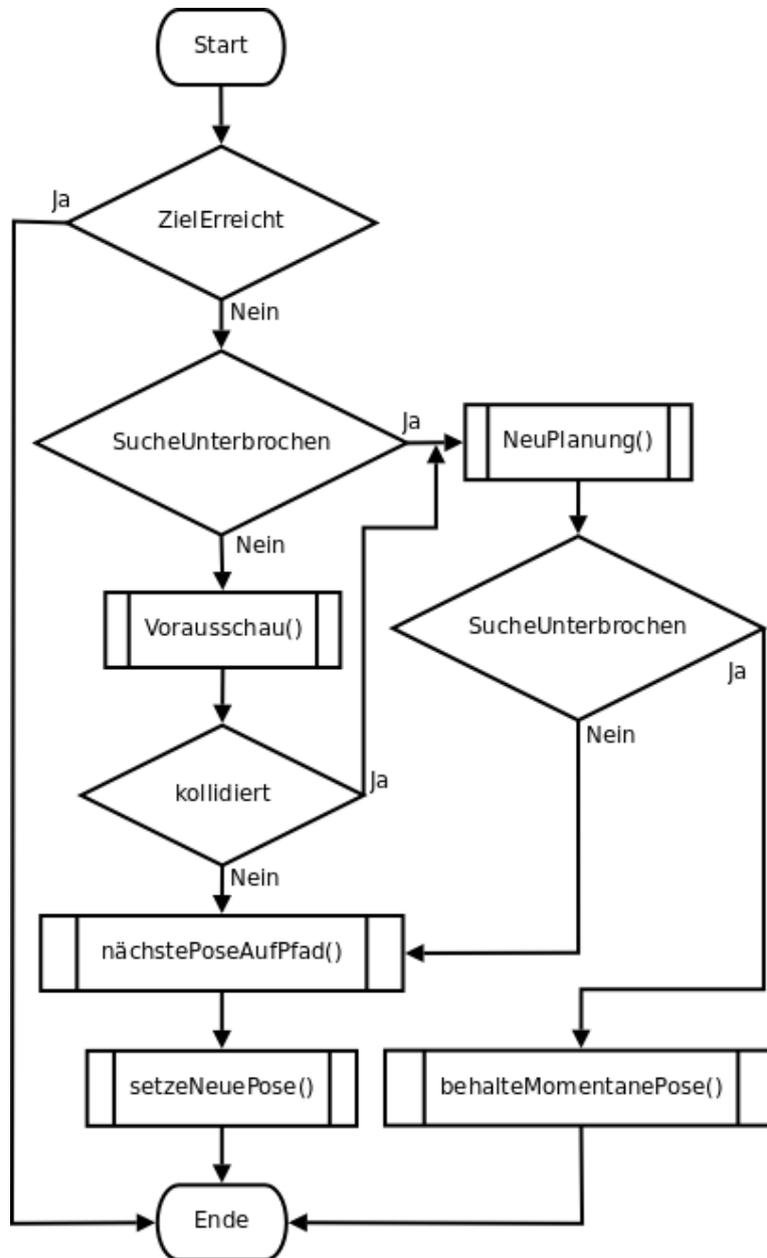


Abbildung 7.1: Vollständige Traversierung der geplanten Trajektorie mit präemptiver Suche.

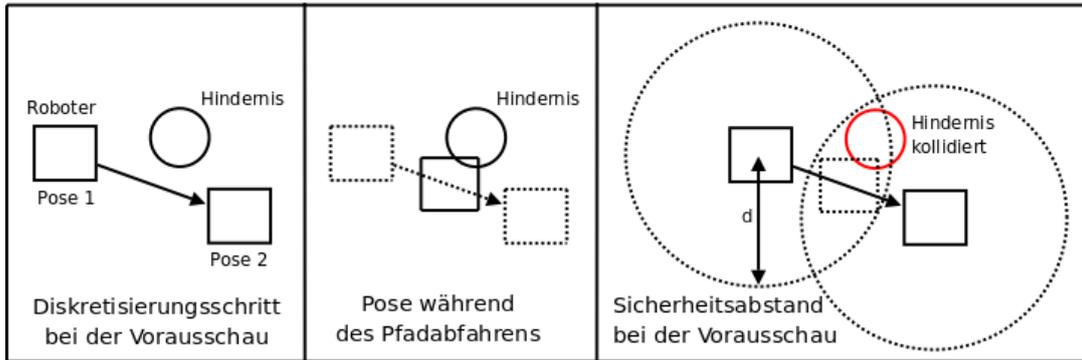


Abbildung 7.2: Vereinfachte Darstellung der Vorausschau mit Konfliktlösung.

Dauert die Berechnung eines Pfades länger als die frei zur Verfügung stehende Zeit, kann das Planungsverfahren unterbrochen werden und im nächsten Zeitschritt an der abgebrochenen Stelle der Suche fortgesetzt werden. Dazu müssen die erstellten *offenen* und *abgeschlossenen Listen* des Bahnplanungsalgorithmus zwischengespeichert werden (siehe Abbildung 7.3).

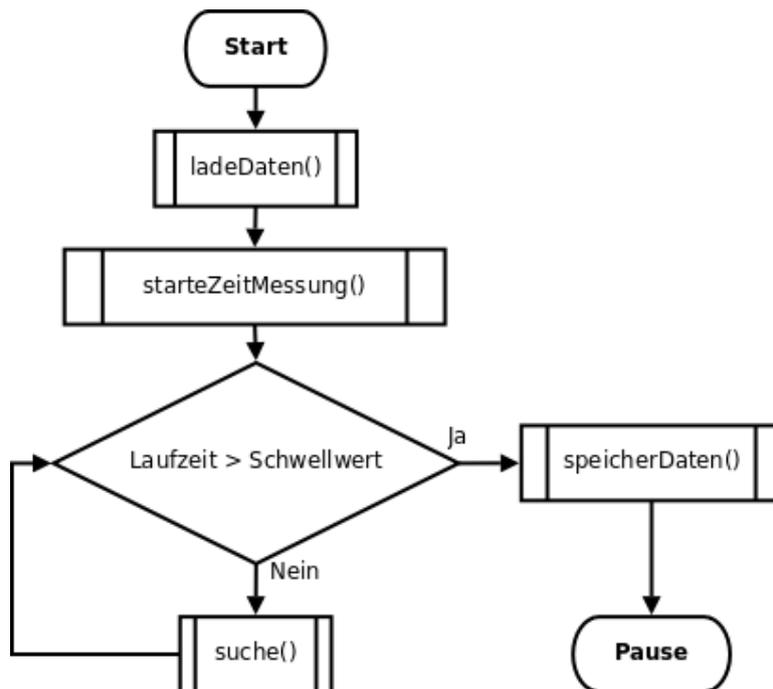


Abbildung 7.3: Präemptiver A*-Algorithmus.

Während der Bahnplanung kann sich eine Person frei in der Zelle bewegen. Die Präsenz einer Person in der Arbeitszelle schränkt den Suchraum ein, da neue, frei gewordenen Knoten nicht wieder in Betracht gezogen werden. Diese Vorgehensweise wird allerdings unpassende Graphen erzeugen, welche als nicht optimal zu bezeichnen sind. Deshalb wird eine Methode zur Zeitmessung eingesetzt, welche nach der Überschreitung eines bestimmten Zeitintervalls die Suche abbricht und anschließend versucht, einen neuen Suchpfad aufzubauen.

Die euklidische Distanz dient als Metrik für die Berechnung eines heuristischen Wertes und der Kosten einer Konfiguration. Die Bewertung wird während der Elternknoten-Expansion für jeden Knoten durchgeführt. Des Weiteren wird diese Vorgehensweise erweitert, um einen Kollisionstest für jede lokale Konfiguration, basierend auf der früher vorgestellten Risikoquantifizierung durchzuführen.

Zusätzlich zur Risikoquantifizierung wird auch der kürzeste Abstand zwischen Mensch und Roboter explizit berücksichtigt. Es wird von der Möglichkeit gebraucht gemacht, dass Knoten, die als nicht riskant eingestuft werden, aber einen geringen Abstand haben, verworfen werden. Die in MAROCO realisierte Risikoschätzung berücksichtigt auch den Bewegungstrend einer sich bewegenden Person. Bei einer schnellen Bewegung in Richtung des Roboters kann eine momentane Roboterkonfiguration als riskoreich eingestuft werden. In diesem Fall verweilt der Roboter in seiner aktuellen Konfiguration, da jeder neu geplante Knoten im Konfigurationsraum mit einem zu hohen Risiko belegt ist. Der Roboter verweilt solange in seiner Konfiguration, bis die Person durch eine Positionsänderung das Risiko selbst minimiert, oder der Roboter eine zulässige Bahn findet. Dieses Verhalten des Roboters bezeichnet man als proaktiv.

Im Rahmen dieser Arbeit wird der Konfigurationsraum auf drei Dimensionen beschränkt. Die ersten drei Achsgelenke repräsentieren die Koordinatenachsen des Konfigurationsraums. Die restlichen Roboterachsen werden außer Acht gelassen, müssen aber für spezielle Anwendungen in der Praxis eventuell mitberücksichtigt werden.

7.2.2 Experimente und Diskussion

Für die experimentelle Analyse stellt das Softwaresystem MAROCO eine ideale Umgebung dar, insbesondere im Hinblick auf die heuristisch einzustellenden Parameter, wie bspw. die Anzahl der abgetasteten Punkte des Konfigurationsraums. Die im Folgenden aufgelisteten Ergebnisse sind Simulationsergebnisse aus MAROCO.

Dazu wurden spezielle Videosequenzen durch MAROCO aufgezeichnet und ausgewertet. Die durch den Menschen erzeugten Bewegungen weisen unterschiedliche Muster auf (konstante Geschwindigkeiten, beschleunigte Bewegungen, etc.). Die Bewegungen unterliegen keinen Beschränkungen.

Während der Auswertung wurden Zeitstempel vergeben, die durch eine von MAROCO bereitgestellte Zeitmessung realisiert wird. Im Folgenden werden zwei ausgewählte Aufzeichnungen aufgelistet, welche Aufschluss geben über mögliche Vorfälle bei der hier vorgestellten Bahnplanung. Alle Zeiten werden dabei in ms angegeben.

Falls sich für die Vorausschau eine Kollisionsgefahr ergibt, wird die gegenwärtige Konfiguration in den existierenden Graphen integriert. Die Vorausschau dauert ca. 15ms. Die Integration dauert weniger als 1ms. Da jedoch ein Knoten integriert werden muss, wird der Schwellwert für die Unterbrechung für den ersten Suchdurchgang auf 15ms herunter gesetzt (üblich sind 30ms). Anschließend kommt es zu einer Umplanung. Die Umplanung muss nur einmal pausieren.

Der zweite Suchzyklus nimmt weniger Zeit in Anspruch als durch die obere Begrenzung erlaubt ist. Aus der Tabelle 7.2 lässt sich ablesen, dass es viele Unterbrechungen gibt und die Pfadsuche deshalb zu viel Zeit benötigt. Deshalb wird sie abgebrochen, um anschließend wieder neu gestartet zu werden.

Eine erfolgreiche Pfadsuche ergibt sich im zweiten Suchzyklus. Die erneute Integration der Roboterkonfiguration in den Graphen nach dem Abbruch ist nicht nötig, da die Dauer der Unterbrechung nur 30 ms in Anspruch nimmt.

Zeit	Funktion	Zustand
16344	traversePath()	start
16344	traversePath()	call isCollidingLookahead()
16359	traversePath()	return isCollidingLookahead()
16359	traversePath()	have collision. connect to graph
16359	traversePath()	connected (return from k_nearest())
16359	rePlanPath()	start
16359	rePlanPath()	start search
16375	A*	interruption: elapsed time: 16
16375	rePlanPath()	interrupted
16375	rePlanPath()	end
16375	traversePath()	end
16422	traversePath()	start
16422	traversePath()	current_path.size()=0
16422	rePlanPath()	start
16422	rePlanPath()	start search
16437	A*	found goal: elapsed time: 15
16437	rePlanPath()	found path
16437	rePlanPath()	end
16437	traversePath()	end – replanPath output length=4

Tabelle 7.1: Traversierung eines Pfades mit Kollision.

Zeit	Funktion	Zustand
59937	traversePath()	start
59937	traversePath()	current_path.size()=0
59937	rePlanPath()	start
59937	rePlanPath()	start search
59937	rePlanPath()	search is interrupted for too long. restart
59969	A*	interruption. elapsed time: 32
59969	rePlanPath()	interrupted
59969	rePlanPath()	end
59969	traversePath()	end – replanPath output length=0
60016	traversePath()	start
60016	traversePath()	current_path.size()=0
60016	rePlanPath()	start
60016	rePlanPath()	start search
60031	A*	found goal. elapsed time: 15
60031	rePlanPath()	found path
60031	rePlanPath()	end
60031	traversePath()	end – replanPath output length=4

Tabelle 7.2: Abbruch der Suche und Wiederaufruf.

Die in 7.3 dargestellten Experimente wurden für die Graphengrößen 600, 1000 und 2000 Knoten durchgeführt. Für jede Graphengröße wurde die Anzahl der Verknüpfungen zwischen den Knoten variiert (5, 10 und 20 Verknüpfungen). Jedes Experiment basiert auf ein und derselben Videosequenz und wurde 10 Minuten lang durchgeführt.

Für die Analyse sind folgende Punkte wesentlich:

- Zykluszeiten, bzw. die Bildrate.
- Die Menge an abgebrochen, erfolglosen und erfolgreichen Umplanungen.
- Das Verhältnis von Stillstandzeit zur Pfadverfahrzeit.

Kurze Zykluszeiten bedeuten hohe Bildraten. In der Simulation ist dieser Wert abhängig von der Geschwindigkeit der Algorithmen und dem Zeitbedarf der Festplattenzugriffe zum Einlesen der Bilddaten. Bei der realen Demonstration stellt die Kamera den Flaschenhals dar, sofern die Software schnellere Zykluszeiten liefert, was bei MAROCO der Fall ist.

Aus der Tabelle ist ersichtlich, dass je nach Konfiguration Bildraten zwischen 11,9 und 14,4 erreicht werden, wobei einige Konfigurationen mehr als 13,6 Bilder pro Sekunde erzielen. Obwohl alle Experimente mit dem gleichen zeitlichen Schwellwert (von 20 ms) ausgeführt wurden, kommt es zu diesen Unterschieden. Ein Zusammenhang zwischen den Zykluszeiten und der Anzahl der Umplanungsaufrufe kann erkannt werden. Daraus ergibt sich, dass dieser Wert für die Analyse indirekt von Bedeutung ist.

Zu beobachten ist auch, dass die Anzahl der erfolgreichen Umplanungen unabhängig von der Anzahl der Umplanungsaufrufe ist. Das selbe gilt für die Anzahl der erfolglosen und abgebrochen Suchen. Abgebrochene Suchen sind Suchen, die nach einer gewissen Zeit komplett abgebrochen und wieder neu gestartet werden. Das kann daran liegen, dass der Suchraum durch die Konfiguration der menschlichen Kinematik stark eingeschränkt wird, wodurch eine Vielzahl von Knoten aus der Suche ausgeschlossen werden. Ein Neustart der Suche ermöglicht eine neue Evaluation aller Knoten. Für erfolglose Suchen ergibt sich, dass kein Pfad zum Ziel gefunden werden konnte. Dieser Fall tritt bspw. ein, falls der Mensch mit dem Rücken zum Roboter steht und dadurch das Risiko für die Zielkonfiguration des Roboters über einen bestimmten Schwellwert steigt. Dieser Knoten wird dann in der Suche nicht betrachtet, im Extremfall kann das erfolgreiche Umplanen dadurch blockiert werden.

Die Werte für die Anzahl der erfolgreichen und abgebrochenen Suchen sind jedoch nicht unabhängig von der Graphenkonfiguration. Mit steigender Nachbarschaftsverknüpfung nehmen auch die abgebrochenen Umplanungen zu, denn der Suchraum wird enorm vergrößert, wodurch mehr Zeit für das Testen der einzelnen Möglichkeiten benötigt wird. Die Anzahl der erfolgreichen Umplanungen scheint unabhängig von der Anzahl der Knoten im Graphen zu sein. Allerdings kommen bei 2000 Knoten in der Regel weniger erfolgreiche Suchen vor. Ein Zusammenhang lässt sich bei Betrachtung der Nachbarschaftsverknüpfungen im Graphen erkennen. Bei 10 Verknüpfungen wird für eine Graphengröße von 1000 Knoten das Maximum erreicht. Der Wert für die erfolglosen Suchen ist stark von der Bildfolge abhängig und hängt nicht von der Konfiguration des Graphen ab.

Pfadzyklen beschreiben, wie häufig der Roboter von seinem Startpunkt A zu seinem Startpunkt B und wieder zurück zum Startpunkt A bewegt wurde. Pfadzyklen hängen deshalb von der Dauer des durchgeführten Experiments und der individuell beanspruchten Zeit jedes einzelnen Zyklus ab. Je länger ein Pfadzyklus dauert, desto weniger Zyklen können während des Experiments durchlaufen werden. Falls eine Umplanung durchgeführt wird, wird der Roboter

Knoten	600				1000				2000			
	#Nachbarn	5	10	20	5	10	20	5	10	20		
Ø Zykluszeit (in ms)	77,68	69,25	72,22	71,64	73,46	81,09	79,7	83,33	83,82			
Ø Bilder pro Sekunde	12,87	14,44	13,85	13,96	13,61	12,33	12,55	12,00	11,93			
Initiale Pfadlänge	39	23	17	46	29	22	65	42	31			
Ø Pfadlänge bei Umplanung	19,9	17,67	14,82	28,86	22,04	12,70	28,90	17,75	11,93			
Ø # Umplanungsaufufe	3992	1134	2408	1751	3097,5	5489,5	4805	5694,5	7015,5			
Ø # Erfolgreiche Suchen	49	52	66	48,5	57,5	40,5	35,5	24	9,5			
Ø # Abgebrochene Suchen	0	0	96	1	61,5	298,5	199	352,5	415			
Ø # Erfolgreiche Suchen	1350	792	410	300,5	769	483	377,5	166,5	1049			
Ø # Pfadzyklen	17	43	34	30	24,5	12	14,5	6,5	1			
Ø Zeit pro Pfadzyklus	34282,24	13865,51	17922,97	19784,26	24609,49	50120,00	47861,29	81806,10	106933,00			
Ø Stillstandzeit je Pfadzyklus	17231,6	1597,7	5355,2	3949,5	9548,7	38757,1	33269,5	68413,4	92915,5			
Stillstandzeit je Pfadzyklus (%)	50,26	11,52	29,88	19,62	38,80	77,33	69,51	83,63	86,89			

Tabelle 7.3: Resultate unterschiedlich dicht abgetasteter Konfigurationsräume.

solange angehalten, bis ein neuer Pfad zum Ziel gefunden ist. Die Summe aller Zeiten in denen der Roboter nicht verfährt, ergibt die Stillstandszeit. Man möchte eine möglichst geringe Stillstandszeit und kurze Pfadverfahrzeiten, wobei das Verhältnis der beiden Größen möglichst klein sein sollte. Die besten Werte sind für die Konfigurationen (600,10) und (1000,5) erzielt worden. Die Experimente mit einer Graphengröße von 2000 Knoten erzielen die schlechtesten Werte.

Die Pfadlängen, die initiale Suche und die Anzahl der Umplanungen sind keine direkten Gütekriterien. Ein dichter Pfad, d.h. ein Pfad mit vielen Knoten, ermöglicht unter Umständen ein Durchdringen von Engstellen im Suchraum. Wenige Knoten ergeben meist eine glattere Trajektorie. Die initialen Pfadlängen hängen von der Anzahl der Knoten im Graphen ab. Die Pfadlängen der Umplanungen sind bspw. abhängig von der Position und Orientierung des Menschen gegenüber dem Roboter und dem Zustand, indem sich der Roboter befindet (Entfernung zum Ziel, initialer Pfadverlauf, etc.).

Mit den hier durchgeführten Experimenten stellt man fest, dass eine Graphenkonfiguration mit 600 Knoten und 10 Nachbarschaftsverknüpfung die besten experimentellen Werte erzielt hat. Alternativen stellen noch die Konfiguration mit 1000 Knoten, 5 Verknüpfungen und 600 Knoten und 20 Verknüpfungen dar.

Weshalb die Konfigurationen (600,10) und (1000,5) ähnlich gut abschneiden, soll im Folgenden erläutert werden.

Die Graphenkonfiguration (600,5) kommt den anderen Konfigurationen mit 600 Knoten nicht sehr nahe, da der Suchraum durch die 5 Verknüpfungen stark eingeschränkt ist. Diesen Nachteil kann die Konfiguration (1000,5) durch mehr Knoten ausgleichen. Die Vergrößerung des Suchraums ist allerdings bei der Konfiguration (1000,10) schon zu stark. Deshalb wird in diesem Fall ein schlechtes Ergebnis erzielt. Für die Konfigurationen (600,10) und (600,20) ist die Vergrößerung weniger signifikant. Es kann sogar passieren, dass die Werte sich verbessern, z.B. bei der Konfiguration (600,10), im Allgemeinen aber nur wenig schlechter werden.

Die Abbildung 7.4(a) zeigt den Roboter bei der Abfahrt der initial geplanten Trajektorie. Dann bewegt sich die Person in Richtung des Roboters. Das Risiko nimmt aufgrund des Bewegungstrends der Person zu. Allerdings kommt es noch nicht zu einer Umplanung der ursprünglichen Trajektorie.

Im unteren Bild 7.4(b) wird schließlich umgeplant, da sich bei der Abfahrt der ursprünglichen Bahn eine Kollision mit der Person ergeben kann.

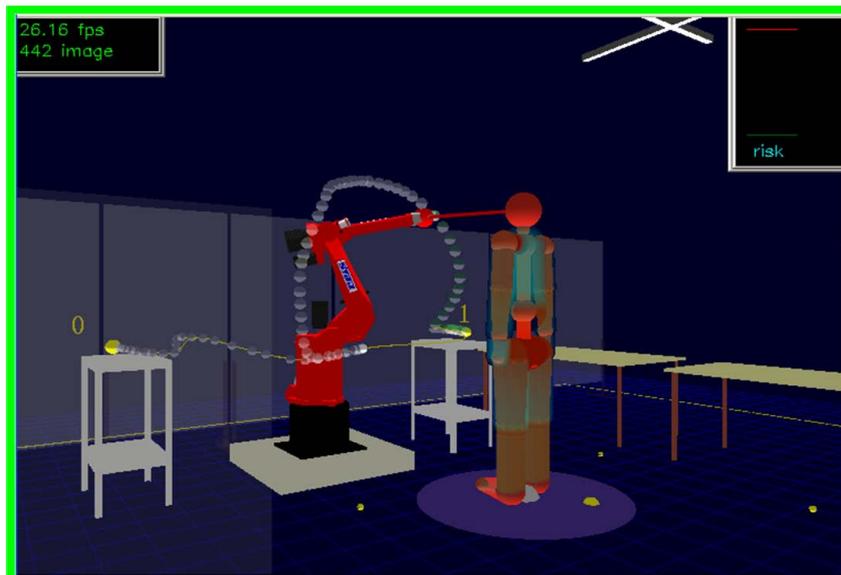
Der resultierende Pfad der Neuplanung hängt stark von der Konfiguration der Kinematiken und insbesondere von den blockierten Knoten des Konfigurationsraumes ab. In Abbildung 7.5 werden einige Resultate von Neuplanungen dargestellt. Die dicken blaue Geradensegmente repräsentiert den initial geplanten Pfad. Die hellgrünen Kreuze repräsentieren die Knoten des Graphen. Die roten Kreuze visualisieren die benachbarten Knoten des Start- bzw. Zielknoten. Die Nachbarknoten werden durch das verwendete Ballungsverfahren ausgewählt. Die verbleibenden Pfade sind durch unterschiedliche Farben zu unterscheiden und repräsentieren die Resultate der Umplanung.

Die hier vorgestellte proaktive Pfadplanung ist in der Lage, auf eine bevorstehende Kollision mit einem Menschen zu reagieren. Eine mögliche Umplanung findet echtzeitfähig zur Laufzeit statt. Die Experimente haben auch gezeigt, dass die Berechnungszeiten für Ausweichbewegungen mehr Zeit in Anspruch nehmen können, als durch die Taktung der Kamera zugelassen.

Um die Echtzeitbedingungen nicht zu verletzen, ist eine modifizierte A*-Suche implementiert, welche die Fähigkeit der Präemption besitzt, d.h. unterbrechbar ist. Das Verfahren kann deshalb während der Planung eingefroren werden und zu einem späteren Zeitpunkt fortgesetzt



(a) Kollisionsgefahr.



(b) Umgeplante Bahn.

Abbildung 7.4: Darstellung der proaktiven Umplanung. (a) Ursprüngliche Bahnplanung mit potentieller Kollisionsgefahr. (b) MAROCO plant die Bahn des Roboters um, um damit das Risiko für den Menschen zu minimieren.)

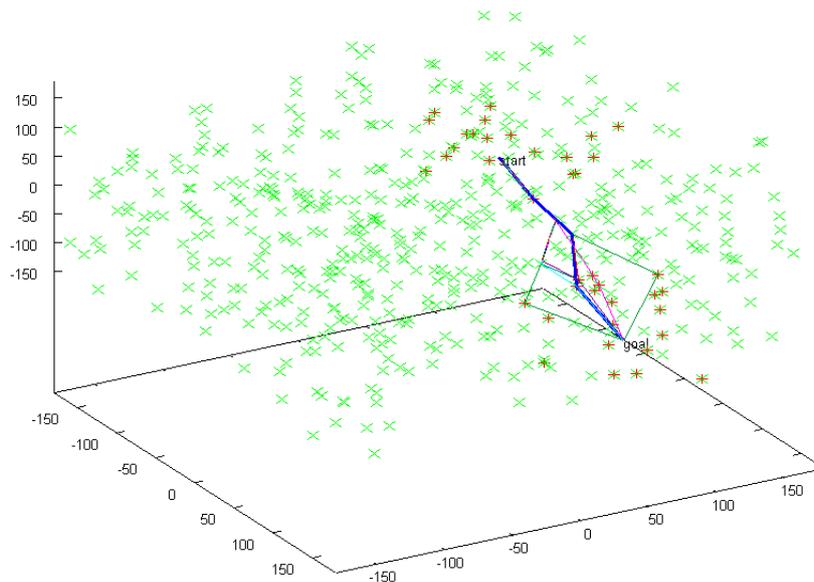


Abbildung 7.5: Visualisierung der Umplanung im Konfigurationsraum.

werden. Die Experimente haben gezeigt, dass das Verfahren nahe an die Taktung der Kamera herankommt. Die Pfadplanung kann auf einem eigenen Prozessorkern parallel ausgeführt werden. Die sich dabei ergebende Latenz (1 Zeitschritt) kann durch Einstellung der Reaktion auf das geschätzte Risiko kompensiert werden. Durch Integration dieser Pfadplanung in MAROCO wird gezeigt [Graf u.a. 2009b], dass eine effizientere Interaktion zwischen Mensch und Roboter möglich ist.

Zusätzlich ist eine Vorausschau integriert (siehe Abbildung 7.6), die eine voreilige Umplanung verhindern kann. Zukünftige Konfigurationen, entlang der geplanten Bahn, werden unter Berücksichtigung des geschätzten Risikos und des kürzesten Abstandes bewertet. Ausgehend von diesen Fähigkeiten, kann das Risiko für den Menschen minimiert werden. In Abbildung 7.6 sind weitere ausgewählte Szenekonfiguration aus einer Bildfolge dargestellt. Die Abbildungen in 7.7 sind die korrespondierenden Risikoschätzungen dargestellt als Knoten im Konfigurationsraum der ersten drei Roboterachsen.

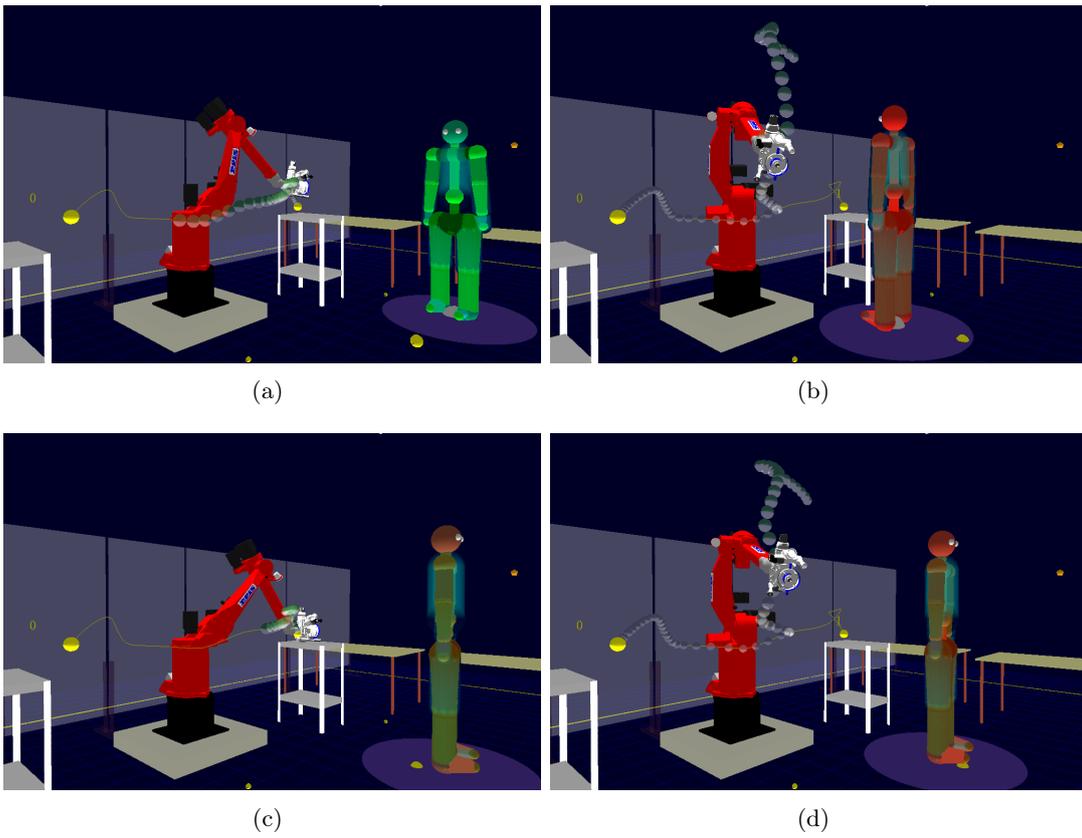
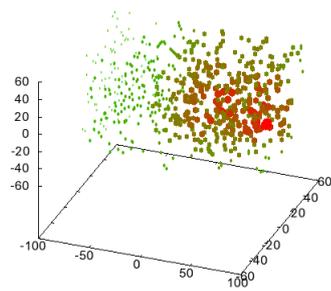
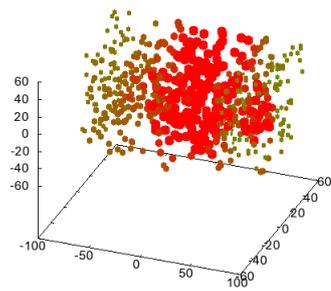


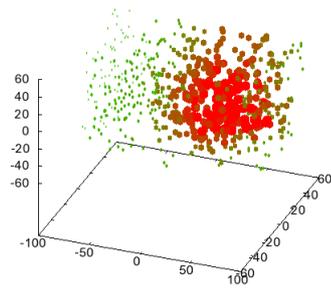
Abbildung 7.6: Szeneabbildungen mit unterschiedlich geschätztem Gefahrenpotential für den Menschen.



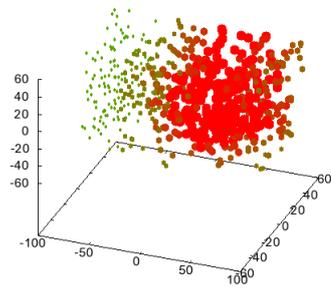
(a)



(b)



(c)



(d)

Abbildung 7.7: Darstellungen der zu Abbildung 7.6 korrespondierenden Konfigurationsräumen mit unterschiedlichen Risikopotentialen.

Kapitel 8

Zusammenfassung und Ausblick

Die im Rahmen dieser Arbeit entwickelten und realisierten Systemarchitekturen für eine sichere Kooperation zwischen Mensch und Roboter, arbeiten echtzeitfähig im Takt der Kamera, d.h. mit mehr als 15 Bildern pro Sekunde. Derzeit ist kein bekanntes vergleichbares System in der Lage, mit nur einem einzelnen tiefenauflösenden Sensor die Szene zu bewerten und anschließend eine Bahnplanung durchzuführen, welche das Risiko für den Menschen minimiert, bei gleichzeitiger Optimierung der Weglänge.

Die Besonderheit der hier vorgestellten Architekturen liegt in der Auswahl und Weiterentwicklung von bekannten oder neu entwickelten Verfahren. Es wurde darauf geachtet, den Stand der Forschung aus benachbarten Disziplinen zu beachten und auf dieser Grundlage Modifikationen und Erweiterungen für die hier vorgestellte Problemstellung zu entwickeln. Dadurch wurden einerseits verschiedene Verfahren die, auch aufgrund ihrer modularen Entwicklung, in zukünftigen Systemen für die Mensch–Roboter–Kooperation eine Rolle spielen werden oder aber in anderen, artverwandten Gebieten zum Einsatz kommen können. Andererseits wurde aber durch die entwickelten Methoden eine in sich abgeschlossene Systemarchitektur realisiert, welche letztendlich nicht nur abstrakte Theorie oder im Stadium einer reinen Simulation blieb, sondern während der Zeit der Doktorarbeit am Lehrstuhl unter realen Bedingungen verschiedenen nationalen und internationalen Forscher- und Entwicklergruppen in Form von realen Demonstrationen vorgestellt wurde.

8.1 Einordnung

Die hier vorliegende Arbeit beschreibt erstmalig, unter Verwendung eines einzelnen tiefenauflösenden Sensors, verschiedene neuartige und weiterentwickelte Methoden, die durch eine Systemarchitektur geordnet, eine echtzeitfähige Mensch–Roboter–Kooperation im industriellen Umfeld realisiert.

Die Realisierung ermöglicht, dass eine einzelne Person mit bis zu drei Robotern in Interaktion tritt. Insbesondere die Minimierung des Risikos für den Arbeiter durch die Bewegungen der Roboter stellt dabei ein herausforderndes Ziel dar.

Vielmehr soll durch diese Forschungsarbeit ein möglicher Weg gezeigt werden, der zukünftige Mensch–Maschine Systeme inspiriert und beim Entwurf der Systeme hilft.

Das Besondere an dem hier vorgestellten System ist die Fähigkeit, die Vorverarbeitung und Auswertung der Bildfolgen, sowie die Merkmalsextraktion und Merkmalsverarbeitung und schließlich die Auswertung der Merkmale für die Rekonstruktion der Teile einer menschlichen

Kinematik und des sich ergebenden Risikos für den Menschen während der Interaktion mit dem Roboter in Echtzeit robust zu schätzen.

Eine wichtige Rolle bei dem vorgeschlagenen Auswertezyklus spielen verschiedene Verfahren aus angrenzenden Forschungsgebieten. Die markerlose Schätzung bestimmter Körperteile aus den Tiefenbildfolgen des Kamerasystems, die effiziente Berechnung des OF durch anspruchsvolle numerische Verfahren und deren Lösung auf heutzutage verfügbaren Grafikprozessoren, die Vorverarbeitung von Tiefenbildfolgen, die Einschätzung des Risikos durch zweisträngige, unscharfe Logiken und die proaktive Bahnplanung, sind nur einige wichtige Errungenschaften, die im Laufe dieser Arbeit (weiter-)entwickelt wurden.

Es ist gelungen, verschiedene Verfahren nicht nur theoretisch zu formalisieren, sondern darüber hinaus praktisch auf heutzutage verfügbaren Grafikkarten mit speziellen Grafikprozessoren zu implementieren. Die Echtzeitfähigkeit der gesamten Architektur wäre ohne Einsatz von effizienten Lösern auf Grafikkartenprozessoren zum heutigen Zeitpunkt nicht möglich gewesen. Abschließend sollen kurz nochmals die wichtigen Entwicklungen der hier vorliegenden Arbeit zusammengefasst werden:

- Entwurfsebene
 - Entwicklung einer Systemarchitektur in zwei Schritten. Daraus resultieren zwei echtzeitfähige Architektur-Entwürfe in Bezug auf den Kameratakt:
 - * Eine auf reiner Bildfolgenauswertung basierende Architektur.
 - * Die Hauptarchitektur, welche merkmalsbasiert arbeitet und u.a. eine proaktive Bahnplanung integriert.
- Bildfolgenauswertung
 - Effiziente, anisotrope Glättung von Bilddaten, basierend auf orientierter Medianfilterung.
 - Weiterentwicklung eines Hintergrundschätzverfahrens, basierend auf Gaußschen-Mischverteilungen, angepasst an die technischen Gegebenheiten des PMD-Sensors.
 - Integration effizienter numerischer Löser für ein hochgenaues OF Schätzverfahren, basierend auf der Methode der konjugierten Gradienten (Echtzeitfähigkeit).
 - Entwicklung der Methode der Dynamischen Konturen, basierend auf einem Variationsansatzes und effizienter Initialisierung.
 - Verallgemeinerung des kombiniert lokal-globalen Ansatzes (CLG) zur Schätzung des OF, basierend auf einem anisotropen Datenterm (XCLG).
 - Effiziente Lösung der 2D XCLG-OF Methode durch Mehrgitterverfahren auf GPU.
 - Weiterentwicklung des 2D XCLG-OF Verfahrens, hin zu einem echtzeitfähigen 3D OF Verfahren.
- Robuste Merkmale
 - Markerlose Schätzung von 13 kinematischen Parametern.
 - Entwicklung eines Verfahrens von sich kreuzenden Armen und einer Verdeckungsauflösung zwischen den Armen eines Menschen und einem Roboter.
 - Systematische Stabilisierung der geschätzten Kinematikparameter, basierend auf einem modifizierten Kalman-Filter mit Wiener-Prozess Beschleunigungsmodell.

- Integration der OF Schätzungen in das KF Modell.
- Entwicklung eines durch ein Kameramodell rückprojizierbaren Hüllkörpermodells zur Bewertung der Güte des geschätzten Kinematikmodells.
- Risikoschätzung
 - Quantifizierung des Risikos durch Adaption zweisträngiger Unscharfen–Logik.
 - Bildung eines Risikopotentialfelds für die sichere Mensch–Roboter–Kooperation.
 - Vorschlag von einem Regelsatz für die Interaktion zwischen Mensch–Maschine, auf der Grundlage des geschätzten Kinematikmodells.
 - Entwicklung eines neuartigen Operators für die Hyperinferenz.
- Risikominimierung
 - Entwicklung einer echtzeitfähigen, inkrementellen und proaktiven Bahnplanung, basierend auf dem geschätzten Risikopotentialfeld.
- Software und Demonstrator
 - Entwicklung eines Software–Rahmenwerks (MAROCO) zur Integration der aufgeführten Methoden und damit zur Realisierung der gesamten Systemarchitektur.
 - Aufbau einer realen Roboterzelle und Test der Verfahren unter realen Bedingungen.
 - Demonstration der Tauglichkeit des PMD–Sensorprinzips für eine sichere Mensch–Roboter–Kooperation.

Darüber hinaus wurden alle wichtigen Verfahren experimentell evaluiert und damit auch eine prinzipielle Tauglichkeit des verwendeten Kamerasystems gezeigt.

8.2 Ausblick

Mensch–Roboter–Kooperation ist ein vergleichsweise junges Forschungsgebiet. Im Rahmen dieser Arbeit konnte nur eine spezielle Grundlagenarbeit geleistet werden, von der zukünftige Systeme profitieren. Allerdings ist man mit dem hier entwickelten Softwaresystem MAROCO in der Lage, spezialisierte Bedürfnisse für Anwendungsfelder der Mensch–Roboter Kooperation zu generieren.

Mit Spannung wird auch zukünftig verfolgt, ob es den Herstellern gelingt, für die 3D Kamerasysteme Sicherheitszertifikate zu erhalten, um damit den Weg für neue innovative Domänen in der Automatisierungstechnik frei zu machen.

Insbesondere die Weiterentwicklung im Hinblick auf komplexere Szenarien stellt eine besondere Herausforderung für zukünftige industrielle Systeme dar. Zukünftige Systeme, insbesondere im industriellen Umfeld, sollten folgende Aufgabenfelder berücksichtigen:

- Die Kooperation mit mehreren beteiligten Akteuren, insbesondere Menschen, stellt eine besondere Herausforderung für die Perzeption der technischen Systeme dar, insbesondere bei der hier vorgeschlagenen Methode, die Kinematik der Menschen markerlos zu rekonstruieren. Die Verfahren zur Auflösung von sich verdeckenden Armen und die durch die Hüllkörper möglich gemachte schnelle Abstandsberechnung und Verdeckungsanalyse mit Robotern sind dabei ein erster Fingerzeig, wie diese Problematik zukünftig behandelt werden könnte.

- Ein weiterer interessanter Aspekt wäre die Berücksichtigung von Roboter–Roboter Kooperation, auch im Hinblick auf die Integration von 3D optischen Systemen, aber insbesondere unter Berücksichtigung der dabei entstehenden multikriteriellen Anforderungen für die Bahnplanung und Optimierung der Arbeitsabläufe.
- Fasst man die ersten beiden Forderungen zusammen und verallgemeinert diese, dann wird der Wunsch laut, dass zukünftige Auswertezyklen in der Lage sein sollen, mehrere Aktoren, d.h. Menschen und Roboter, in komplexen Szenarien zusammen arbeiten zu lassen.
- Weiterentwicklung des Szeneverstehens, z.B. durch Integration nicht–klassischer Logiksysteme. Dabei könnte die Erzeugung natürlich–sprachlicher Protokolle durch das technische System eine besonders nützliche und herausfordernde Aufgabe sein.
- Ein besonders wichtiger Punkt für zukünftige Systeme ist die Integration mehrerer 3D Kameras in die Zelle um die Abschattungsbereiche zu minimieren. Dabei entstehen interessante wissenschaftliche Herausforderungen für die Fusion von Daten, aber auch für die Auswertung.
- Sicherlich werden zukünftig die nützlichen 3D Sensoren mit anderen Sensoren kombiniert, wodurch neuartige technische Sehsysteme entstehen, die biologischen Vorbilder nahe kommen oder diese übertreffen und von deren Möglichkeiten die kognitive Industrierobotik profitieren wird.
- Auch im Hinblick auf die Handlungsplanung wird es zukünftig Herausforderung geben, die sich an der Komplexität der Anforderungen messen lassen müssen. Insbesondere auf Logiken-basierende Handlungsplanung kann ein geeignetes Mittel sein, um diese Aufgabe zu beherrschen.

Abbildungsverzeichnis

2.1	Abbildung eines achsparallelen Stereosystems.	8
2.2	Darstellung zur Epipolargeometrie.	9
2.3	Darstellung eines Menschen im Raum, aufgezeichnet mit dem PMD CamCube 2.0 (204 × 204-Bildpunkte Auflösung).	13
2.4	Vier-phasige Abtastung eines sinusmodulierten Signals.	15
2.5	Modifizierte Systemarchitektur nach [Ebert 2003b].	23
2.6	Einordnung in die Sicherheitskategorien nach [DIN13849].	30
2.7	Einordnung in die Sicherheitskategorien nach [DIN954-1].	31
3.1	Abstandsberechnung zwischen Mensch und Roboter, basierend auf einem Hüllmodell.	37
3.2	Eine Architektur für die sichere Mensch-Roboter-Kooperation als Auswertezyklus von Bildfolgen.	39
3.3	Illustration zur Berechnung ausgewählter kinematischer Parameter, basierend auf Tiefendaten zur markerlosen Rekonstruktion einer vereinfachten menschlichen Kinematik.	41
3.4	Zwei unterschiedliche Aufnahmen einer Person mit der Tiefenbildkamera und ihre Rekonstruktion in das virtuelle Umgebungsmodell von MAROCO.	43
3.5	(a) Vermessung der erweiterten Demonstratorzelle. (b) Visualisierung des zweiten Demonstratoraufbaus.	45
3.6	Darstellung einer komplexen rekonstruierten Szene durch MAROCO.	46
4.1	Abbildungen für das lineare Modell nach [Perona und Malik 1990].	49
4.2	Abbildungen für das anisotrope Modell nach [Perona und Malik 1990].	50
4.3	Visualisierung einer Instenitätsstruktur mit eingezeichneten Eigenvektoren, resultierend aus einer Strukturtensor-Analyse.	52
4.4	Visualisierung eines Eigenvektorfeldes mit unterschiedlichen Grauwertvariationen.	53
4.5	Verschiedene Iterationsstufen für ein medizinisches Bild mit starken Grauwertvariationen.	54
4.6	Resultate der bilateralen Filterung.	58
4.7	Überblick zu den untersuchten nicht-linearen Filtern.	60
4.8	Darstellung der Resultate nach der Anwendung ausgewählter nicht-linearer Filter.	61
4.9	Darstellung der Tiefen- und Amplitudeninformation.	70
4.10	Ausgewählte Bilder der Experimente zu unterschiedlichen Zeitpunkten. Die unterschiedlichen Tiefendaten für die Bewertung wurden an den gekennzeichneten Stellen in den Bildern entnommen.	73

4.11	Vergleich der Mittelwert–Verläufe der einzelnen Gaußmodelle für den Punkt P1 zwischen dem Verfahren von [Stauffer und 1999] und dem hier weiterentwickelten Verfahren.	74
4.12	Vergleich des Prioritätenverlaufs der einzelnen Gaußmodelle im Fixpunkt 1 zwischen dem Verfahren von [Stauffer und 1999] und dem hier weiterentwickelten Verfahren.	74
4.13	Vergleich des Varianzenverlaufs der einzelnen Gaußmodelle im Fixpunkt 1 zwischen dem Verfahren von [Stauffer und 1999] und dem hier weiterentwickelten Verfahren.	75
4.14	Vergleich des Mittelwertverlaufs der einzelnen Gaußmodelle im Fixpunkt 2 zwischen dem Verfahren von [Stauffer und 1999] und dem hier weiterentwickelten Verfahren.	76
4.15	Vergleich des Prioritätenverlaufs der einzelnen Gaußmodelle im Fixpunkt 2 zwischen dem Verfahren von [Stauffer und 1999] und dem hier weiterentwickelten Verfahren.	76
4.16	Vergleich des Varianzenverlaufs der einzelnen Gaußmodelle im Fixpunkt 2 zwischen dem Verfahren von [Stauffer und 1999] und dem hier weiterentwickelten Verfahren.	77
4.17	(Links) Hintergrundsubtraktion ohne Berücksichtigung der Intensitätsbilder. (Rechts)Hintergrundsubtraktion mit Berücksichtigung der Intensitätsbilder. . . .	77
4.18	Anpassung einer parametrischen Kurve, ausgehend von einer Ellipse, an die Kontur einer menschlichen Silhouette.	86
4.19	Divergenter Fall, bei Initialisierung mit der Kontur des vorhergehenden Tiefenbildes ohne zusätzliche externe Kraft.	88
4.20	Konvergenter Fall, erzielt mit der in [Graf und Wörn 2008a] entwickelten Methode.	89
4.21	Visualisierung der Treppenfunktion (linke Abbildung) und der Exponentialfunktion mit $\sigma = 0.25$ (rechte Abbildung).	90
4.22	Ausgewählte Bilder, erzeugt durch die Bewegungen eines menschlichen Arms. Die blauen Punkte repräsentieren Knotenpunkte der Kontur des vorhergehenden Zeitschrittes. Die gelben Punkte repräsentieren korrespondierende Knotenpunkte der Kontur nach Anwendung der Gradientenabstiegsgleichung. (Oben) Zwei Iterationen, d.h. $\chi_1(t - 2)$. (Unten) Konvergenzfall.	91
4.23	Ausgewählte Resultate aus einer Bildfolge mit einer sich in der Szene bewegendem Person.	92
4.24	Ausgewählte Resultate aus einer Bildfolge mit einer sich in der Szene bewegendem Person.	93
4.25	Projektion der Verschiebung eines Punktes in der Szene auf die Bildebene. . . .	94
4.26	(Oben): Homogene Bewegung des Oberkörpers und der Arme. (Unten): Bewegung des Oberkörpers und der Arme in unterschiedliche Richtungen.	94
4.27	(a) – (c): An lokale Strukturen angepasste Faltungsfunktionen \tilde{K}	98
4.28	(b) Punkte in orange beschreiben initiale Punktmenge des DBSCAN Algorithmus. (c) Starke Gewichtung der Winkeldifferenz. (d) Der Gewichtungparameter ω_α wurde sehr klein gewählt.	99
4.29	(a) Original. In (b), (c) wurden unterschiedliche Schwellwerte t_1 und t_2 festgelegt. Dadurch konnten mehr oder weniger strukturierte Bildpunkte identifiziert werden, die später der Ballungsanalyse übergeben werden.	100

4.30	(a) Darstellungen zur Dichtezugehörigkeit (Definition 1), (b) Dichteerreichbarkeit (Definition 2) und Dichteverbundenheit (Definition 3).	101
4.31	(a) Mit DBSCAN gefundene Kante. (b) Schätzung der Orientierung und Ausbreitung. (c) Umschließendes Rechteck der Ballung. (d) Bestimmung der Faltungsfunktion \tilde{K} für einen Ballungspunkt.	103
4.32	(a), (b): Diskrete, isotrope Gauß-Funktion und der daraus gebildete Ableitungsoperator. (c), (d): Diskrete, anisotrope gedrehte Gauß-Funktion und der daraus gebildete Ableitungsoperator.	105
4.33	Restriktion mit der Methode der vollen Gewichtung (1D).	110
4.34	Lineare Interpolation (1D).	111
4.35	V-Zyklus mit der maximalen rekursiven Tiefe: $t_{max} = 3$	113
4.36	Ablaufdiagramm des μ -Zyklus mit der maximalen rekursiven Tiefe: $t_{max} = 3$ und $\mu = 2$	114
4.37	Position eines Punktes im 3D-Koordinatensystem der Kamera.	117
4.38	Verschiebung eines Objekts im 2D-Koordinatensystem der Kamera.	118
4.39	Ausgewählte Bilder einer Bildfolge zur Visualisierung des 3D-OF.	119
4.40	Stillstehender Oberkörper, mit sich aufeinander zu bewogender Arme. (a) Frame zum Zeitpunkt t . (b) Frame zum Zeitpunkt $t + 1$. (c) Ergebnis der Ballungsanalyse. (d) OF-Schätzung der CLG-Methode, mit $\rho = 2.5$. (e) OF-Schätzung der XCLG-Methode. Beide Schätzungen wurden nachträglich bzgl. des Hintergrunds segmentiert. (f) Visualisierung der Verteilung der Betragsdifferenzen, mit $d = 1.26$	120
4.41	Vergrößerter Ausschnitt der OF-Schätzung. Der Bereich entspricht Abbildung 4.40 (f). Die roten Pfeile beschreiben die Schätzung der CLG-Methode, die grünen Pfeile die der XCLG-Methode. Der OF wurde an jeder zweiten Position in horizontale und vertikale Richtung eingezeichnet.	121
4.42	Homogene Bewegung des gesamten Körpers. (a) Frame zum Zeitpunkt t . (b) Frame zum Zeitpunkt $t + 1$. (c) Ergebnis der Ballungsanalyse. (d) OF-Schätzung der CLG-Methode, mit $\rho = 2.5$. (e) OF-Schätzung der XCLG-Methode. Beide Schätzungen wurden nachträglich bzgl. des Hintergrunds segmentiert. (f) Visualisierung der Verteilung der Betragsdifferenzen, mit $d = 1.47$	122
4.43	Vergrößerter Ausschnitt der OF-Schätzung. Der Bereich entspricht Abbildung 4.42 (f). Die roten Pfeile beschreiben die Schätzung der CLG-Methode, die grünen Pfeile die der XCLG-Methode. Der OF wurde an jeder dritten Position in horizontale und vertikale Richtung eingezeichnet.	123
5.1	Beispiel zur Berechnung der Orientierung des Kopfes und der Schultern. In (a) und (b) ist das Originalbild und das segmentierte Bild zu sehen. In (c) bzw. (e) und (d) bzw. (f) sind die Bildpunkte des segmentierten Körperteils und die an die Daten angepasste Ellipse zu sehen. Abbildung (g) fusioniert die Resultate.	130
5.2	Sich ergebende Teilpfade nach der Skelettierung durch Einsatz des rekursiven Pfadfinders.	131
5.3	Richtungskodierung.	132
5.4	Richtungsmasken. Die Bedeutung der Matrixeinträge ist folgendermaßen: '0' → Bildpunkt darf kein Element des Pfades sein, '1' → Bildpunkt muss Element des Pfades sein und '-' → Bildpunkt kann Element des Pfades sein.	132
5.5	Drei mögliche Konfigurationen von Armstellungen nach der Segmentierung.	133

5.6	(a) Annahme, dass keine Verdeckung vorliegt. (b) Annahme, dass eine Verdeckung vorliegt. (c) Schätzung des Armkreuzungspunktes.	135
5.7	(a) Resultat einer geschätzten Kreuzung der Arme. (b) Resultat einer Armverdeckung.	136
5.8	Analyse der Armstellungen.	137
5.9	Schätzung der Position im Zentrum der Bildfläche bei Tageslicht und bei Dunkelheit.	140
5.10	Regulierende Funktion für den Ausgleich kleiner und größerer Messstörungen.	141
5.11	Experiment mit fixiertem Oberarmwinkel (ca. 320°).	142
5.12	Schätzung der Position im Zentrum der Bildfläche bei Tageslicht.	143
5.13	Schätzung der Position im Zentrum der Bildfläche bei Dunkelheit.	144
5.14	Zeit-diskretes System. Linker Block: Prozess-Modell. Rechter Block: Mess-Modell	145
5.15	Visualisierung des Mensch-Modells.	146
5.16	(a) OF-Schätzung zum Zeitpunkt $t - 1$. (b) Umgebung zur OF-Schätzung des Körpermittelpunktes. (c) Prädiktion (blau) des Körpermittelpunktes zum Zeitpunkt t	150
5.17	(a) OF-Schätzung zum Zeitpunkt $t - 1$. (b) Umgebungen zur OF-Schätzung der Hand, Ellbogen und Schulter. (c) Prädiktionen (blau) der Hand, Ellbogen und Schulter zum Zeitpunkt t	150
5.18	Repräsentation der kinematischen Strukturen mit Kugelmodellen.	153
5.19	Experimentelle Auswertung des Verdeckungsgrades. Auf der Abszisse sind die einzelnen Bilder abgebildet, die Ordinate gibt den Verdeckungsgrad wieder.	156
5.20	Ausgewählte Bilder zur Visualisierung des Verdeckungsgrades.	157
6.1	Adaptierte zweisträngige unscharfe Logik aus [Graf u.a. 2010b].	163
6.2	Starkes Veto.	164
6.3	Schwaches Veto.	165
6.4	Eigenes Veto.	165
6.5	Visualisierung des Risikos für einzelne Roboterkonfigurationen.	166
6.6	Ausgewählte Experimente aus [Graf u.a. 2010b].	167
7.1	Vollständige Traversierung der geplanten Trajektorie mit präemptiver Suche.	171
7.2	Vereinfachte Darstellung der Vorausschau mit Konfliktlösung.	172
7.3	Präemptiver A*-Algorithmus.	172
7.4	Darstellung der proaktiven Umplanung. (a) Ursprüngliche Bahnplanung mit potentieller Kollisionsgefahr. (b) MAROCO plant die Bahn des Roboters um, um damit das Risiko für den Menschen zu minimieren.)	178
7.5	Visualisierung der Umplanung im Konfigurationsraum.	179
7.6	Szeneabbildungen mit unterschiedlich geschätztem Gefahrenpotential für den Menschen.	180
7.7	Darstellungen der zu Abbildung 7.6 korrespondierenden Konfigurationsräumen mit unterschiedlichen Risikopotentialen.	181

Tabellenverzeichnis

4.1	Kategorisierung von Intensitätsstrukturen, modifiziert nach [Jähne 1993].	52
4.2	Zusammenfassender, qualitativer Vergleich ausgewählter Filtermethoden.	59
4.3	Benchmark: OF-Berechnung der PMD Tiefenbilder.	124
4.4	Benchmark: OF-Berechnung der Yosemite Sequenz.	124
4.5	Benchmark: Vorverarbeitung	125
4.6	Benchmark: Ballungsanalyse der PMD Tiefenbilder	125
4.7	Benchmark: Ballungsanalyse der Yosemite-Sequenz	125
4.8	Benchmark: ST-Erstellung der PMD Tiefenmessungen	125
4.9	Benchmark: ST-Erstellung der Yosemite Sequenz	125
4.10	Benchmark: Lösung des LGS – PMD Messungen	126
4.11	Benchmark: Lösung des LGS – Yosemite Sequenz	126
5.1	Die diskreten Kalman-Formeln, modifiziert nach [Gelb 1974].	138
5.2	Empirisch ermittelte Werte für ϵ_1 und ϵ_2	142
5.3	Quantifizierung der Überdeckungsgrade durch die beiden ersten statistischen Momente.	156
6.1	Positive und negative Regeln für ein in MAROCO realisiertes zwei-strängiges unscharfes Logik System.	162
7.1	Traversierung eines Pfades mit Kollision.	174
7.2	Abbruch der Suche und Wiederaufruf.	174
7.3	Resultate unterschiedlich dicht abgetasteter Konfigurationsräume.	176

Literaturverzeichnis

- [AnisotropeDiffusion] www.mathworks.de/matlabcentral/fileexchange/24549 Letzter Besuch: September, 2009.
- [Asfour u.a. 2006] T. Asfour, K. Regenstein, P. Azad, J. Schröder, A. Bierbaum, N. Vahrenkamp und R. Dillmann: *Armar-III: An Integrated Humanoid Platform for Sensory-Motor Control*. In Proc. IEEE-RAS International Conference on Humanoid Robots. Genoa, Italy, December 2006, S. 169–175.
- [Azad u.a. 2008] P. Azad, T. Asfour und R. Dillmann: *Robust Real-time Stereo based Markerless Human Motion Capture*. In Proc. IEEE-RAS International Conference on Humanoid Robots. Dajeon, Korea, 2008, 8 Seiten.
- [Baker u.a. 2007] S. Baker, D. Scharstein, J.P. Lewis, S. Roth, M.J. Black and R. Szeliski: *A Database and Evaluation Methodology for Optical Flow*. In Proc. 11th International Conference on Computer Vision, ICCV, 2007, S. 1–8.
- [Bar u.a. 2001] Y. Bar-Shalom, X. R. Li, T. Kirubarjan: *Estimation with Application to Tracking and Navigation*. John Wiley and Sons Inc., New York, USA, 2001.
- [Barron u.a. 1994] J.L. Barron, D.J. Fleet, S.S. Beauchemin and T.A. Burkitt: *Performance of Optical Flow Techniques*. International Journal of Computer Vision, 12:1, 1994, S. 43–77.
- [Bigün u.a. 1991] J. Bigün, G. Granlund, and J. Wiklund: *Multidimensional Orientation Estimation with Application to Texture Analysis and Optical Flow*. In IEEE Trans. Pattern Analysis and Machine Intelligence, 13:8, 1991, S. 775–790.
- [Bilmes 1998] Jeff A. Bilmes: *A Gentle Tutorial of the EM Algorithm and its Application to Parameter Estimation for Gaussian Mixture and Hidden Markov Models*. U.C. Berkeley, 1998.
- [Böhme 1993] G. Böhme: *Fuzzy-Logik – Einführung in die algebraischen und logischen Grundlagen*. Springer Verlag, Berlin, Heidelberg, Deutschland, 1993.
- [Boult u.a. 1999] Boult, T.E.; Micheals, R.; Gao, X.; Lewis, P.; Power, C.; Yin, W.; Erkan, A.: *Frame-rate Omnidirectional Surveillance and Tracking of camouflaged and occluded Targets*. Second IEEE Workshop on Visual Surveillance, S. 48–55, Fort Collins, Colorado. 26 June 1999.

- [Briggs u.a. 2000] W.L. Briggs, V.E. Henson, and S.F. McCormick: *A Multigrid Tutorial: Second Edition*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2000.
- [Bronstein u.a. 1999] I.N. Bronstein, K.A. Semendjajew, G. Musiol, and H. Mühlig: *Taschenbuch der Mathematik*. 4. überarbeitete und erweiterte Auflage der Neubearbeitung, Verlag Harri Deutsch, Frankfurt am Main, Thun, 1999.
- [Brox u.a. 2004] T. Brox, A. Bruhn, N. Papenberg and J. Weickert: *High Accuracy Optical Flow Estimation Based on a Theory for Warping*. In Proc. 8th European Conference on Computer Vision, Lecture Notes in Computer Science 3024, T. Pajdla und J. Matas (Eds.), vol. 4, S. 25-36, Prague, Czech Republic, Springer-Verlag Berlin, Heidelberg, May 2004.
- [Bruhn u.a. 2005a] A. Bruhn, J. Weickert and C. Schnörr: *Combining Local and Global Optical Flow Methods*. International Journal of Computer Vision 61:3 (2005) 211–231.
- [Bruhn u.a. 2005b] A. Bruhn, J. Weickert, C. Feddern, T. Kohlberger, and C. Schnörr: *Variational Optical Flow Computation in Real Time*. IEEE Transactions on Image Processing 14:5 (2005) 608–615.
- [Bruhn 2006] A. Bruhn: *Variational Optical Flow Computation – Accurate Modelling and Efficient Numerics*. Dissertation, Universität des Saarlandes, Saarbrücken, Juli 2006.
- [Burghart u.a. 2002] C. Burghart, S. Yigit, O. Kerpa, D. Osswald und H. Wörn: *Concept for Human Robot Co-Operation Integrating Artificial Haptic Perception*. In Proceedings of IAS 7. Marina del Rey, USA, 2003, S. 38–45.
- [Calion 2009] S. Calion: *Robot Programming by Demonstration*. EPFL Press, Lausanne, Switzerland, 2009.
- [Chwatal u.a. 2001] W.Chwatal, P. Borejko: *Eine Untersuchung der Amplituden refraktierter Wellen*. Institut für Geodäsie und Geophysik, TU Wien, 2001.
- [Ebert und Henrich 2001] D. Ebert und D. Henrich: *Safe Human-Robot-Cooperation: Problem Analysis, System Concept and Sensor Fusion*. In IEEE International Conference on Sensor Fusion. Baden-Baden, Germany, August 2001, 6 Seiten.
- [DIN12100] Deutsches Institut für Normung: *Sicherheit von Maschinen - Allgemeine Gestaltungsleitsätze, Risikobeurteilung und Risikominimierung*. DIN EN ISO 12100, März, 2009.
- [DIN13849] Deutsches Institut für Normung: *Sicherheit von Maschinen - Sicherheitsbezogene Teile von Steuerungen*. DIN EN ISO 13849, Oktober, 2006.
- [DIN13855] Deutsches Institut für Normung: *Sicherheit von Maschinen - Anordnung von Schutzeinrichtungen im Hinblick auf Annäherungsgeschwindigkeiten von Körperteilen*. DIN EN ISO 13855, Februar, 2002.
- [DIN292] Deutsches Institut für Normung: *Sicherheit von Maschinen*. DIN EN 292, November, 1991.

- [DIN10218] Deutsches Institut für Normung: *Industrieroboter Sicherheit*. DIN EN ISO 10218, Februar, 2007.
- [DIN775] Deutsches Institut für Normung: *Industrieroboter Sicherheit*. DIN EN 775, August, 1993.
- [DIN953] Deutsches Institut für Normung: *Trennende Schutzeinrichtungen*. DIN EN 953, November, 1997.
- [DIN954-1] Deutsches Institut für Normung: *Sicherheitsbezogene Teile von Steuerungen*. DIN EN 954-1, November, 1997.
- [DIN1050] Deutsches Institut für Normung: *Leitsätze zur Risikobeurteilung*. DIN EN 1050, Januar, 1997.
- [DIN12453] Deutsches Institut für Normung: *Nutzungssicherheit kraftbetätigter Tore*. DIN EN 12453, Februar, 2002.
- [Do u.a. 2008] M. Do, P. Azad, T. Asfour und R. Dillmann: *Imitation of Human Motion on a Humanoid Robot using Nonlinear Optimization*. In Proc. IEEE-RAS International Conference on Humanoid Robots. Dajeon, Korea, 2008, S 545–552.
- [Durand und Dorsey 2002] F. Durand und J. Dorsey: *Fast Bilateral Filtering for the Display of High-Dynamic-Range Images*. In Proc. ACM SIGGRAPH 2002, San Antonio, USA, Juli 2002, S. 257–266.
- [Ebert und Henrich 2002a] D. Ebert und D. Henrich: *Safe Human-Robot-Cooperation: Image-Based Collision Detection for Industrial Robots*. In IEEE International Conference on Intelligent Robots and Systems. Lausanne, Switzerland, September 2002, 6 Seiten.
- [Ebert und Henrich 2002b] D. Ebert und D. Henrich: *Das SIMERO-System*. In OTS-Systeme in der Robotik. OTS 2002. München, Germany 2002, Kap. 5, S. 1–17.
- [Ebert und Henrich 2003a] D. Ebert und D. Henrich: *SIMERO: Sichere Mensch-Roboter-Koexistenz*. Technischer Bericht, AG Eingebettete Systeme und Robotik (RESY), FB Informatik, Universität Kaiserslautern, Mai 2003, S. 1–14.
- [Ebert 2003b] D. Ebert: *Bildbasierte Erzeugung kollisionsfreier Transferbewegungen für Industrieroboter*, Dissertation, Fachbereich Informatik der Technischen Universität Kaiserslautern, 2003.
- [EG2006] Richtlinie 2006/42/EG des Europäischen Parlaments und des Rates: *Maschinenrichtlinie 2006/42/EG*. Mai, 2006.
- [EG98] Richtlinie 98/37/EG des Europäischen Parlaments und des Rates: *Maschinenrichtlinie 98/37/EG*. Juni, 1998.
- [Elad 2002] M. Elad: *On the Origin of the Bilateral Filters and Ways to Improve It*. In IEEE Trans. on Image Processing, 11(10), S. 950–953, 2002.
- [Enkelmann 1987] W. Enkelmann: *Investigation of multigrid algorithms for the estimation of optical flow fields in image sequences*. Computer Vision, Graphics and Image Processing 43: (1987) 150–177.

- [Ester et. al. 1996] M. Ester, H. Kriegel, J. Sander and X. XU: *A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise*. E. Simoudis, J. Han and U. M. Fayyad: Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining (KDD-96), 1996. S. 226-231.
- [Faugeras 1993] O. Faugeras: *Three Dimensional Computer Vision: A Geometric Viewpoint*, The MIT Press, Cambridge, Massachusetts, London England, 1993.
- [Faugeras und Luong 2001] O. Faugeras und Q. T. Luong: *The Geometry of Multiple Images*, The MIT Press, Cambridge, Massachusetts, London England, 2001.
- [Förstner und Gülch 1987] W. Förstner und E. Gülch: *A Fast Operator for Detection and Precise Location of Distinct Points, Corners, and Centres of Circular Features*. In Proceedings of the ISPRS Conference on Fast Processing of Photogrammetric Data, S. 281–305, Interlaken, Juni 1987.
- [Forsyth und Ponce 2003] D. A. Forsyth und J. Ponce: *Computer Vision: A Modern Approach*, Prentice Hall, Pearson Education, Inc., Upper Saddle River, NJ 07458, 2001.
- [Fuchs und Hirzinger 2008] S. Fuchs und G. Hirzinger: *Extrinsic and Depth Calibration of ToF-Cameras*. In IEEE Conference on Computer Vision and Pattern Recognition (CV-PR), Anchorage, USA, Juni 2008, 6 Seiten.
- [Galvin u.a. 1998] G. Galvin, B. McCane, K. Novins, D. Mason, and S. Mills: *Recovering Motion Fields: An Analysis of Eight Optical Flow Algorithms*. J.N. Carter and M.S. Nixon (Eds.): Proceedings of the British Machine Vision Conference 1998, S. 195–204, Southampton, UK. British Machine Vision Association 1998.
- [Gecks und Henrich 2004] T. Gecks und D. Henrich: *SIMERO: Camera Supervised Workspace for Service Robots*. In ASER 2nd International Workshop on Advances in Service Robotics. Feldafing, Germany, May 2004, 5 Seiten.
- [Gecks und Henrich 2005] T. Gecks und D. Henrich: *Human-Robot Cooperation: Safe Pick-and-Place Operations*. In 14th IEEE International Workshop on Robot and Human Interactive Communication. Nashville, Tennessee, USA, August 2005, 6 Seiten.
- [Gecks und Henrich 2007] T. Gecks und D. Henrich: *Path Planning and Execution in Fast-Changing Environments with Known and Unknown Obstacles*. International Conference on Intelligent Robots and Systems. San Diego, USA, October 2007, 6 Seiten.
- [Gelb 1974] A. Gelb (Ed.): *Applied Optimal Estimation*. MIT Press, Cambridge, MA, 1974.
- [Glazer 1984] F. Glazer: *Multilevel Relaxation in Low-Level Computer Vision*. In A. Rosenfeld (Ed.): *Multiresolution Image Processing and Analysis*. S. 312–330. Springer, Berlin, Heidelberg, 1984.
- [Göger u.a. 2006] D. Göger, K. Weiss, C. Burghart und H. Wörn: *Sensitive Skin for a Human Robot*. Human Centered Robotic Systems (HCRS 2006). München, Deutschland, 2006, S. XX–YY.
- [Golub und Van Loan 1996] G.H. Golub und C.F. Van Loan: *Matrix Computations*. Johns Hopkins University Press, Baltimore and London, 1996.

- [Gonzalez und Woods 2002] R. C. Gonzalez und R. E. Woods: *Digital Image Processing*. 2nd Edition, Prentice Hall Inc., Upper Saddle River, New Jersey, USA, 2002.
- [Graf 2006] J. Graf: *Adaption eines PDG-basierten Optischen-Fluss-Schätzverfahrens an das MOTRIS-System*. Diplomarbeit, Fakultät für Informatik der Universität Karlsruhe (TH), 2006.
- [Graf und Wörn 2007] J. Graf und H. Wörn: *Oriented Median Filtering Preserving Significant Image Structures*. In Proceedings of the 9th International Workshop on Computer Science and Information Technologies (CSIT 2007), Vol. 2. Ufa, Russland, 2007, S. 4–8.
- [Graf und Wörn 2008a] J. Graf und H. Wörn: *Incorporating Optical Flow Estimation into Dynamic Snake Model*. In Proceedings of the 10th International Workshop on Computer Science and Information Technologies (CSIT 2008), Vol. 2. Antalya, Türkei, 2008, S. 1–6.
- [Graf und Wörn 2008b] J. Graf und H. Wörn: *Combined Optical Flow Estimation and Active Contours for Safe Human-Robot Cooperation*. In Proceedings of the 8th Asia-Pacific Conf. on Control and Measurement (APPCM), Harbin, China, 2008, S. 76–82.
- [Graf und Wörn 2008c] J. Graf und H. Wörn: *An Image-Sequence Analysis System with Focus on Human-Robot-Cooperation using PMD-Camera*. In VDI-Berichte 2012, Robotik 2008. VDI/VDE-Gesellschaft Mess- und Automatisierungstechnik. München, Deutschland, 2008, S. 1–8.
- [Graf und Wörn 2009a] J. Graf und H. Wörn: *Safe Human-Robot Interaction using 3D Sensor*. In VDI-Berichte 2067, Automation 2009. VDI/VDE-Gesellschaft Mess- und Automatisierungstechnik. Baden-Baden, Deutschland, 2009, S. 1–12.
- [Graf u.a. 2009b] J. Graf, S. Puls und H. Wörn: *Incorporating Novel Path Planning Method into Cognitive Vision System for Safe Human-Robot Interaction*. In Proc. of IARIA Computation World: Cognitive 2009. Athen, Griechenland, November 2009, S. 443-447.
- [Graf u.a. 2010a] J. Graf, F. Dittrich und H. Wörn: *High Performance Optical Flow Serves Bayesian Filtering for Safe Human-Robot Interaction*. **In Proc. Joint 41th International Symposium on Robotics and 6th German Conference on Robotics. München, Deutschland, Juni 2010, S. 325-332.**
- [Graf u.a. 2010b] J. Graf, P. Czapiewski und H. Wörn: *Evaluating Risk Estimation Methods and Path Planning for Safe Human-Robot-Cooperation*. **In Proc. Joint 41th International Symposium on Robotics and 6th German Conference on Robotics. München, Deutschland, Juni 2010, S. 579-585.**
- [Hackbusch 1994] W. Hackbusch: *Iterative Solution of Large Sparse Systems of Equations*. Applied mathematical sciences 95, Springer, Berlin, Heidelberg, 1994.
- [Hackbusch 2003] W. Hackbusch: *Multi-Grid Methods and Applications*. Springer Series in Computational Mathematics, Second Printing, Springer, Berlin, Heidelberg, 1994.
- [Haque u.a. 2008] M. Haque, M. Murshed, and M. Paul: *Improved Gaussian Mixtures for Robust Object Detection by Adaptive Multi-Background Generation*. In Proc. of 19th Int. Conf. on Pattern Recognition (ICPR), 2008. ICPR 2008, S. 1–4.

- [Hartikainen und Särkkä 2008] J. Hartikainen und S. Särkkä: *Optimal Filtering with Kalman Filters and Smoothers*. Technical Report, University of Technology Helsinki, Februar 2008.
- [Hartley und Zisserman 2004] R. I. Hartley und A. Zisserman: *Multiple View Geometry in Computer Vision*, Second Edition, Cambridge University Press, 2004.
- [Heikkilä u.a. 1999] Janne Heikkilä and Olli Silvén: *A Real-Time System for Monitoring of Cyclists and Pedestrians*. Second IEEE Workshop on Visual Surveillance, S. 74–81, Fort Collins, Colorado. 26 June 1999.
- [Heiligensetzer 2003] P. Heiligensetzer: *Sichere Mensch-Roboter Kooperation durch Fusion haptischer und kapazitiver Sensorik*, Dissertation, Fakultät für Informatik der Universität Karlsruhe (TH), März 2003.
- [Heinol 2001] H. G. Heinol: *Untersuchung und Entwicklung von modulationslaufzeitbasierten 3D-Sichtsystemen*. Dissertation, Fakultät für Elektrotechnik und Informatik der Universität Siegen, Januar, 2001.
- [Henrich u.a. 2004] D. Henrich, T. Gecks und D. Ebert: *Bildbasierte Kollisionstests für Randomized-Roadmap-Bahnplaner*. In Robotik 2004. München, Germany, 2004, 8 Seiten.
- [Henrich und Kuhn 2006] D. Henrich und S. Kuhn: *Modeling Intuitive Behavior for Safe Human-Robot Coexistence and Cooperation*. In IEEE International Conference on Robotics and Automation, May 2006. Orlando, Florida, USA, S. 3929–3934.
- [Henrich und Gecks 2008a] D. Henrich und T. Gecks: *Multi-Camera Collision Detection Between Known and Unknown Objects*. In 2nd ACM/IEEE International Conference on Distributed Smart Cameras - ICDS-C 2008. Stanford, USA, September 2008, S. XX–YY.
- [Henrich u.a. 2008b] D. Henrich, M. Fischer und T. Gecks: *Sichere Mensch/Roboter-Koexistenz und Kooperation*. In Robotik 2008. München, Germany, 2008, 10 Seiten.
- [Horn und Schunck 1981] B.K.P. Horn und B.G. Schunck: *Determining Optical Flow*. Artificial Intelligence 17:1-3 (1981) 185–203
- [Horn 1986] B. K. P. Horn: *Robot Vision*, The MIT Press, Cambridge, Massachusetts, London, England, 1986.
- [Horn 1987] B.K.P. Horn: *Closed-Form Solution of Absolute Orientation using Unit Quaternions*. Optical Society of America Journal A, vol. 4, S. 629-642, April 1987.
- [Huang 1981] T.S. Huang: *Two-Dimensional Signal Processing II: Transforms and Median Filters*. Springer-Verlag, Berlin, 1981. S. 209–211
- [Isard und Blake 1998] M. Isard und A. Blake: *Condensation-Conditional Density Propagation for Visual Tracking*. In International Journal of Computer Vision 29 (1998), Nr 1, S. 5–28.
- [Jähne 2006] B. Jähne: *Digitale Bildverarbeitung*, 6. Auflage, Springer-Verlag, Berlin, Heidelberg, 2006.
- [Jähne 1993] B. Jähne: *Spatio-Temporal Image Processing - Theory and Scientific Applications*. Springer-Verlag, Berlin, Heidelberg, 1993.

- [Justen 2001] D. Justen: *Untersuchung eines neuartigen 2D- gestützten 3D-PMD Bildverarbeitungssystems*. Dissertation, Fakultät für Elektrotechnik und Informatik der Universität Siegen, Januar, 2001.
- [Kaewtrakulpong und Bowden 2001] P. Kaewtrakulpong, R. Bowden: *An Improved Adaptive Background Mixture Model for Realtime Tracking with Shadow Detection*. Proceedings of the Second European Workshop on Advanced Video Based Surveillance Systems, 2001.
- [Kalmoun und Rüde 2003] M. Kalmoun and R. Rüde: *A Variational Multigrid for Computing the Optical Flow*. In T. Ertl, B. Girod, H.-P. Seidel, E. Steinbach and R. Westermann (Eds.): *Vision, Modelling and Visualization*, S. 577–584, Genf, Schweiz, 2003.
- [Kass u.a. 1988] M. Kass, A. Witkin und D. Terzopolous: *Snakes: Active Contour Models*. In *International Journal of Computer Vision*, 1988, 1(4):321–331.
- [Kavraki und Latombe 1998] L. E. Kavraki and J.-C. Latombe: *Probabilistic Roadmaps for Robot Path Planning*. In *Practical Motion Planning in Robotics: Current Approaches and Future Directions*. 1998, S. 33–53.
- [Kiendl 1997] H. Kiendl: *Fuzzy Control methodenorientiert*. R. Oldenbourg Verlag, München, Deutschland, 1997.
- [Knoop u.a. 2006] S. Knoop, S. Vacek and R. Dillmann: *Sensor Fusion for 3D Human body Tracking with an Articulated 3D Body Model*. Proceedings of the 2006 IEEE International Conference on Robotics and Automation, ICRA 2006, May 15-19, 2006, Orlando, Florida, USA.
- [Krüger u.a. 2005] J. Krüger, B. Nickolay, P. Heye: *Image based 3D Surveillance for flexible Man-Robot-Cooperation*. In *CIRP Annals - Manufacturing Technology*, vol. 54:1, 2005. S. 19-22.
- [Kuhn u.a. 2006] S. Kuhn, T. Gecks und D. Henrich: *Velocity Control for Safe Robot Guidance based on fused Vision and Force/Torque Data*. In *IEEE Conference on Multisensor Fusion and Integration*. Heidelberg, Germany, September 2006, 8 Seiten.
- [Kuhn u.a. 2007] S. Kuhn und D. Henrich: *Fast Vision-Based Minimum Distance Determination Between Known and Unknown Objects*. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*. San Diego, USA, October 2007, 6 Seiten.
- [Kulić 2005] D. Kulić: *Safety for Human-Robot Interaction*. Dissertation, Faculty of Graduate Studies, Mechanical Engineering, University of British Columbia, December 2005.
- [Kulić und Croft 2006] D. Kulić und E. A. Croft: *Real-Time Safety for Human-Robot Interaction*. *Robotic and Autonomous Systems* 54, 2006, S. 1–12.
- [Kulić und Croft 2007] D. Kulić und E. A. Croft: *Affective State Estimation for Human-Robot Interaction*. *IEEE Transactions on Robotics*, Vol. 54(5), 2007, S. 991–1000.
- [Lam u.a. 1994] L. Lam, S. W. Lee und C.Y. Suen: *Thinning Methodologies – A Comprehensive Survey*. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 14(9), September, 1994, S. 869–885.

- [Lange u.a. 1999] R. Lange, P. Seitz und R. Schwarte: *Time-Of-Flight Entfernungskamera in CMOS/CCD-Technik mit pixelintegrierten Lock-In Verstärkern*. In VDI/VDE Berichtsband 70 zur Fachtagung Optische Formerfassung, Stuttgart, Deutschland, Oktober 1999, S. 271-280.
- [Lange 2000] R. Lange: *3D Time-of-Flight Distance Measurement with Custom Solid-State Image Sensors in CMOS/CCD-Technology*. Dissertation, Fakultät für Elektrotechnik und Informatik der Universität Siegen, Juni, 2008.
- [Latombe 1991] J. C. Latombe: *Robot Motion Planning*. Kluwer Academic Publishers, Norwell, MA, USA, 1991.
- [Lee 2005] D.-S. Lee: *Effective Gaussian Mixture Learning for Video Background Subtraction*. In IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 27, May, 2005, S. 827–832.
- [Löschet u.a. 2009] M. Lösch, S. Gärtner, S. Knoop, S.R. Schmidt–Rohr und R. Dillmann: *A human body model initialization approach made real-time capable through heuristic constraints*. In Proc. 14th International Conference on Advanced Robotics (ICAR 2009), Munich, Germany, 6 Seiten.
- [Longuet–Higgins 1981] H. Longuet-Higgins: *A Computer Algorithm for Reconstructing a Scene From Two Projections*. Nature, No. 293, 1981, pp. 133-135.
- [Lucas und Kanade 1981] B.D. Lucas and T. Kanade: *An Iterative Image Registration Technique With an Application to Stereo Vision*. Proc. of the 7th International Joint Conference on Artificial Intelligence, pp. 674-679, Vancouver, Canada, August 1981.
- [Ma u.a. 2006] Y. Ma, S. Soatto, J. Košecá und S. S. Sastry: *An Invitation to 3-D Vision*, Springer, 2006.
- [Mamaev u.a. 2008] I. Mamaev, K. Safronov, I. Tchouchenkov und H. Wörn: *Real-time Framework for Bin-Picking Problem using Advantages of PMD-Technology*. In Proceedings of the 10th International Workshop on Computer Science and Information Technologies (CSIT 2008), Vol. 2. Antalya, Türkei, 2008, 6 Seiten.
- [Mrázek et al. 2006] P. Mrázek, J. Weickert und A. Bruhn: *On Robust Estimation and Smoothing with Spatial and Tonal Kernels*. In R. Klette, R. Kozera, L. Noakes und J. Weickert (Herausgeber), Geometric Properties from Incomplete Data, S. 335–352, Springer, Dordrecht, 2006.
- [McKenna u.a. 1998] S. J. McKenna, Y. Raja, S. Gong: *Object Tracking using Adaptive Colour Mixture Models* Third Asian Conference on Computer Vision, vol. 1, 1998.
- [Mémin und Perez 1998] E. Mémin and P. Perez: *Hierarchical Estimation and Segmentation of Dense Motion Fields*. International Journal of Computer Vision, 46(2):129–159, 2002.
- [Meyberg und Vachenaer 1999] K. Meyberg und P. Vachenaer: *Höhere Mathematik 2*. 2. Auflage, Springer, Berlin, Heidelberg, New York, 1999.

- [Michoud u.a. 2007] B. Michoud, E. Guillou, and S. Bouakaz: *Real-Time and Markerless 3D Human Motion Capture Using Multiple Views*. In Human Motion - Understanding, Modeling, Capture and Animation, 2007. Lecture Notes in Computer Science, Vol. 4814 , Springer Berlin, Heidelberg, S. 88–103.
- [Middendorf und Nagel 2002] M. Middendorf und H.–H. Nagel: *Empirically Convergent Adaptive Estimation of Grayvalue Structure Tensors*. In L. Van Gool (Ed.), Pattern Recognition, volume 2449 of Lecture Notes in Computer Science, pp. 66–74. Springer, Berlin, 2002.
- [Nagel und Gehrke 1998] H.–H. Nagel und A. Gehrke: *Spatiotemporally Adaptive Estimation and Segmentation of OF-Fields*. In H. Burkhardt und B. Neumann (Eds.): Computer Vision - ECCV'98, Springer LNCS 1407, S. 86–102, 1998.
- [Nickel 2008] K. Nickel: *Visuelle Benutzermodellierung mit Tracking und Zeigegestenerkennung für einen humanoiden Roboter*. Dissertation, Fakultät für Informatik der Universität Karlsruhe (TH), November 2008.
- [O3D-PMD] www.ifm-electronic.com Letzter Besuch: Februar, 2010.
- [Perona und Malik 1990] P. Perona und J. Malik: *Scale-Space and Edge Detection Using Anisotropic Diffusion*. IEEE Trans. Pattern Analysis and Machine Intelligence 12:7 (1990), S. 629–639.
- [Pilz 2007] <http://www.pilz.de/products/sensors/camera/f/safetyeye/index.de.jsp> Letzter Besuch: 10.08.2009.
- [Potmesil 1986] M. Potmesil: *Generating Octree Models of 3D objects from their silhouettes in a Sequence of Images*. Computer Vision and Image Processing 40, 1991. S. 1–20.
- [Power und Schoones 2002] P. Wayne Power und Johann A. Schoones: *Understanding Background Mixture Models for Foreground Segmentation*. Proceedings Image and Vision Computing New Zealand, 2001. 6 Seiten.
- [Prassler et al. 2005] E. Prassler et al. (Eds.): *Advances in Human-Robot Interaction*. Springer Tracts in Advanced Robotics 14, Springer-Verlag, Berlin, Heidelberg, 2005.
- [Rosenfeld und Thurston 1971] A. Rosenfeld und M. Thurston: *Edge and Curve Detection for Visual Scene Analysis*. In IEEE Transactions on Computation, vol. C-20, S. 562–569, Mai 1971.
- [Russell und Norvig 2003] S. J. Russell und P. Norvig: *Artificial Intelligence: A Modern Approach*. Prentice Hall, Upper Saddle River, N. J., 2003.
- [Rymel u.a. 2004] J. Rymel, J. Renno, D. Greenhill, J. Orwell und G.A. Jones: *Adaptive Eigen-Backgrounds for Object Detection*. 2004 International Conference, pp.1847-1850, Vol.3, Oktober 2004, S. 24–27.
- [Särkkä 2007] S. Särkkä: *Unscented Rauch-Tung-Striebel Smoother*. In IEEE Trans. on Automatic Control, 53(3), April 2008. S. 845–849.

- [Schneider 2003] B. Schneider: *Der Photomischdetektor zur schnellen 3D-Vermessung für Sicherheitssysteme und zur Informationsübertragung im Automobil*. Dissertation, Fakultät für Elektrotechnik und Informatik der Universität Siegen, Juli 2003.
- [Schreer 2005] O. Schreer: *Stereoanalyse und Bildsynthese*. Springer-Verlag, Heidelberg, New York, Tokyo, 2005.
- [Schwarz & Köckler 2009] H.R. Schwarz and N. Köckler: *Numerische Mathematik*. Teubner Verlag, 7. Auflage, Wiesbaden, 2009.
- [Siegert 1999] H. J. Siegert: *Weg- und Trajektorienplanung*. Skriptum zur Vorlesung, AG Echtzeitsysteme und Robotik, TU München, 1999.
- [Stauffer und 1999] C. Stauffer and W. E. L. Grimson: *Adaptive Background Mixture Models for Real-Time Tracking*. Computer Vision and Pattern Recognition, Fort Collins, Colorado, Juni 1999, S. 246–252.
- [Stauffer 2002] C. P. Stauffer: *Perceptual Data Mining: Bootstrapping Visual Intelligence from Tracking Behavior*. Dissertation, Departement of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Mai, 2002.
- [Stelling 1999] D. Stelling: *Teamarbeit in Mensch-Maschine-Systemen*. Hogrefe-Verlag, Reihe: Lehr- und Forschungstexte Psychologie, Band 8, Göttingen, 1999.
- [Som 2003] F. Som: *Sichere Steuerungstechnik für den OTS-Einsatz von Robotern*. 2. Workshop für OTS-Systeme in der Robotik, IPA 2003.
- [Som 2005] F. Som: *Sichere Steuerungstechnik für den OTS-Einsatz von Robotern*. 4. Workshop für OTS-Systeme in der Robotik, IPA 2005.
- [Som 2007] F. Som: *Robotersteuerung für sichere Mensch-Roboter-Kooperation*. Visionen im Arbeitsschutz, Fachtagung, Berufsgenossenschaft Metall Nord Süd (BGM), 2007.
- [Som 2008] F. Som: *Robotersteuerung für sichere Mensch-Roboter-Kooperation*. Visionen im Arbeitsschutz, Fachtagung, Berufsgenossenschaft Metall Nord Süd (BGM), 2008.
- [Spirig u.a. 1995] T. Spirig, P. Seitz, O. Vietze und F. Heitger: *The Lock-In CCD – Two-Dimensional Synchronous Detection of Light*. In IEEE Journal of Quantum Electronics, 31(9), S. 1705–1708, 1995.
- [Spirig und Seitz 1996] T. Spirig und P. Seitz: *Vorrichtung und Verfahren zur Detektion und Demodulation eines intensitätsmodulierten Strahlungsfeldes*. Deutsches Patent No. DE 44 40 613 C 1, 1996.
- [Spirig u.a. 1997] T. Spirig, M. Marley und P. Seitz: *The Multitap Lock-In CCD with Offset Subtraction*. In IEEE Journal of Quantum Electronics, 44(10), S. 1643–1647, 1997.
- [Terzopoulos 1986] D. Terzopoulos: *Image analysis using multigrid relaxation*. IEEE Transaction on Pattern Analysis and Machine Intelligence, 8(2), 1986. S. 129–139.
- [Thiemermann 2005] S. Thiemermann: *Direkte Mensch-Roboter-Kooperation in der Kleinteilemontage mit einem SCARA-Roboter*. Dissertation, Fakultät für Maschinenbau der Universität Stuttgart, Februar 2005.

- [Thrun u.a. 2006] S. Thrun, W. Burgard und D. Fox: *Probabilistic Robotics*. The MIT Press, Cambridge, MA, USA, 2006.
- [Tomasi und Manduchi 1998] C. Tomasi und R. Manduchi: *Bilateral Filtering for Gray and Color Images*. In Proceedings of the 6th International Conference on Computer Vision, S. 839–846, Bombay, Indien, 1998.
- [Turau 1996] V. Turau: *Algorithmische Graphentheorie*. Addison–Wesley, Bonn, Deutschland, 1996.
- [Tuytelaars und Mikolajczyk 2007] T. Tuytelaars und K. Mikolajczyk: *Local Invariant Feature Detectors: A Survey*. *Foundation and Trends in Computer Graphics and Vision* 3:3, 2007, S. 177–280.
- [Valgerats u.a. 2008] L. Valgerats, A. Bruhn and J. Weickert: *A Variational Approach for the Joint Recovery of the Fundamental Matrix and the Optical Flow*. In Pattern Recognition, Proc. 30th DAGM Symposium DAGM 2008, Munich, Germany, June 2008 - G.Rigoli (Ed.) *Lecture Notes in Computer Science*, Vol. 5096, Springer, Berlin, S. 314–324.
- [Wachter und Nagel 1999] S. Wachter and H.-H. Nagel: *Tracking Persons in Monocular Image Sequences*. In *Computer Vision and Image Understanding* 74(3) (1999), S. 174–192.
- [Weickert 1998] J. Weickert: *Anisotropic Diffusion in Image Processing*. ECMI series, Teubner-Verlag, Stuttgart, Deutschland, 1998.
- [Weickert 1999] J. Weickert: *Coherence Enhancing Diffusion Filtering*. In *International Journal for Computer Vision* 31(2/3) (1999), S. 111–127.
- [Weiss 2006] B. Weiss: *Fast Median and Bilateral Filtering*. In Proc. of ACM SIGGRAPH 2006, vol. 25(3), 2006, S. 519–526.
- [Wesseling 2004] P. Wesseling: *An Introduction to Multigrid Methods*. John Wiley & Sons Ltd., 1992. Corrected Reprint. Philadelphia: R.T. Edwards, Inc., 2004.
- [Winkler 2006a] A. Winkler: *Ein Beitrag zur kraftbasierten Mensch–Roboter–Interaktion*. Dissertation, Fakultät für Elektrotechnik und Informationstechnik der Technischen Universität Chemnitz, September 2006.
- [Winkler 2006b] G. Winkler: *Image Analysis, Random Fields and Markov Chain Monte Carlo Methods*. Springer-Verlag, 2. Auflage, Berlin und Heidelberg, Deutschland, 2006.
- [Winkler 2008] B. Winkler: *Konzept zur Sicheren Mensch–Roboter–Kooperation auf Basis von Schnellen 3-D Time-of-Flight Sensoren*. A. Verl und M. Hägele (Eds.), VDI/VDE-Gesellschaft für Meß- und Automatisierungstechnik (GMA), Düsseldorf, Deutsche Gesellschaft für Robotik: Robotik 2008, S. 147–151. Tagung Münschen, Juni 2008.
- [Wiktin 1983] A. Witkin: *Scale Space Filtering*. In International Joint Conference on Artificial Intelligence, Karlsruhe, West-Deutschland, 1983, S. 1019–1021.
- [Wörn u.a. 2009] H. Wörn, J. Graf, K. Safronov und I. Mamaev: *3D–Umwelterfassung in der Produktion mit neuer 3D–Kamera*. In *Allgemeine Vermessungs–Nachrichten, Zeitschrift für Geodäsie und Geoinformation*, Wichmann Verlag, Mai, 2009, S. 200–206.

- [Wren u.a. 1999] C. Wren, A. Azabayejani, T. Darrell and A. Pentland: *Pfinder: Real-Time Tracking of the Human Body*. IEEE Transactions on Pattern Analysis and Machine Intelligence, S. 780-785, 1997.
- [Xu und Prince 2000] C. Xu und J. L. Prince: *Gradient Vector Flow Deformable Models*. In I. Bankman (Ed.), Handbook of Medical Imaging, Academic Press, London, 2000, S. 159–170.
- [Yigit 2005] S. Yigit: *Methoden zur taktil gekoppelten Kooperation zwischen Mensch und anthropomorphen Roboterarm*, Dissertation, Fakultät für Informatik der Universität Karlsruhe (TH), März 2005.
- [Young 2003] D.M. Young: *Iterative Solution of Large Linear Systems*. Dover Publication Inc., 2003. 31 East 2nd Street, Mineola, N.Y. 11501, USA.
- [Zhou 2009] N. Zhou: *Bewegungsschätzung für die sichere Mensch–Roboter–Kooperation auf der Basis von Bayesscher Filterung*. Diplomarbeit, Fakultät für Informatik der Universität Karlsruhe (TH), April 2009.
- [Zini u.a. 1997] G. Zini, A. Sarti and C. Lamberti: *Application of Continuum Theory and Multi-Grid Methods to Motion Evaluation From 3D echocardiography*. In IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control, 44(2): 1997 297–308.