

**Entwicklung eines Softwaremoduls zur  
ansichtenbasierten Erkennung und Verfolgung  
hochdynamischer Objekte für einen  
anthropomorphen Roboterassistenten**

**Diplomarbeit**

**Bernadus Ingo Vetter**

**Fakultät für Physik  
Ruhr-Universität Bochum**

**durchgeführt am  
Institut für Neuroinformatik  
Ruhr-Universität Bochum**

**Juni 2002**



Hiermit versichere ich, dass die vorgelegte Arbeit von mir selbstständig verfasst wurde und ich keine anderen als die angegebenen Quellen und Hilfsmittel verwendet habe.



meiner Tochter gewidmet



# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>5</b>
1.1	Problemstellung . . . . .	6
1.2	Gliederung . . . . .	7
<b>2</b>	<b>Tracking und Positionsbestimmung</b>	<b>9</b>
2.1	Differenzbilder und Rauschunterdrückung . . . . .	10
2.2	Clustern . . . . .	11
2.3	Kantenfilter . . . . .	14
2.3.1	Eindimensionaler Differenzoperator . . . . .	15
2.3.2	Prewitt-Operator . . . . .	16
2.3.3	Canny-Operator . . . . .	17
2.4	Das Stereosystem . . . . .	19
2.4.1	Korrespondenzproblem . . . . .	20
2.4.2	Korrespondenzsuche . . . . .	24
2.4.3	Rekonstruktion der dreidimensionalen Szene . . . . .	27
2.5	Objektgeschwindigkeit . . . . .	29
2.5.1	Optischer Fluss . . . . .	29
2.5.2	Objektmatching . . . . .	31
2.5.3	Hausdorffdistanz . . . . .	32
2.5.4	Bestimmung der Geschwindigkeit mittels Hausdorffdi- stanz . . . . .	34
<b>3</b>	<b>Positionsprädiktion</b>	<b>35</b>
3.1	Prädiktion der Objektposition . . . . .	35
3.1.1	Bestimmung der Objektposition . . . . .	36
3.2	Bestimmung einer Greifposition . . . . .	38
<b>4</b>	<b>CoRA</b>	<b>43</b>
4.1	Autonomes Assistenzsystem CoRA . . . . .	43
4.2	Bildaufnahme . . . . .	45

<b>5</b>	<b>Ergebnisse</b>	<b>47</b>
5.1	Erste Beispielsequenz . . . . .	47
5.2	Zweite Beispielsequenz . . . . .	55
5.3	Dritte Beispielsequenz . . . . .	58
5.4	Vierte Beispielsequenz . . . . .	59
5.5	Fünfte Beispielsequenz . . . . .	62
<b>6</b>	<b>Zusammenfassung und Ausblick</b>	<b>69</b>
6.1	Allgemeine Bewertung der Ergebnisse . . . . .	69
6.2	Mögliche Verbesserungen und Weiterentwicklung . . . . .	70
<b>A</b>	<b>Tabellen</b>	<b>73</b>
<b>B</b>	<b>Sequenzen</b>	<b>79</b>
<b>C</b>	<b>Technische Daten</b>	<b>85</b>

# Kapitel 1

## Einleitung

Die visuelle Verfolgung von Gegenständen spielt sich bei Mensch und Tier im Alltagsleben weitgehend ohne Schwierigkeiten ab. Ein Säugling kann Gegenstände mit seinen Augen verfolgen, bevor er lernt zu greifen. Die dabei vom Gehirn geleistete Arbeit wird leicht unterschätzt, da gerade beim Sehen größere Datenmengen vom Gehirn verarbeitet werden müssen. Die Gegenstandsbewegung muss wahrgenommen werden, das Gehirn muss entscheiden, ob sich der Beobachter oder der Gegenstand bewegt (Beispiel: eine sich im anfahrenden Zug befindliche Person hat Schwierigkeiten, zu beurteilen, ob sich der eigene Zug bewegt oder das Gleis oder vielleicht ein anderer benachbarter Zug), das Gehirn muss eventuell abschätzen, wohin sich ein Gegenstand bewegen wird. Soll gleichzeitig der Gegenstand aus seiner Bewegung gegriffen werden, erschwert das die Aufgabe für das Gehirn. Zusätzlich zur Gegenstandsverfolgung muss das Gehirn parallel dazu den Bewegungsablauf für den Körper oder einzelne Körperteile koordinieren und planen.

Diese Fähigkeit des Gehirns, verschiedene Aufgaben und Daten hochgradig parallel zu verarbeiten, sind den Computern in dieser Form und Leistung noch verwehrt. Zwar sind heutige Computersysteme in der Lage, Daten parallel zu verarbeiten, diese erreichen momentan aber bei weitem noch nicht die Leistung des menschlichen Gehirns.

Aus der augenblicklichen Entwicklung der Computerhardware besteht aber die Aussicht, dass die Leistung der Computersysteme derart zunimmt, um Aufgaben dieser Art mit Hilfe autonomer Systeme in Zukunft problemlos lösen zu können. Vorausgegangene Arbeiten auf dem Gebiet der Objektverfolgung und dem Greifen aus der Bewegung zeigen große Fortschritte (vgl. [1], [2], [3] und [4]) und geben den Antrieb, neue Verfahren zu entwickeln, die die Leistungen und Ressourcen heutiger Computer optimal ausnutzen.

## 1.1 Problemstellung

Am Institut für Neuroinformatik an der Ruhr-Universität Bochum wird im Rahmen des vom *bmb+f* geförderten Projektes „Morpha“ das autonome Robotersystem CoRA (Cooperative Robot Assistant) entwickelt. Der momentane Entwicklungsstand des Roboters ist in Abschnitt 4 dargestellt. Unter anderem ist CoRA bereits in der Lage, gelernte stationäre Gegenstände wiederzuerkennen und diese anschließend zu greifen. Der geplante Einsatzort des Roboters ist an einem Transportband eines Montageassistenzszenarios. Der Roboter soll von dem Transportband Gegenstände greifen, um sie danach in eine eventuelle Montageposition zu bringen.

Die Analyse des dynamischen Ablaufs dieser Szene bildet den Schwerpunkt dieser Arbeit. Die Position des Gegenstandes soll schnell geschätzt und prädiziert werden. Anschließend erfolgt die Bestimmung der Griffposition. Schließlich soll der Gegenstand von einem 7-Freiheitsgrade-Arm aus der Bewegung gegriffen werden. Die Geometrie des Gegenstandes muss hierbei so beschaffen sein, dass ein Greifen möglich ist.

Die Positionsschätzung erfolgt aus der Tiefenschätzung eines Stereoalgorithmus. Die Objektgeometrie fließt in die Positionsschätzung mit ein. Auf diese Weise ist es möglich, Fehler bei der Tiefenschätzung zu kompensieren.

Die Positionsprädiktion basiert auf der Hausdorffdistanz. Die Hausdorffdistanz erlaubt eine relativ schnelle und zuverlässige dreidimensionale Analyse der Objektverschiebung. Darauf aufbauend werden die zukünftigen Positionen des Gegenstandes prädiziert.

Aus diesen Informationen lässt sich ein minimaler Abstand zwischen Greifer und Gegenstand schätzen. Das ungerichtete Greifen kann dann eingeleitet werden, wenn die Orientierung des Gegenstands im Raum bekannt ist.

Weitere denkbare Einsatzgebiete von Objektverfolgungssystemen:

- In Zeiten, in denen chirurgische Eingriffe immer größere Präzision verlangen, können Objektverfolgungssysteme dieser Art eine unterstützende oder überwachende Funktion übernehmen.
- In sportlichen und sportlich medizinischen Bereichen könnten Objektverfolgungssysteme eingesetzt werden, um Bewegungsabläufe zu verfolgen oder um Ganganalysen zu erstellen. Diese könnten dazu verhelfen, beispielsweise für einen Sportler die richtige Wahl eines Sportschuhs zu erleichtern. Denkbar ist auch der Einsatz von Ganganalysen, um Trickfiguren in Trickfilmen natürlicher erscheinen zu lassen.

- Im Bereich der Mensch-Maschine-Kommunikation sind Objektverfolgungssysteme eine mögliche Hilfe für die Interpretation etwa der Gebärden (vgl. [5]) und können so eine Kommunikation zwischen Gehörlosen und nicht Gebärden Sprechenden erleichtern.

## 1.2 Gliederung

In Kapitel 2 werden Verfahren für das Verfolgen von Objekten und die Schätzung von Objektpositionen vorgestellt. Das Kapitel 3 befasst sich mit der Prädiktion von Objektpositionen. In Kapitel 4 wird der Roboter CoRA vorgestellt. Kapitel 5 geht auf die verwendeten Verfahren zur Objektverfolgung und -prädiktion genauer ein. Daneben werden in diesem Kapitel Ergebnisse für die Objektprädiktion präsentiert und diskutiert. Das Kapitel 6 umfasst schließlich eine Zusammenfassung der Ergebnisse und einen Ausblick mit möglichen weiteren Schritten und denkbaren Optimierungen.



# Kapitel 2

## Tracking und Positionsbestimmung

Das Verfolgen eines Objekts in einem Video oder Film bezeichnet man als Tracken. Die hierbei zu lösenden Aufgaben umfassen generell die folgenden Punkte:

- die Registrierung des zu verfolgenden Objektes bzw. die Segmentierung des Objekts vom Hintergrund (leitet das Tracken ein),
- das Wiedererkennen des zu verfolgenden Objektes in nachfolgenden Bildern und
- das Unterscheiden des zu verfolgenden Objektes von anderen Objekten im Bildbereich.

Das Detektieren von Bewegung lässt sich erreichen, indem die Veränderungen eines Bildes zu einem zeitlich später aufgenommenen Bild analysiert werden. Durch bewegte Kameras ist diese Aufgabe erschwert, da dann im gesamten Bild eine Veränderung zu erkennen ist. Bei der Analyse der Bildveränderung erzeugt das bewegte Objekt aber ein anderes Bewegungsmuster (Flussfeld) im Bild, als das durch die Eigenbewegung der Kamera entstandene Bewegungsmuster. Somit lassen sich zu verfolgende Objekte beispielsweise über den optischen Fluss segmentieren.

Um ein einzelnes, einmal segmentiertes Objekt zu tracken, muss es in Folgebildern wiedergefunden werden. Diese Aufgabe lässt sich durch Matchingverfahren lösen, bei denen Teile von Bildflächen miteinander verglichen werden (siehe Kapitel 2.5.2).

In dieser Arbeit sollen die Kameras ihre Position und Stellung im Raum beibehalten. Zu Beginn jeder Szene befindet sich nur das zu trackende Objekt

in Bewegung. Auf diese Weise kann eine Detektion von Objektbewegung durch die Bildung von Differenzbildern erfolgen (siehe Kapitel 2.1). Durch die raumfesten Kameras entfällt die Bildveränderung, die aus der Eigenbewegung der Kameras resultiert.

Da sich in dieser Arbeit nur das zu trackende Objekt in Bewegung befindet, ist eine Unterscheidung von anderen sich bewegenden Objekten nicht nötig. Daher soll hier nur erwähnt werden, dass sich das Unterscheiden von mehreren sich bewegenden Objekten zum Beispiel durch die Bildung von Flussfeldern erreichen ließe (siehe Kapitel 2.5.1) oder durch das Segmentieren eines Bildes (einzelne Segmente sind zum Beispiel durch ihre Grenzen oder ihre Flächen definiert) (siehe [6]).

Als vorverarbeitende Schritte für eine Bildanalyse, ist es in dieser Arbeit sinnvoll, folgende Maßnahmen durchzuführen:

- das Clustern eines Differenzbildes, um Bildstörungen zu eliminieren (Kapitel 2.1),
- eine Kantenextraktion, um Objektmerkmale zu extrahieren und das Tracken leistungsfähiger zu gestalten (Kapitel 2.3),

Aus diesen Informationen lässt sich die Geschwindigkeit des Objekts bestimmen (siehe Kapitel 2.5), zumal aus der 3D-Ansicht und den Disparitäten die Entfernung des Objekts zur Kamera für jedes Bild ermittelt werden kann. Der Ort eines sich bewegenden Objekts zu einem späteren Zeitpunkt lässt sich demnach mit Hilfe des Trackings vorhersagen, wenn die Startposition des Objekts bekannt ist.

## 2.1 Differenzbilder und Rauschunterdrückung

Ein Differenzbild ergibt sich aus der pixelweisen Differenz zweier Bilder. Differenzbilder setzt man beispielsweise ein, um Veränderungen zu detektieren, die auf eine Bewegung schließen lassen. Dazu erfolgt ein Vergleich von zwei zeitlich nacheinander aufgenommenen Bildern. Das Ergebnis ist ein Differenzbild, das die Bereiche im Bild kennzeichnet, in denen Veränderungen im Bild oder in der Umwelt, z.B. Bewegungen, Helligkeitsveränderungen oder Rauschen, stattgefunden haben. Durch die pixelweise Differenz der Grauwerte von zwei zeitlich unterschiedlich aufgenommenen Bildern  $f(x, y, t_1)$  und  $f(x, y, t_2)$  erhält man ein Differenzbild  $D(x, y, t_2, t_1)$ :

$$D(x, y, t_2, t_1) = \begin{cases} 0 & : f(x, y, t_2) - f(x, y, t_1) < \text{Schwelle} \\ 1 & : f(x, y, t_2) - f(x, y, t_1) \geq \text{Schwelle} \end{cases} .$$

Zusammenhängende Bereiche im Differenzbild, die aus weniger als zehn Pixeln bestehen, werden als Störung interpretiert. Dieses Rauschen ist nicht erwünscht und wird aus dem Differenzbild entfernt. Das kann etwa durch eine Bildglättung mittels Gaußfunktion geschehen oder über eine Bildsegmentierung, die in Abschnitt 2.2 vorgestellt wird.

## 2.2 Clustern

Da das Tracken in Echtzeit geschehen soll, muss die zu verarbeitende Datenmenge auf ein Minimum reduziert werden. Ein Verfahren hierfür ist die Bildung einer *Region of Interest* (ROI). Diese ROI ergibt sich aus dem Differenzbild. Die ROI ist ein ausgewählter Bereich eines Bildes, in dem die interessierenden Ereignisse für die jeweilige Aufgabe stattfinden. Dadurch reduziert sich der zu verarbeitende Bildbereich auf die Region of Interest.

Wie bereits erwähnt, werden durch das Differenzbild auch unerwünschte Störpixel hervorgehoben. Eine geeignete Bestimmung einer ROI wird durch diese Störpixel verhindert, da in dieser Arbeit nur die Bereiche eines Bildes relevant sind, in denen die Objektbewegung stattgefunden hat. Vereinte Bereiche, die aus weniger als 10 Pixeln bestehen, werden als Bildstörungen aufgefasst. Um diese Störpixel zu entfernen, untersucht man die Größe von sogenannten Clustern im Differenzbild. Cluster sind Anhäufungen oder Gruppen von Pixeln, die zusammenhängend sind. Das Clustern ist ein Prozess, bei dem einzelne Pixel einer bestimmten Gruppe von Pixeln zugewiesen werden. Cluster, die aus weniger als zehn Pixeln bestehen, werden aus dem Differenzbild entfernt.

Aus dem bearbeiteten Differenzbild lässt sich die Region of Interest (ROI) bilden, innerhalb der eine Kantenextraktion des Objekts erfolgt. In dem störungsfreien Differenzbild wird nach dem ersten und dem letzten horizontalen Pixel und nach dem ersten und dem letzten vertikalen Pixel gesucht. Aus diesen vier Bildkoordinaten lässt sich die ROI bilden. Der hier benutzte Clusteralgorithmus basiert auf einem in [7] beschriebenen Verfahren.

### Beschreibung des verwendeten Clusterverfahrens

Das Differenzbild wird zeilenweise abgetastet, beginnend im Bild oben links. Ist ein gesetztes Pixel gefunden, bekommt das gesetzte Pixel einen neuen Wert zugewiesen. Dieser Wert ist die Stelle, an der sich das Pixel im Bild befindet (siehe Abb. 2.1). Da das zeitweise Trennen verbundener Objekte hier unkritisch ist, genügt es, bei jedem gesetztem Pixel vier Nachbarpixel zu betrachten, um ein Cluster zu bilden. Dieses sind das linke Pixel und die Pi-

xel links oberhalb, oberhalb und rechts oberhalb des momentan untersuchten Pixels. Gesetzte Pixel und damit bereits verbundene gesetzte Pixel erhalten ebenfalls den Wert, der dem momentan untersuchten Pixel zugewiesen wurde (siehe Abb. 2.2). Nach der Abtastung des gesamten Bildes besitzt jedes vereinte Gebiet eine gemeinsame kanonische Objektnummer.

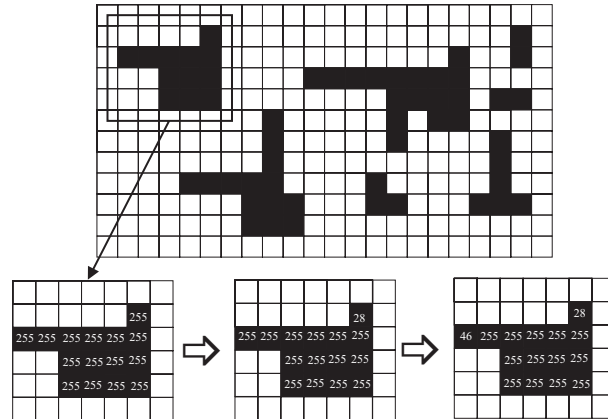


Abbildung 2.1: **Unteres mittleres Bild:** Betrachtet wird das oberste rechte Pixel: Dieses Pixel ist gesetzt (es hat den Wert „255“) und erhält den Wert „28“, da es sich an der 28. Stelle im Bild befindet. Die Pixel in der Vierernachbarschaft (siehe Text) sind ungesetzt und bleiben daher unverändert. **Unteres rechtes Bild:** Das Pixel an der 46. Bildstelle ganz links ist ebenfalls gesetzt und erhält daher den Wert „46“.

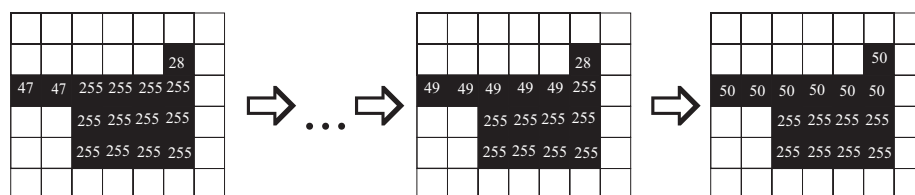


Abbildung 2.2: Umnummerierung der bereits geänderten Pixelwerte, wenn sie Nachbarn des untersuchten Pixels sind: **Linkes Bild:** Das 47. Pixel ist gesetzt. Da sich das 46. Pixel in der Vierernachbarschaft (siehe Text) des 47. Pixels befindet, erhalten das 46. und das 47. Pixel beide den Wert „47“. **Mittleres und rechtes Bild:** das 28. und das 49. Pixel befinden sich in der Nachbarschaft des 50. Pixels. *Allen* Pixeln mit dem Wert „49“ und *allen* Pixeln mit dem Wert „28“ wird der Wert „50“ zugewiesen.

In einem zweiten Durchlauf findet eine Umnummerierung der Objektnummern in aufsteigender Reihenfolge statt, so dass die Pixel des ersten Clusters den Wert eins zugewiesen bekommen, die Pixel des nächsten Clusters erhalten den Wert zwei, die Pixel des dritten Clusters den Wert drei und die Pixel des  $n$ -ten Clusters den Wert  $n$  (siehe Abb. 2.3). Die Umnummerierung beginnt mit dem zuletzt gebildeten Cluster.

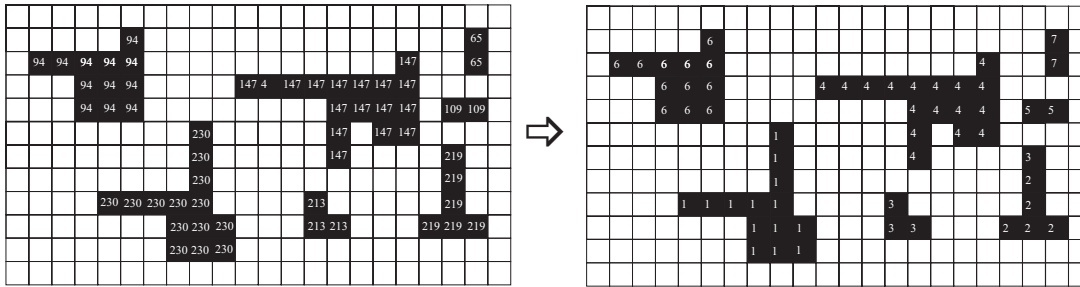


Abbildung 2.3: **Linkes Bild:** Fertiges Clusterbild mit Pixelstellen als Pixelwerte. **Rechtes Bild:** Fertiges Clusterbild mit umnummerierten Objektnummern in aufsteigender Reihenfolge.

Sind die Cluster gebildet, untersucht man die Cluster auf ihre Größe hin. Dazu erfolgt eine Analyse des Clusterbildes bezüglich der Häufigkeit der unterschiedlichen Indizes. Die Anzahl der Pixel mit dem Index  $n$  ist entsprechend klein, wenn es sich um Störungen handelt. So lassen sich die Störpixel eliminieren. Für diesen Schritt werden die Pixel ungesetzt (d.h. sie erhalten den Wert „0“), die zu einem Cluster gehören, dessen Größe einen bestimmten Schwellwert unterschreitet (siehe Abb. 2.4). Das Ergebnis ist ein nahezu störungsfreies Differenzbild, in dem nur noch die Pixel gesetzt sind, die durch Bewegungen von Objekten entstanden sind. Dadurch ist die Bildung einer geeigneten ROI ermöglicht (siehe Abb. 2.5).

Das Objekttracking erfolgt innerhalb der ROI. Zum Teil liegen die Pixel in homogenen Flächen, die schwer zu tracken sind. Eine gute Methode, um dieses Problem zu umgehen, ist das Verfolgen der Objektkanten. Das Kantentracking parallel zur Kantenlage bereitet Schwierigkeiten. Daher sind die Ecken des Objekts bessere Merkmale, die sich für das Tracken eignen (siehe [8]). Ecken sind in einem Bild seltener und sind untereinander leichter voneinander zu unterscheiden als Kanten. Das Verfolgen von Kanten und Ecken in einem Bild reduziert den Rechenaufwand, da nicht die gesamte Objektfläche betrachtet wird.

Die ROI enthält neben dem Objekt den Bildhintergrund. Die Kanten und Ecken dieser Bereiche sind für das Tracken des Objekts aber irrelevant und

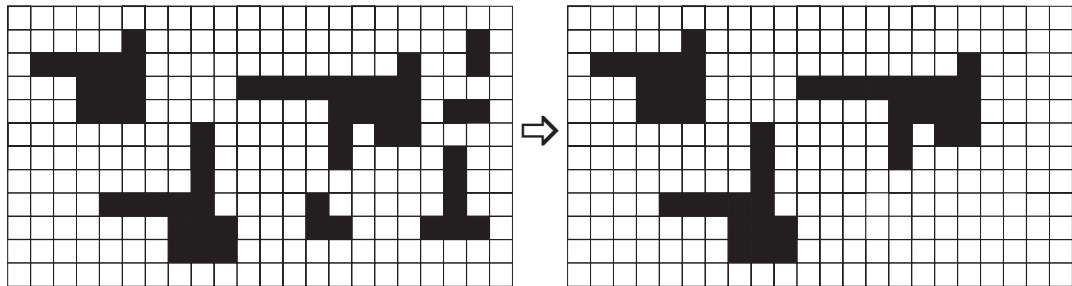


Abbildung 2.4: **Linkes Bild:** Unbearbeitetes Clusterbild, das noch Störungen enthält. **Rechtes Bild:** Bearbeitetes Clusterbild, in dem Cluster, die aus weniger als zehn Pixeln bestehen, entfernt sind.

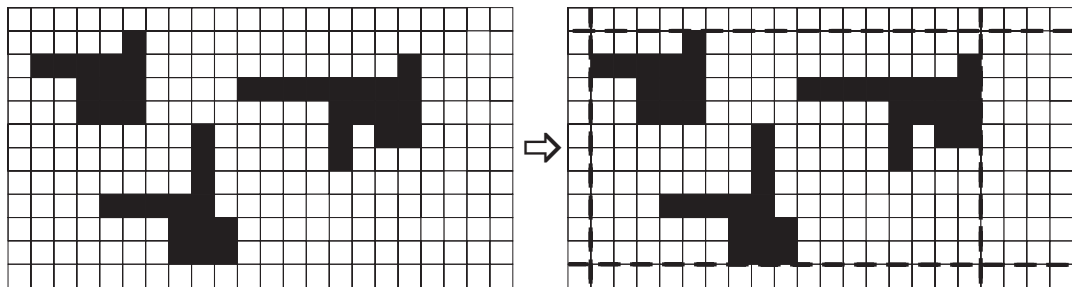


Abbildung 2.5: Der gestrichelt eingerahmte Bereich im rechten Bild ist die *Region of Interest* (Erläuterungen siehe Text).

sind gegebenenfalls Kandidaten für Fehlmatches. Dieses Problem lässt sich beheben, indem man nur an Stellen im Grauwertbild nach Kanten und Ecken sucht, sobald ein Pixel im Differenzbild gesetzt ist. Ein Verlust an Kanten- bzw. Eckeninformation ist nicht zu befürchten, da sich das störungsfreie Differenzbild aus der Positionsänderung der Objektkonturen ergibt, also aus den Kanten und Ecken. Objektpunkte, die an eine Bildstelle wandern, wo vorher ein anderer Objektpunkt mit gleichem Pixelwert lag, werden im Differenzbild nicht deutlich gemacht. Dies trifft häufig für solche Objektpunkte zu, die zu gleichen homogenen Flächen gehören.

## 2.3 Kantenfilter

Um Kanten in einem Bild zu suchen, muss zunächst gefragt werden, was eine Kante in einem digitalen Bild ist. Intuitiv entspricht eine Kante scheinbar

der Grenze zwischen zwei Pixeln, deren Helligkeitswerte sich signifikant unterscheiden. Diese zwei einzelnen Pixel gehören aber häufig unterschiedlichen benachbarten Regionen an, wo ein Gebiet heller erscheint als das daneben liegende. Zwischen jedem Paar benachbarter Pixel, wo ein Pixel in dem helleren Gebiet liegt und das andere Pixel außerhalb davon, findet sich ein prägnanter Unterschied des Grauwerts. In diesem Fall entspricht eine Kante einem Ort im Bild, an dem sich die Helligkeit abrupt oder wenigstens schnell ändert (siehe [9] und [10], Kapitel 7.4). Die Helligkeitswerte scheinen zu springen. Dies äußert sich in einem sich stark und schnell ändernden Grauwert (Grauwertgradient) an der Stelle der Kante (siehe Abb. 2.6). Dies ist die Art von Kanten, die L.G. Roberts 1965 ursprünglich entdeckt hat (vgl. [11]).

Zwischen den Strukturen des Objekts und den Kanten im Bild besteht nicht unmittelbar ein Zusammenhang. So können Grauwertgradienten in einem Bild etwa durch Schattenränder entstehen.

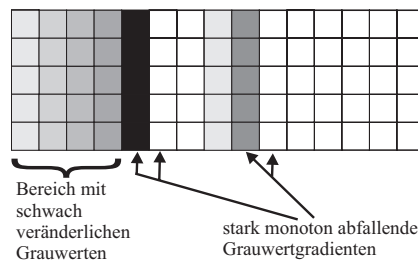


Abbildung 2.6: Eine Kante in einem Bild entspricht einem starken oder zumindest schnell veränderlichen Grauwertgradienten.

Die Gradienten in einem Bild lassen sich durch den Vergleich des Grauwertes zweier benachbarter Pixel oder mehrerer in der Nachbarschaft befindlicher Pixel ermitteln. Das Ergebnis ist ein Grauwertgradient. Überschreitet der Grauwertgradient einen bestimmten Schwellwert, handelt es sich um eine Kante. Je nach Aufgabe, d.h. je nach gewünschter Genauigkeit der Kantengerichtung, der Kantenstärke oder der Übereinstimmung der ermittelten Objektkante mit der realen Objektkante, bieten sich verschiedene Kantenfilter an. In der hier gestellten Aufgabe ist es erforderlich, einen einfachen und daher schnellen Kantenfilter einzusetzen.

### 2.3.1 Eindimensionaler Differenzoperator

Der eindimensionale Differenzoperator untersucht die zwei Pixel links und rechts bzw. oberhalb und unterhalb des Pixels  $i$  auf ihren Grauwertgradienten hin. Der eindimensionale Differenzoperator hat die Form (vgl. [12])

$$\vec{D}_x = \frac{1}{2} [1 \ 0 \ -1] \quad \text{und} \quad \vec{D}_y = \frac{1}{2} \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}.$$

Durch die Faltung von  $\vec{D}$  und dem Bild  $f(x, y)$

$$f(x, y) * \vec{D}_x \text{ bzw. } f(x, y) * \vec{D}_y.$$

werden Kanten detektiert, die senkrecht zur Operatorrichtung liegen:  $\vec{D}_x$  durchläuft das Bild spaltenweise und erkennt daher vertikale Kanten.  $\vec{D}_y$  durchläuft das Bild zeilenweise und erkennt daher horizontale Kanten. Wenn der Grauwertgradient zwischen dem  $Pixel_{i-1}$  und dem  $Pixel_{i+1}$  groß genug ist, liegt an der Stelle  $i$  eine Kante vor. Das Ergebnis wird in einem Kantenbild  $k_i(x, y)$  markiert:

$$k_i(x, y) = \begin{cases} 0 : & f_i(x-1, y) - f_i(x+1, y) < Schwelle \\ & \vee f_i(x, y-1) - f_i(x, y+1) < Schwelle \\ 1 : & f_i(x-1, y) - f_i(x+1, y) \geq Schwelle \\ & \vee f_i(x, y-1) - f_i(x, y+1) \geq Schwelle \end{cases}.$$

Um die Kantensuche zu beschleunigen, kann dieser Pixelvergleich in einem Bildbereich erfolgen, der für die jeweilige Aufgabe relevant ist. In dieser Arbeit erfolgt der Pixelvergleich an jeder Stelle der *Region of Interest* (siehe Seite 13). Um den Kantensuche leistungsfähiger zu gestalten, sucht man im Grauwertbild nur dann nach Kanten, wenn ein Pixel im Differenzbild gesetzt ist. Auf diese Weise sind unwichtige Bereiche im Hintergrund von der Kantendetektion ausgeschlossen.

führen

### 2.3.2 Prewitt-Operator

Eine Erweiterung des eindimensionalen Differenzoperators ist der Prewitt-Operator. Dieser hat die Form (vgl. [10])

$$P_x = \begin{pmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{pmatrix} \quad \text{und} \quad P_y = \begin{pmatrix} -1 & -1 & -1 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{pmatrix}.$$

$P_x$  bestimmt vertikale Kanten gut, dafür horizontale Kanten schlecht und  $P_y$  bestimmt horizontale Kanten gut, dafür aber vertikale Kanten schlecht.

Deswegen bietet sich zur Bestimmung des Grauwertgradientenbetrags  $M$  eine Kombination von  $P_x$  und  $P_y$  an:

$$M \approx \sqrt{P_x^2 + P_y^2}.$$

In der Praxis übergeht man das Ausführen der Wurzel jedoch meist aus Geschwindigkeitsgründen. Stattdessen genügt die Betrachtung der Betragssumme

$$|P_x| + |P_y| .$$

Im Vergleich zum eindimensionalen Differenzoperator hat dieser Kantenoperator den Vorteil, dass ein Störpixel, das große Kantenwerte in alle Richtungen erzeugt, unterdrückt wird. Bei dieser Art der Kantendetektion werden nämlich mehrere Nachbarpixel betrachtet, die das Gradientenergebnis bestimmen. Der Prewitt-Operator gibt zudem Aufschluss über die Kantenorientierung. Variationen des Prewitt-Operators sind z.B. der Sobel-Operator und der Roberts-Kantendetektor, die sich vom Prewitt-Operator lediglich in der Gewichtung der Filtermaske unterscheiden (siehe [12]).

### 2.3.3 Canny-Operator

Die zwei wichtigsten Kriterien für Kantenextraktionsverfahren sind die Kantenerkennung und die Positionsbestimmung der Kante (siehe [9]). Ein drittes Kriterium, das von John Canny aufgestellt wurde, ist das „multiple response critereon“. Dieses Kriterium gibt an, wie hoch die Wahrscheinlichkeit ist, dass ein Verfahren eine Kante mehrmals entdeckt.

Unter diesem Aspekt hat John Canny drei Bedingungen an einen Kantendetektor gestellt (siehe [13]):

- Wichtige Kanten sollten nicht verfehlt werden, es soll nicht mehrere Antworten auf eine Kante geben (Detektionskriterium).
- Der Abstand zwischen der tatsächlichen Kante und der lokalisierten Kante soll minimal sein (Lokalisationskriterium).
- Mehrfache Antworten sollen auf eine einzelne Kante minimiert werden (Eine-Antwort-Kriterium).

Dies bedeutet im Einzelnen, dass alle in einem Bild vorkommenden Kanten erkannt werden sollen, aber keine weiteren oder falschen Kanten. Da die Erkennung vom Signal-Rausch-Verhältnis des untersuchten Bildes abhängt, ist ein geeignetes Verfahren zu finden, das den Rauschanteil möglichst gering

hält. Die zweite Bedingung beinhaltet die Lokalisation der Kante nach Möglichkeit auf „Sub-Pixel-Ebene“, um die exakte Position der Kante zu erhalten.

Die Erfüllung dieser Bedingungen lässt sich erreichen durch die Kombination einer Glättung mittels Gaußfunktion  $g(x)$  mit einer Kantendetektion. Die Begründung hierfür liegt darin, dass die erste Ableitung der Gaußfunktion eine gute Annäherung an den von J. Canny gefundenen optimalen Kantendetektor ist, mit der im Zweidimensionalen effizienter gerechnet werden kann.

Das Bild  $f(x, y)$  wird also zunächst mit einem Gauß-Filter der Form

$$g(x) = e^{-\frac{x^2}{2\sigma^2}}$$

jeweils entlang der Zeilen und entlang der Spalten des Bildes geglättet. Die Standardabweichung  $\sigma$  bestimmt die Stärke der Glättung. Wegen horizontaler und vertikaler Faltung des Bildes ergeben sich für jedes Pixel im Bild zwei Werte, die groß sind, falls es sich um einen Gradienten handelt und klein, falls kein Gradient vorliegt. Anschließend erfolgt die Bestimmung der Gradientennormalenrichtung  $n$  für jedes Pixel im Bild :

$$n = \frac{\nabla(g * f)}{|\nabla(g * f)|}$$

Als nächstes wird der Betrag des Gradienten und daraufhin dessen Lage berechnet:

$$\frac{\partial^2(g * f)}{\partial n^2} = 0 .$$

Dieses ist die *nonmaximum suppression*, bei der alle Pixel auf Null gesetzt werden, deren Gradientenbeträge keine lokalen Maxima sind. Der Grund für die nonmaximum suppression liegt in dem Verhalten der Pixel, die neben einem Kantenpixel liegen: da die Gradientenbeträge der Kantenpixel größer sein sollten als die Gradientenbeträge für neben dem Kantenpixel liegende Pixel, lassen sich auf diese Weise alle Gradienten auf die Breite eines einzelnen Pixels verdünnen. Somit ergeben sich für die Detektion nur echte Kanten.

Abschließend wird ein *Hystere-Schwelwertverfahren* zur Unterdrückung von Falschanzeigen eingeführt. Das Hystere-Schwelwertverfahren besitzt einen oberen Schwellwert  $T_{hoch}$  und zusätzlich einen unteren Schwellwert  $T_{tief}$ . Jedes Pixel, das  $T_{hoch}$  überschreitet, wird als Kantenpixel gesetzt. Alle Pixel, die von diesem Pixel aus erreichbar sind und einen Wert größer als  $T_{tief}$  haben, werden ebenfalls als Kantenpixel gesetzt. Indem man das Bild in mehreren Durchgängen untersucht, erhält man alle Kantenpixel.

Dadurch ist das sichere Antworten auf alle echten Kanten gewährleistet und es wird nur eine Kante detektiert, wenn nicht mehr als eine Kante vorliegt.

Der Vorteil des Canny-Opertors liegt neben dem Erhalt der Kantensorientierung zusätzlich in dem direkten Wissen über die tatsächliche Kantestärke. Kantenfilter wie z.B. der Prewitt- oder Sobeloperator liefern nur die Kantensorientierung.

Ein Einsatz des Canny-Operators ist in dieser Arbeit nicht sinnvoll, da die Kenntnis über die Kantestärke entbehrlich ist, um den Rechenaufwand gering zu halten.

## 2.4 Das Stereosystem

Dem Stereosystem stehen zwei raumfeste und lokal benachbarte Kameras zur Verfügung, die zwei Bilder aus unterschiedlicher Perspektive zeitgleich aufnehmen. Die Kameras sollten so zueinander angeordnet sein, dass ein Objekt nach Möglichkeit immer auf einen binokularen Bereich abgebildet wird (siehe Abb. 2.7). Wenn ein Objekt auf die Bildebene der einen Kamera abgebildet wird, aber nicht auf die andere Kamerabildebene, ist ein stereoptisches Sehen nicht möglich.

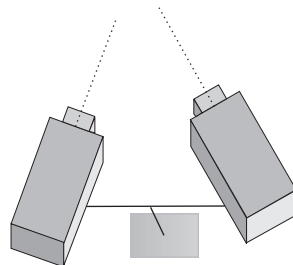


Abbildung 2.7: Anordnung des Stereokamerasystems

Für das Stereosehen müssen die Punkte in dem einen Kamerabild gefunden werden, die mit Punkten des anderen Kamerabildes korrespondieren. Korrespondierende Pixel sind Abbildungen des gleichen Objektpunktes in der Umwelt. Der Abstand dieser Punkte auf den Kamerasensoren wird als Disparität bezeichnet. Aus den Disparitäten ist eine Rekonstruktion der dreidimensionalen Ansicht einer Szene möglich. Die Disparität bildet man aus den unterschiedlichen Bildpositionen zweier korrespondierender Pixel und aus der Kenntnis der Kameraparameter (siehe Abb.2.8).

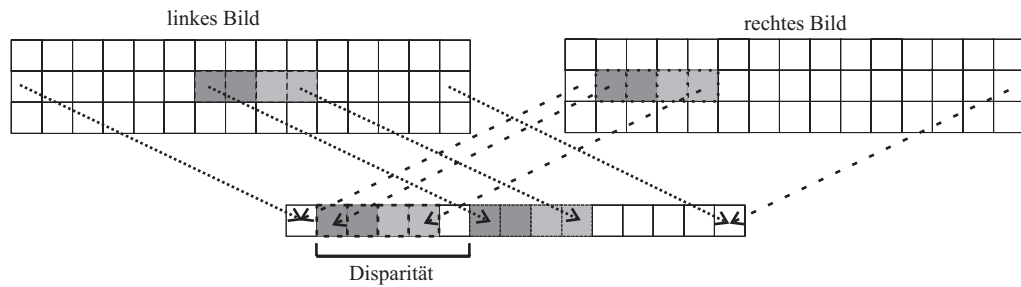


Abbildung 2.8: Die Disparität ergibt sich aus dem Abstand korrespondierender Pixel.

### 2.4.1 Korrespondenzproblem

Die Qualität der Suche nach korrespondierenden Pixeln (Disparitäts- bzw. Korrespondenzsuche) bestimmt die Genauigkeit der später zu ermittelnden Weltkoordinaten. Die Schwierigkeit korrespondierende Pixel zu finden liegt darin, verschiedene Erscheinungen wie Wiederholungsmuster, Helligkeitsschwankungen, Verdeckungen etc. zu bewältigen. Hinzu kommt, dass die Korrespondenzsuche eine echtzeitfähige Objektverfolgung zulassen muss.

Die Suche nach Korrespondenzpixeln lässt sich in zwei Gruppen einteilen (vgl. [14], Kapitel 2): die flächenbasierten Ansätze und die merkmalsbasierten Ansätze.

Die flächenbasierten Ansätze beruhen auf dem Vergleich von Pixelmustern in beiden Kamerabildern. Die Flächen können sehr kleine Bildausschnitte sein bis hin zu einzelnen Pixeln. Verglichen wird dann beispielsweise die Intensität eines jeden Pixels.

Die merkmalsbasierten Ansätze wählen aus den Bildern abstrakte, spärlichere Merkmale aus, das können u.a. Linien, Ecken, Figuren oder Formen sein. Die Merkmale des einen Bildes werden dann mit den Merkmalen des anderen Bildes verglichen. Dabei beschränkt sich die Lokalisation der Korrespondenz nicht auf einen einzelnen Punkt, sondern auf das gesamte Merkmal. Daraus ergeben sich nur den Merkmalen und nicht den einzelnen Pixeln zugeordnete Disparitätswerte. Man erhält also keine vollständige Disparitätskarte, wie es bei den flächenbasierten Ansätzen in der Regel der Fall ist. Wenn genauere Disparitäten nötig sind, ist eine zweite Analyse nötig, in der die exakten Disparitäten ermittelt werden. Die merkmalsbasierte Korrespondenzsuche ist nur für eine grobe Disparitätsschätzung geeignet.

Bei der Disparitätssuche treten u.a. folgende Korrespondenzprobleme auf:

- Aperturproblem

- Verdeckungsproblem
- Wiederholungsmuster
- Helligkeitsschwankungen
- Rauschen.

Betrachtet man eine Kontur, etwa eine Kante, durch einen kleinen Ausschnitt (Apertur), so ist die Bewegung der Kontur nicht eindeutig bestimmbar. Dies bezeichnet man als Aperturproblem. Bei einer Bewegung der Kontur hinter einer Apertur besitzt die sichtbare Konturbewegung nicht ausreichend Information. Die Ermittlung der Geschwindigkeit und der Bewegungsrichtung der Kontur ist erschwert oder sogar unmöglich. Lediglich die Bewegung senkrecht zum Grauwertgradienten lässt sich untersuchen. Diese wiederum kann von mehreren möglichen Bewegungen aus unterschiedlichen Richtungen und Geschwindigkeiten resultieren (siehe Abb. 2.9). Damit ist eine sichere Zuordnung korrespondierender Pixel unzuverlässig.

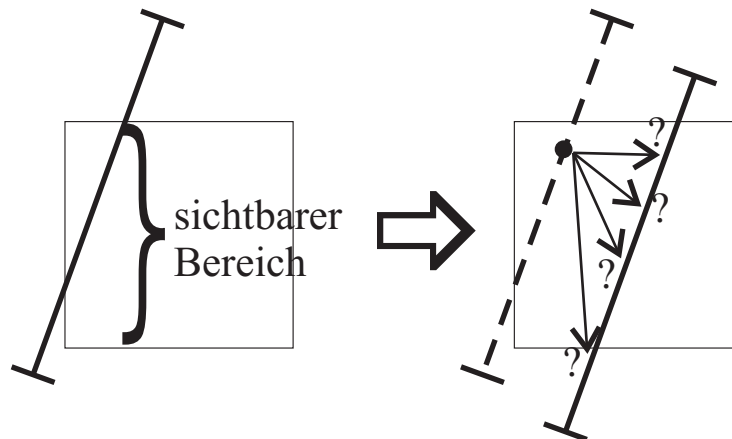


Abbildung 2.9: Die Zuordnung gleicher Objektpunkte ist nicht eindeutig.

Wenn ein Objektpunkt auf einem der beiden Bilder aufgrund einer Verdeckung nicht abgebildet ist, so spricht man vom Verdeckungsproblem (siehe Abb. 2.10). In diesem Fall können korrespondierende Pixel in den verdeckten Bereichen nicht gefunden werden, da ein Objektpunkt nur auf eine der zwei Bildebenen abgebildet wird.

Treten im Bild mehrere identische Objekte oder Flächen auf, muss die Korrespondenzsuchmethode diese Schwierigkeit bewältigen, um gute Disparitätswerte zu liefern (siehe Abb. 2.11). Das Problem beim Auffinden von



Abbildung 2.10: Verdeckung: Der Karton im mittleren Bereich des linken Bildes ist auf dem rechten Bild durch die weiße Fläche zur Hälfte verdeckt.

Disparitäten in diesem Fall scheint zunächst ähnlich zum Aperturproblem. Der Unterschied zum Aperturproblem liegt darin, dass kein periodisches Muster vorliegt, sondern zufällig verteilte Objekte, deren Verschiebungsvektoren unterschiedliche Richtungen haben können. Das Korrespondenzproblem kann gelöst werden, indem die Aufnahme aufeinanderfolgender Bilder so schnell erfolgt, dass der mittlere Verschiebungsvektor der Objekte oder Flächen kleiner ist als der Objektabstand.

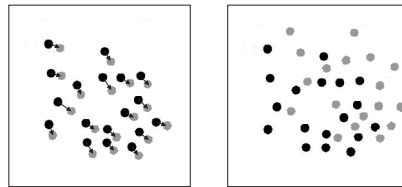


Abbildung 2.11: In beiden Bildern sind die schwarzen Kreise bei der ersten Bildaufnahme entstanden und die grauen Kreise stellen die Verschiebung der schwarzen Kreise in einer zweiten Bildaufnahme dar. Im linken Bild ist der mittlere Objektabstand größer als die Verschiebungsvektoren. Im rechten Bild ist der mittlere Teilchenabstand kleiner als die Verschiebungsvektoren.

Es kann vorkommen, dass die beiden Kameras Bilder mit unterschiedlicher Helligkeit liefern (siehe Abb. 2.12). Die Gründe für solche *Helligkeitsschwankungen* sind u.a.

- schlecht eingestellte Kameraobjektive,
- unterschiedliche Kamerapositionen und die damit verbundene Richtungsabhängigkeit der Reflexion,

- leichte Toleranzabweichungen bezüglich der technischen Grunddaten der Kameras, wie zum Beispiel Verschlusszeit, Sensitivität, minimale Beleuchtung,
- eine Umgebung mit variabler Helligkeit, beispielsweise flackernde Neonröhren oder
- die technische Grundausstattung der Kameras.

Hier kann ein Korrelationskriterium, das die Bilder bezüglich ihrer Helligkeit angleicht, Abhilfe schaffen. Einige Digitalkameras sind bereits so ausgestattet, dass sie die Bildhelligkeiten automatisch der Umgebung anpassen (etwa *automatic gain control*), wodurch die Helligkeit in beiden Kamerabil- dern annähernd gleich sein sollte.

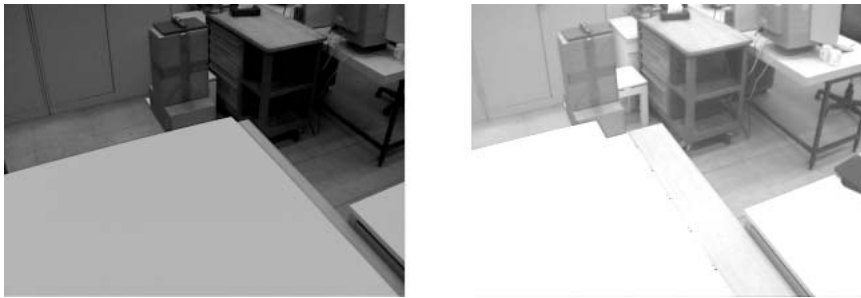


Abbildung 2.12: Das linke Bild hat eine geringere Helligkeit als das rechte. Abbildungen des gleichen Raumpunktes haben unterschiedliche Grauwerte.

Wie in Abschnitt 2.1 angesprochen, verursachen die CCD-Chips geringe Bildstörungen. Diese Bildstörungen lassen sich durch Modifikatoren oder Filter beheben.

Das *Rauschen* resultiert zum einen daraus, dass der Bestrahlungswert auf den CCD-Sensoren für dunkle Flächen und Objekte sehr gering bis verschwindend ist.

Zum anderen kommt neben dem oben beschriebenen Rauschen noch thermisches Rauschen hinzu. Das thermische Rauschen entsteht durch Kristallgitterbewegungen in den CCD-Sensoren, die sich aus Wärme und daraus resultierenden Elektronenübergängen auf höhere Energiezustände ergeben. Einige Digitalkameras regeln die Bildhelligkeit automatisch über die maximal eintreffende Helligkeit. Das thermische Rauschen ist bei diesen Kameras nur bei Dunkelheit erkennbar.

Die unterschiedlichen Farbkanäle reagieren verschieden auf die Wellenlängen der Photonen: das ist eine dritte Ursache für das Rauschen. Wegen

der geringeren Quanteneffizienz (Verhältnis der mittleren Anzahl ausgelöster Photoelektronen zur mittleren Anzahl der einfallenden Photonen) löst blaues Licht in einem Halbleiter weniger Elektronen aus als rotes Licht und daher muss der blaue Farbkanal einer CCD-Kamera mehr verstärkt werden als der rote Farbkanal. Entsprechend groß ist dann das Rauschen für blaue Farbtöne (siehe Abb. 2.13).

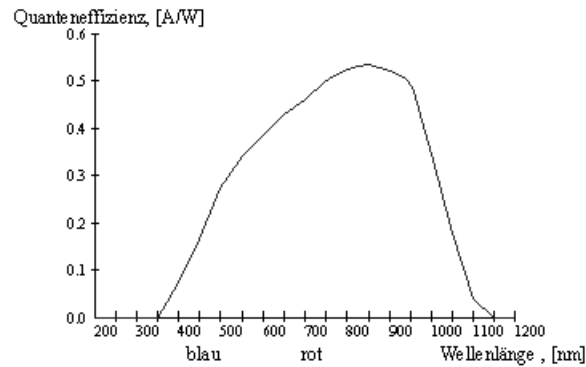


Abbildung 2.13: Quanteneffizienz in Abhängigkeit von der Wellenlänge  $\lambda$ .

## 2.4.2 Korrespondenzsuche

Die Korrespondenzsuche umfasst den Vergleich eines jeden Pixels des einen Bildes mit jedem Pixel des anderen Bildes. Bei dieser Methode können viele Fehlmatches entstehen. Daher ist es erforderlich, die Suche auf weniger Pixel zu reduzieren.

Da korrespondierende Pixel des einen Bildes auf korrespondierenden Bildzeilen des anderen Bildes liegen, lässt sich die Suche optimieren. Bei diesen Zeilen handelt es sich um Epipolarlinien. Eine Epipolarlinie im rechten Bild ergibt sich aus dem Schnitt von zwei Ebenen: Die erste Ebene (die Epipolarebene) ergibt sich aus den vier Punkten (vgl. Abb. 2.14)

- optisches Zentrum  $c_l$  der linken Kamera,
- optisches Zentrum  $c_r$  der rechten Kamera,
- Gegenstandspunkt  $P(x, y, z)$  im Raum und
- Abbildungspunkt  $f_l(x, y)$  des Gegenstandspunktes auf die Bildebene der linken Kamera.

Die zweite Ebene ist die Bildebene der rechten Kamera. Die Epipolarlinie des linken Bildes ist die Verbindungslinie des Abbildungspunktes  $f_l(x, y)$  mit dem Epipol der linken Kamera. Die korrespondierende Epipolarlinie des rechten Bildes ist die Verbindungslinie des Schnittpunktes aus Epipolarebene und Bildebene der rechten Kamera mit dem Epipol der rechten Kamera. Ein Epipol ist der Punkt, in dem sich alle Epipolarlinien schneiden. Der Epipol der linken Kamera ist die Abbildung des optischen Zentrums der rechten Kamera auf die linke Bildebene (siehe Abb. 2.15). Umgekehrt ist der Epipol der rechten Kamera die Abbildung des optischen Zentrums der linken Kamera auf die Bildebene der rechten Kamera. Sind die Epipolarlinien parallel, liegt der Epipol im Unendlichen.

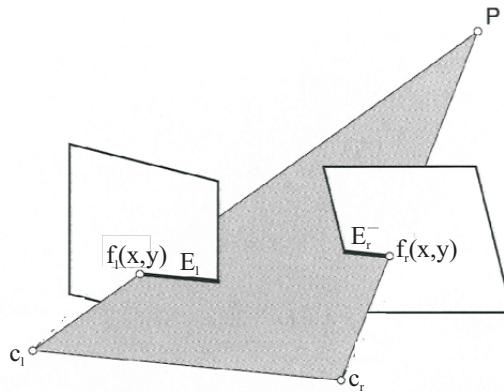


Abbildung 2.14: Zur Entstehung der Epipolarlinien.  $c_l$ =optisches Zentrum der linken Kamera,  $c_r$ =optisches Zentrum der rechten Kamera,  $E_l$ =Epipolarlinie der linken Bildebene,  $E_r$ =Epipolarlinie der rechten Bildebene,  $f_l(x, y)$ =Pixelkoordinate des linken Korrespondenzpixels,  $f_r(x, y)$ =Pixelkoordinate des rechten Korrespondenzpixels,  $P = P(x, y, z)$ =Umwelt punkt.

Um die Korrespondenzsuche zu vereinfachen, ist es sinnvoll, die optischen Achsen der beiden Kameras

- parallel zueinander und
- auf gleicher Höhe

auszurichten.

Durch diese Maßnahme ist eine Einschränkung des Disparitätssuchbereiches auf eine Zeile im Bild möglich, da korrespondierende Pixel in diesem Fall auf derselben Bildzeile liegen müssen (siehe Abb. 2.16).

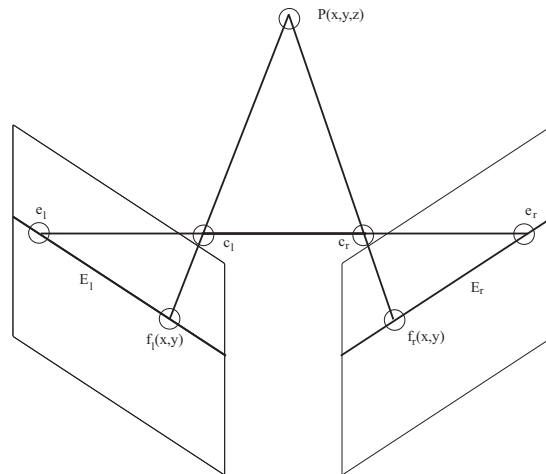


Abbildung 2.15: Die Bezeichnungen in der Skizze sind analog zu den Bezeichnungen in der Abbildung 2.14. Der Abbildungspunkt des optischen Zentrums  $c_r$  der rechten Kamera auf die linke Bildebene ist der Epipol  $e_l$  der linken Kamera.

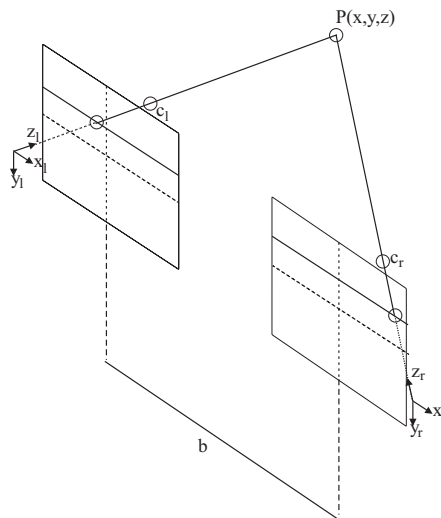


Abbildung 2.16: Sind die CCD-Chips parallel und auf gleicher Höhe, so liegen die Epipolarlinien auf gleichen Bildzeilen.

Aufgrund der Kalibrierungsungenauigkeit ist es sinnvoll, anstatt auf einzelnen Epipolarlinien zu operieren, die Suche auf Epipolarbänder mit einer Breite von zwei bis sechs Pixeln auszudehnen.

Eine weitere Maßnahme für die Leistungsoptimierung der Korrespondenzsuche ist eine Einschränkung des Suchbereichs.

In dieser Arbeit erfolgt die Korrespondenzsuche innerhalb der ROI. Ferner wird im Referenzbild nach korrespondierenden Pixeln gesucht, sofern im Modellbild eine Kante vorliegt. Das Referenzbild entspricht dem linken Kamerabild und das Modellbild dem rechten Kamerabild. Im rechten Grauwertbild werden für den Fall einer Kante alle Nachbarpixel eines 3x3-großen Bereiches um einen Pixel extrahiert. Diese Bildmatrix wird schrittweise über das linke Grauwertbild bewegt und dort mit einer 3x3-Bildmatrix verglichen. Der Vergleich beinhaltet für jeden Schritt die elementweise Subtraktion der Pixelgrauwerte und die anschließende Summation der betragsmäßigen Differenzen derselben. Der kleinste Wert, der daraus resultiert, ist der beste Match und ist somit das korrespondierende Pixel. Ist aber die Differenz des zweitbesten Matches nicht wesentlich größer als die des besten Matches, dann sind sich die Matches zu ähnlich. In diesem Fall kann kein korrespondierendes Pixel gefunden werden.

Für den Abstand der zwei korrespondierenden Pixel, die Disparität  $d(x, y)$ , ergibt sich:

$$d(x, y) = | \vec{p}_r - \vec{p}_l |, \quad (2.1)$$

$$\vec{p}_r = \begin{pmatrix} x_{korr,r} \\ y_{korr,r} \end{pmatrix}, \vec{p}_l = \begin{pmatrix} x_{korr,l} \\ y_{korr,l} \end{pmatrix}$$

Hierin bezeichnen  $x_{korr,r}$  und  $x_{korr,l}$  die Positionen der korrespondierenden Pixel innerhalb einer Zeile,  $y_{korr,r}$  und  $y_{korr,l}$  die Positionen der korrespondierenden Pixel innerhalb einer Spalte, wobei

$$(x, x_{korr,r}, x_{korr,l}) \in [0, \text{Bildbreite}] \text{ und } (y, y_{korr,r}, y_{korr,l}) \in [0, \text{Bildhöhe}].$$

Wenn die korrespondierenden Pixel in der gleichen Bildzeile liegen, vereinfacht sich Gleichung 2.1 zu:

$$d(x, y) = | x_{korr,r} - x_{korr,l} |. \quad (2.2)$$

Mit Hilfe der gewonnenen Disparitäten können Aussagen bezüglich der Raumkoordinaten des Objekts getroffen werden.

### 2.4.3 Rekonstruktion der dreidimensionalen Szene

Die Rekonstruktion beinhaltet die Transformation der Weltkoordinaten aus den Disparitäten  $d(x, y)$ , aus der Lage der korrespondierenden Pixel im Kamerabild (das sind die Sensorkoordinaten) und den Kameraparametern. Die

beiden Kameras sind im Abstand  $b$  voneinander entfernt und die Brennweite ist durch  $f$  gegeben. Die Kamerakoordinaten  $(R_x, R_y, R_z)$  stehen dann mit den Bildkoordinaten  $f_r(x_r, y_r)$  und  $f_l(x_l, y_l)$  in folgender Beziehung zueinander:

$$\frac{x_l}{f} = \frac{R_x + \frac{b}{2}}{R_z}; \quad \frac{x_r}{f} = \frac{R_x - \frac{b}{2}}{R_z}. \quad (2.3)$$

Da die optischen Achsen der Kameras parallel zueinander sind und die Bildzeilen der Kameras kollinear zueinander ausgerichtet sind, ergibt sich weiter:

$$\frac{y_l}{f} = \frac{y_r}{f} = \frac{R_y}{f}. \quad (2.4)$$

Die Disparität ist der Abstand korrespondierender Punkte in einer Zeile, damit ergibt sich für die Disparität der Zusammenhang:

$$\frac{x_l - x_r}{f} = \frac{d}{f} = \frac{b}{z}. \quad (2.5)$$

Die Kamerakoordinaten ergeben sich damit zu

$$\begin{aligned} R_x &= b \frac{x_l + x_r}{2(x_l - x_r)}, \\ R_y &= b \frac{y_l + y_r}{2(x_l - x_r)}, \\ R_z &= b \frac{f}{x_l - x_r}. \end{aligned} \quad (2.6)$$

Ausnutzung der Gleichung 2.5 und die Voraussetzung, dass die Disparitätswerte im linken Bild enthalten sind, liefert für die Kamerakoordinaten

$$\begin{aligned} R_x &= \frac{b}{d} \left( x_l - \frac{d}{2} \right), \\ R_y &= \frac{b}{d} y_l, \\ R_z &= \frac{b}{d} f. \end{aligned} \quad (2.7)$$

Zum Schluss erfolgt die Transformation der Kamerakoordinaten in Weltkoordinaten. Der Winkel  $\phi$  beschreibt den Azimutwinkel der Kamera, was einer Drehung der Kamera um die  $x$ -Achse gleicht.  $\theta$  bezeichnet den Neigungswinkel, das entspricht einer Drehung der Kamera um die  $z$ -Achse. Für die Transformation der Kamerakoordinaten in Weltkoordinaten  $(W_x, W_y, W_z)$  gilt (vgl. [12], Kapitel 7.2):

$$\begin{pmatrix} W_x \\ W_y \\ W_z \end{pmatrix} = \begin{pmatrix} \cos(\phi) & -\sin(\phi)\sin(\theta) & -\sin(\phi)\cos(\theta) \\ \sin(\phi) & \cos(\phi)\sin(\theta) & \cos(\phi)\cos(\theta) \\ 0 & -\cos(\theta) & \sin(\theta) \end{pmatrix} \begin{pmatrix} R_x \\ R_y \\ R_z \end{pmatrix}. \quad (2.8)$$

## 2.5 Objektgeschwindigkeit

Um ein Objekt aus der Bewegung zu greifen, sind Vorhersagen über die Raumpositionen und die Kenntnis über die Orientierung des Objekts im Raum erforderlich. Das Ziel des Objekttrackings ist es, die Position eines Objekts zu späteren Zeitpunkten zu schätzen und leichter zu finden. Hierfür ist das Wissen über das Bewegungsverhalten, also Bewegungsrichtung und Objektgeschwindigkeit des Objekts, notwendig. Diese Punkte können nicht als bekannt vorausgesetzt werden. Um das System flexibel zu gestalten, sollte das System das Bewegungsverhalten des Objekts und die Objektorientierung selbstständig ermitteln. Für die Objektgeschwindigkeit allein genügt die Kenntnis über die Bewegungsrichtung. Das Wissen hierüber lässt sich sowohl über die zweidimensionale Bildebene als auch über den dreidimensionalen Raum erreichen. In beiden Fällen empfehlen sich Matchingverfahren.

Eine Möglichkeit, das Bewegungsverhalten eines Objekts über die zweidimensionale Bildebene zu ermitteln, ist die Bestimmung des Flussfeldes (optischer Fluss) in einem Bild. Dies ist ein direktes Verfahren, die Geschwindigkeit zu bestimmen (siehe Abschnitt 2.5.1). Hier besteht aber der Nachteil, dass noch eine Umrechnung der Bildinformation in eine dreidimensionale Ansicht erfolgen muss.

Die Hausdorffdistanz dagegen liefert nicht nur eine Möglichkeit, zweidimensionale Bilder zu vergleichen. Vielmehr lassen sich mit ihr die dreidimensionalen Räume bzw. die dreidimensionalen Objekte miteinander vergleichen. Die Hausdorffdistanz ist dabei nicht im Sinne einer euklidischen Distanz zu verstehen. Sie liefert eine Möglichkeit, zwei Bilder oder zwei Punktmengen miteinander zu vergleichen und deren Ähnlichkeit zu bewerten. Hierdurch lässt sich die Verschiebung des Objekts ermitteln (siehe Abschnitt 2.5.4).

### 2.5.1 Optischer Fluss

Der optische Fluss behandelt die Bewegung eines Pixels im Bild. Dabei kann sich sowohl die Kamera bewegen als auch das Objekt. Ein Raumpunkt  $\vec{P} = (X_s, Y_s, Z_s)^T$  werde auf einen Punkt  $\vec{p} = (x, y)^T$  in die Bildebene abgebildet. Die Kamera bewege sich in einer stationären Umgebung mit der

Translationsgeschwindigkeit  $\vec{T} = (T_x, T_y, T_z)^T$  und der Winkel- bzw. Rotationsgeschwindigkeit  $\vec{R} = (R_x, R_y, R_z)^T$  bezüglich des Kameraframes  $R_s$ . Somit hat das Objekt sechs Freiheitsgrade, die beschrieben werden durch die Komponenten der Translationsgeschwindigkeit (Beschreibung der Bewegungsrichtung) und der Rotationsgeschwindigkeit (Beschreibung der Orientierungsänderung). Die Gleichungen für den optischen Fluss lauten dann:

$$u = [x \frac{T_z}{Z_s} - T_x Z_s] + [xyR_x - (1 + x^2)R_y + yR_z]$$

$$v = [y \frac{T_z}{Z_s} - T_y Z_s] + [(1 + y^2)R_x - xyR_y - xR_z],$$

mit  $u = \dot{x}$  und  $v = \dot{y}$ . Für den Fall der stationären Kamera und bewegten Objekten ergeben sich die gleichen Beziehungen, nur mit vertauschten Vorzeichen.

Die Rekonstruktion der Objektbewegung ist nun auf drei Weisen möglich (siehe [15]) :

- direkt
- Tiefenvariation
- Matched Filter.

### Direktes Verfahren

Da das Objekt sechs Freiheitsgrade besitzt, ist die Objektbewegung durch sechs Unbekannte gegeben. Sind sechs Flussvektoren bestimmt, so lässt sich die Gleichung für den optischen Fluss lösen. Hieraus ergibt sich direkt die Geschwindigkeit bzw. das Bewegungsverhalten des Objekts. Die Rekonstruktion der Objektbewegung erfolgt durch die Inversion der Optik, d.h. eine Inversion der Flussgleichung.

### Tiefenvariation

Im Gegensatz zum Translationsfeld hängt das Rotationsfeld nicht von den Abständen der bewegten Bildelemente ab. Dies ermöglicht eine Bestimmung der Translation und der Rotation des Objektes an isolierten Punkten, sofern an diesen Stellen ein dichtes Flussfeld vorliegt. Aus der Differenz der Flussvektoren von zwei eng benachbarten Bildpunkten ist es daher möglich, auf die Tiefendifferenz der erzeugenden Objektpunkte zu schließen. Da die Rotationsanteile unabhängig von der Tiefe sind, zeigt dieser Differenzvektor in die Richtung des Expansions- bzw. Kontraktionspunktes des Translationsanteils. Der Schnitt mehrerer Differenzvektoren liefert den Expansionspunkt.

### Matched Filter

Ein Matched Filter errechnet zunächst erwartete Flussfelder für bestimmte Bewegungsweisen. Sobald ein zu analysierendes Flussfeld vorliegt, wird dieses mit den errechneten Flussfeldern verglichen. Das am besten mit dem aktuellen Flussfeld übereinstimmende Feld ist die gesuchte Objektbewegung. Die Rekonstruktion erfolgt durch den Abgleich mit einer erwarteten Nachbildung des Flussfeldes.

Der optische Fluss als Verfahren für die Bestimmung der Geschwindigkeit im Raum ist aus folgendem Grund für weit entfernte Objekte kompliziert: Da das Objekt aufgrund seiner Entfernung auf den Kameras klein abgebildet wird, sinkt die Auflösung der Objektabbildungen. Dadurch können fehlgeordnete Korrespondenzpunkte entstehen, die wiederum falsche Weltkoordinaten liefern. Eine verlässliche Messung der Objektverschiebung über den optischen Fluss ist deswegen ungewiss.

### 2.5.2 Objektmatching

Die Qualität des Objektmatchings muss den geforderten Bedingungen genügen. Ein sehr gutes Objektmatching ist oft mit viel algorithmischen Aufwand und zu überprüfenden Kriterien verbunden. In dieser Arbeit muss das Matching schnell abgeschlossen sein, um echtzeitfähig zu bleiben. Dem Matchingverfahren sind für diese Arbeit aber Grenzen gesetzt, da eine leichte Abweichung von der gematchten zur realen Objektposition eine große Abweichung für die prädierte Objektposition bedeuten kann.

In der Regel beinhaltet das Objektmatching einen Vergleich von zwei Objekten (d.h. hier gewöhnlich ein Bild dieser Objekte). Diese Objekte sind das Modellobjekt und das Referenzobjekt. Das Referenzobjekt ist gewissermaßen das Ursprungsobjekt und das Modellobjekt ist gleichzusetzen mit dessen Nachbildung. Im Bereich der *Computer Vision* gibt es zahlreiche Methoden, um diese Vergleiche durchzuführen. Beispiele sind das *Graph-Matching*, bei dem beispielsweise die Abstände zwischen den Punkten in einem Referenzbild mit den Punktabständen in einem Modellbild verglichen werden. Eine andere Methode ist das *Feature-Matching*. Bei dieser Methode extrahiert man Merkmale aus einem Referenzbild und sucht diese Merkmale in einem Modellbild. Dazu gehört u.a. das *Template-Matching*. Eine weitere Methode ist das *Intensity-Matching*, bei dem ein Vergleich von Intensitätsprofilen stattfindet. Die Methoden lassen sich beliebig weiterführen. Das Objektmatching dient in dieser Arbeit dazu, den neuen Raumpunkt eines verlagerten Objekts zu finden.

Im Folgenden soll als Beispiel das Template-Matching als Repräsentant des Feature-Matching vorgestellt werden. Anschließend wird auf die Hausdorffdistanz eingegangen, mit der ein Vergleichsverfahren zwischen zwei Punktwolken zur Verfügung steht (siehe [16]), das dem Graphmatching ähnlich ist.

### Template Matching

Das Template Matching ist ein direkter Vergleich eines Referenzobjekts (template) mit einem Modellobjekt. Ein Template (engl. Vorlage) entspricht einem einzelnen Pixel, einer Linie, einer Ecke oder ganzen Objekten. Der Vergleich oder die Suche dieser Vorlage erfolgt, indem man das Template pixelweise über das Modellbild bewegt. An jeder Stelle des Bildes findet eine Berechnung der Ähnlichkeit zwischen Template und Modellobjekt statt. Ähnlichkeitsmaße sind etwa die Kreuzkorrelation, der euklidische Abstand, der Laplace-Abstand oder die Minimierung der kleinsten Fehlerquadrate. Das Ergebnis ist ein Templatebild, in dem hohe Werte zu erkennen sind, wenn die Ähnlichkeit hoch ist und kleine Werte, wenn nur sehr wenig Ähnlichkeit besteht.

Häufig wird dieses Templatebild noch einem Matching-Kriterium unterzogen. Dieses Kriterium kann ein Schwellwert sein, der auf ein Templatebild wirkt. Durch diese Maßnahme werden nur solche Merkmale im Modellbild als übereinstimmend mit dem Template bewertet, die eine genügend hohe Antwort auf den Schwellwert liefern.

Das größte Problem des Template Matchings ist der hohe Rechenaufwand, der schon bei kleinen Bildbereichen (ca. 20x20 Pixel) in einem 700x512 großen Bild für rotationsinvariante Objekte ca.  $10^8$  Pixelvergleiche erfordert. Ein weiteres Problem bei diesem Verfahren ist, dass Störpixel nicht ignoriert werden und das Ergebnis daher verfälschen können.

### 2.5.3 Hausdorffdistanz

#### Gerichtete Hausdorffdistanz

Die Hausdorffdistanz zwischen zwei endlichen Punktmengen  $A$  und  $B$  ist definiert als (siehe [16])

$$h(A, B) = \max_{a_i \in A} \min_{b_j \in B} \|a_i - b_j\|.$$

Dabei ist  $B$  eine endliche Bildpunktmenge mit Punkten  $b_j$ , mit  $j = 1, \dots, N_B$ , und  $A$  eine endliche Modellpunktmenge mit Punkten  $a_i$ , mit  $i = 1, \dots, N_A$ .  $N_x$  bezeichnet hierbei die Anzahl aller Punkte in einer Menge.

Die gerichtete Hausdorffdistanz  $h(A, B)$  von  $A$  nach  $B$  ist klein, wenn sich jeder Punkt von  $A$  in der Nähe eines Punktes aus  $B$  befindet. Diese Matchingmethode ist in der Praxis allerdings sehr anfällig für Rauschen, Unvollständigkeit und für Punktmengen mit geringen Dichten. Ein einzelner verrauschter Punkt aus  $A$  etwa, der weit entfernt von  $B$  ist, verursacht ein großes  $h(A, B)$ . Punkte, die aus Unvollständigkeit nicht in der Punktwolke  $A$  enthalten sind, verursachen ein zu kleines  $h(A, B)$ . Die *partielle Hausdorffdistanz* kann dieses Problem behandeln (vgl. [17]).

### Partielle Hausdorffdistanz

Seien  $\min_p(\vec{a}_1, B), \dots, \min_p(\vec{a}_N, B)$  die Punktabstände für die Punkte  $\vec{a}_i$  aus  $A$  zu den Punkten  $\vec{b}_i$  aus  $B$ . Nach Sortierung dieser Punktabstände erhält man eine Folge

$$\min_p(\vec{a}_{s(1)}, B)_{1:N_A}, \dots, \min_p(\vec{a}_{s(N)}, B)_{N_A:N_A}.$$

Die Notation  $(i : N_A)$  bedeutet der  $i$ -te Wert von  $N_A$ , wenn die  $N_A$  Werte der Größe nach sortiert werden. D.h., dass  $1 : N_A$  das Minimum und  $N_A : N_A$  das Maximum der  $N_A$  Werte bezeichnet.

Die partielle Hausdorffdistanz ist definiert als

$$h_p(A, B) = \max_{a_i \in A} (\min_{b_i \in B} \|a_i - b_i\|)_{p:N_A},$$

wo  $\max_{x \in X} (g(x))_{p:N}$  den  $p$ -ten Abstand aus der Menge  $X$  beschreibt.

Die partielle Hausdorffdistanz ist derart zu verstehen, dass zu jedem Punkt  $a_i$  aus  $A$  der Punkt mit dem kleinsten euklidischen Abstand zu  $B$  zu finden ist, anschließend die Abstände der Größe nach zu sortieren sind und den gewünschten  $p$ -ten Abstand zu betrachten. Auf diese Weise werden die größten, eventuell durch Störungen entstandenen Abstände ausgeschlossen.

Der Vorteil der partiellen Hausdorffdistanz als Matchingverfahren liegt darin, dass der Einfluss von Rauschpixeln auf das Matchingergebnis klein bleibt. Aus diesem Grund eignet sich die partiellen Hausdorffdistanz bei klein abgebildeten Objekten gut als Matchingmethode.

Die Punktwolke ist in dem Bereich, wo sich das Objekt befindet, dicht verteilt. Aufgrund dieser dichten Verteilung ist die partielle Hausdorffdistanz ein geeignetes Mittel, um die Abstände zweier Punktwolken zu bestimmen. Es bleibt jedoch die Schwierigkeit, ein sinnvolles  $p$  zu finden.

Da die meisten Störpunkte einen großen Abstand voneinander haben und die Punkte des zu verfolgenden Objekts dicht beieinander liegen, bietet sich in dieser Arbeit folgende Ermittlung für  $p$  an: Von den sortierten Punktabständen werden die Differenzen benachbarter Punktabstände gebildet. Wenn

die Differenz eine Schwelle  $\text{Diff}_{\min}$  überschreitet, erhält ein Index „ $c$ “ eine negative Bewertung. Ist die Differenz kleiner als die Schwelle  $\text{Diff}_{\min}$ , erhält der Index  $c$  eine positive Bewertung. Das Subtrahieren benachbarter Punktabstände erfolgt solange, bis der Index  $c$  eine Schwelle  $c_{\min}$  erreicht hat. Der Punktabstand, der an dieser Stelle erreicht wird, ist die gesuchte partielle Hausdorffdistanz.

#### 2.5.4 Bestimmung der Geschwindigkeit mittels Hausdorffdistanz

Mit Hilfe der Hausdorffdistanz lässt sich die räumliche Verschiebung einer Raumwolke ermitteln. Hieraus kann die Punktwolkengeschwindigkeit direkt bestimmt werden. Die Punktwolke  $A(\vec{W}_A) = A(W_{Ax}, W_{Ay}, W_{Az})$  aus dem Bild  $f_A(x_A, y_A, t_A)$  wird über den durch das Bild  $f_B(x_A + u, y_A + v, t_B)$  mit der Punktwolke  $B(\vec{W}_B) = B(W_{Bx}, W_{By}, W_{Bz})$  erzeugten Raum geschoben. Hierbei bezeichnen  $u = \dot{x}$  und  $v = \dot{y}$  den in Abschnitt 2.5.1 beschriebenen optischen Fluss. Die Hausdorffdistanz gibt die Übereinstimmung zwischen den beiden Punktwolken wieder. Eine kleine Hausdorffdistanz entspricht einer hohen Übereinstimmung. Die Hausdorffdistanz ist nicht zu verwechseln mit der euklidischen Distanz. Sie ist lediglich ein Gütemaß für die Übereinstimmung zweier Objekte. Der räumliche Abstand hingegen wird bestimmt durch die Verschiebung, die die Punktwolke  $A(\vec{W}_A)$  aus dem Bild  $f_A(x_A, y_A, t_A)$  in dem durch das Bild  $f_B(x_A + u, y_A + v, t_B)$  erzeugten Raum erfährt, damit die Hausdorffdistanz minimal wird. Die Geschwindigkeit  $\vec{v}_{HD}$  ergibt sich aus der Punktwolkenverschiebung und der Zeit, die zwischen zwei Bildaufnahmen verstrichen ist:

$$\vec{v}_{HD} = \frac{\vec{s}_{AB}}{t_{AB}},$$

mit

$$\vec{s}_{AB} = A(\vec{W}_A) - B(\vec{W}_B),$$

$$t_{AB} = t_A - t_B.$$

Da mit dieser Methode ein Verfahren zur Verfügung steht, das eine recht präzise Bestimmung der Objektverschiebung im dreidimensionalen Raum bei Anwesenheit von Störpunkten erlaubt, wird dieser Weg in dieser Arbeit als Objekterkennungsverfahren angewendet.

# Kapitel 3

## Positionsprädiktion

Die Positionsprädiktion soll den Ort eines bewegten Objekts zu dem Zeitpunkt schätzen, an dem das Objekt von einem 7-Freiheitsgrade-Manipulator gegriffen werden wird. Es ist zum einen die gegenwärtige Position des bewegten Objekts zum Zeitpunkt  $t$  gesucht, wobei  $t$  die Zeit zwischen zwei aufeinander folgenden Bildern ist. Zum anderen ist die Position des Objekts zum Zeitpunkt  $kt$  zu bestimmen, wobei  $k = 1, 2, \dots, n$  das Bild angibt, für das die Prädiktion erfolgen soll,  $k = 1$  ist das momentan betrachtete Bild. Die Verschiebung des Objekts lässt sich beispielsweise erreichen durch die Bestimmung des optischen Flusses (siehe 2.5.1), diverse Matchingverfahren (siehe 2.5.2) und durch die Hausdorffdistanz (siehe 2.5.3).

Die Bestimmung der gegenwärtigen Objektposition ist aufgrund der verhältnismäßig großen Entfernung zur Kamera nicht einfach, da das Objekt klein auf die Bildebene abgebildet wird. Die Disparität nimmt mit zunehmender Entfernung exponentiell ab. Da die Disparitätswerte bei diesen Verfahren ganzzahlig sind, ergibt sich daraus eine weitere Ungenauigkeit für die Positionsbestimmung. Für die gegenwärtige Objektposition bietet sich daher eine Näherung mit der Bildung eines geeigneten Konfidenzmaßes an.

### 3.1 Prädiktion der Objektposition

Die Objektposition zu einem späteren Zeitpunkt resultiert aus der momentanen Objektposition und der Objektverschiebung bzw. der Objektgeschwindigkeit. Hieraus wird deutlich, dass die Bestimmung der Objektposition entscheidend ist für die Qualität der Positionsvorhersage des Objekts. Ist die Positionsbestimmung nicht präzise genug, dann ist die Positionsprädiktion nur bedingt ausreichend.

Da das Greifobjekt aufgrund seiner Entfernung zum Kamerapaar klein auf

die Bildebene abgebildet wird, lassen sich nur wenig korrespondierende Pixel finden. Die Qualität der Korrespondenzsuche hat allerdings großen Einfluss auf die Güte der Positionsbestimmung. Folglich muss die Korrespondenzsuche einer erfolgreichen Ortsbestimmung angemessen sein (siehe Abschnitt 2.4.1).

Die korrespondierenden Punkte liefern die Raumkoordinaten für das Objekt (siehe Abschnitt 2.4.3). Aus diesen Raumkoordinaten lässt sich der Mittelpunkt des Objekts bestimmen. Der Mittelpunkt soll der Greifpunkt sein. Bei einer geringen Anzahl korrespondierender Pixel lassen sich nur wenige Raumkoordinaten bestimmen. Dadurch wird der Greifpunkt unscharf. Hinzu kommen verfälschte Raumkoordinaten, die sich aus fehlerhaft ermittelten korrespondierenden Pixeln ergeben haben (siehe Abschnitt 2.4.1). Diese fehlerhaften Raumkoordinaten dürfen bei der Bestimmung der Objektposition nicht berücksichtigt werden. Ein geeignetes Verfahren hierfür wird im folgenden Abschnitt vorgestellt.

### 3.1.1 Bestimmung der Objektposition

Das hier verwendete Verfahren zur Objektpositionsbestimmung beruht darauf, dass das Objekt klein genug ist, um ein Greifen zu ermöglichen. Zunächst wird die vorliegende dreidimensionale Punktwolke unterabgetastet (siehe Abb. 3.1). Aus der Unterabtastung ergeben sich  $n_x * n_y * n_z$  Quader, wobei  $n_i$ , mit  $i = x, y, z$ , die Anzahl der Quader in  $x$ -,  $y$ - und  $z$ -Richtung angibt. Ermittelt wird der Quader, der die maximale Anzahl an Raumkoordinaten enthält. Nur solche Koordinaten, die in einem bestimmten Bereich um diesen Quader herum liegen, gehen in die Bestimmung des Objektmittelpunktes mit ein. Der ausgewählte Bereich stellt näherungsweise eine definierte Objektgeometrie dar, die ein Greifen ermöglicht. Auf diese Weise lassen sich stark fehlerhaft bestimmte Raumkoordinaten ausschließen, die weiter entfernt von der Objektposition liegen, als das Objekt groß ist. Eine zu hohe Verzerrung des Objektmittelpunktes ist somit ausgeschlossen.

Für die Bestimmung des Objektmittelpunktes gehen weiter nur die Quader in die Berechnung mit ein, die mehr als einen Raumpunkt enthalten, um einzelne Störpunkte unberücksichtigt zu lassen (siehe Abb. 3.2). Damit ergibt sich für die Mittelpunktskoordinate  $\vec{P}(x, y, z)$  des Objekts:

$$\vec{P}(x, y, z) = \frac{1}{n} \sum_{i=1}^n \vec{P}_i,$$

$$\forall \vec{P}_i \in [\vec{P}_{max} - \vec{O}(x, y, z), \vec{P}_{max} + \vec{O}(x, y, z)] \wedge N(K_i) > 1.$$

Hierbei bezeichnet  $n$  die Anzahl der Raumpunkte,  $\vec{P}_{max}$  den Häufungs-

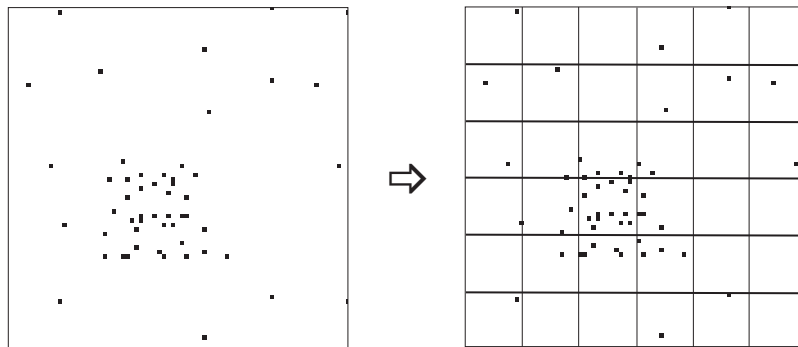


Abbildung 3.1: Grobrasterung des Raumes (aus Gründen der Vereinfachung nur zweidimensional dargestellt): Die Punkte stellen Raumpunkte des Objekts dar.

punkt bzw. den Quader mit der maximalen Anzahl an Raumkoordinaten,  $\vec{O}(x, y, z)$  die definierte Ausdehnung des Objekts und  $N(K_i)$  die Anzahl an Raumpunkten, die der Quader  $K_i$  enthält.

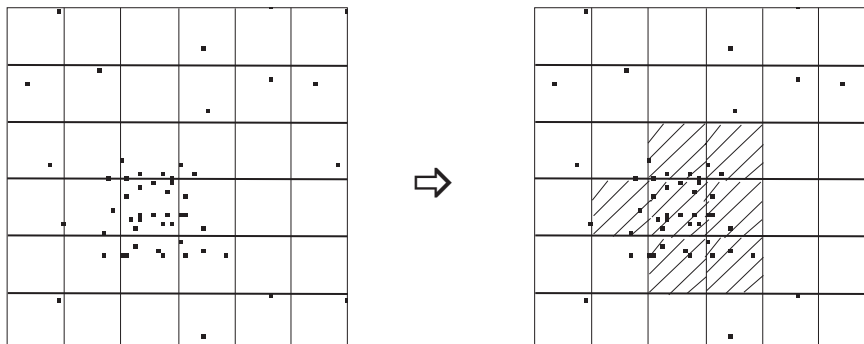


Abbildung 3.2: Nur die schraffierten Flächen werden bei der Objektmittelpunktsberechnung berücksichtigt.

Aus Kombination des Objektmittelpunktes  $\vec{P}(x, y, z, t = 0)$  und der errechneten Geschwindigkeit mittels Hausdorffdistanz  $\vec{v}_{HD}$  (siehe 2.5.4) lässt sich die Objektposition  $\vec{P}_{pred}(x, y, z, t)$  zu allen späteren Zeitpunkten  $t$  vorhersagen:

$$\vec{P}_{pred}(x, y, z, t) = \vec{P}(x, y, z, t = 0) + \vec{v}_{HD}(x, y, z)t \quad (3.1)$$

Hiermit ist abzuschätzen, zu welcher Zeit  $t_{greif}$  das Objekt den greifbaren Bereich erreicht haben wird. Ist die Geschwindigkeit  $\vec{v}_{arm}$  des Roboterarms

groß genug, um den Ort  $\vec{s}_{greif}$  innerhalb der Zeit  $t_{greif}$  zu erreichen, kann der Greifvorgang erfolgen. Für den Fall, dass die Roboterarmgeschwindigkeit nicht ausreicht, müssen eine Zeit und ein Ort bestimmt werden, zu denen das Objekt durch den Roboterarm greifbar wird.

## 3.2 Bestimmung einer Greifposition

Mit Gleichung 3.1 ist der Ort eines Objekts für jeden Zeitpunkt bekannt. Bevor das Greifen eingeleitet wird, bleibt zu untersuchen, ob der Manipulator genügend Zeit hat, einen gegebenen Greifort anzufahren. Daraus ergeben sich als Bedingungen, dass der Greifort innerhalb des durch den Manipulator greifbaren Bereiches liegt. Ferner darf die Zeitspanne, die der Manipulator benötigt, um den Greifort anzufahren, einen bestimmten Wert nicht überschreiten. Dies ist die Zeit, in der das Objekt den Greifort erreicht haben wird (time-to-contact).

In dieser Arbeit ergeben sich für die Bestimmung des Greifortes daraus folgende zwei Ansätze:

- Der Greifvorgang erfolgt möglichst bald, nachdem das Objekt den greifbaren Bereich erreicht hat;
- Der Greifvorgang erfolgt, sobald das Objekt einen bestimmten Abstand vom Roboterarm hat;

### Greifen erfolgt bei Eintritt in den greifbaren Bereich

In dieser Variante wird die Objektposition für jedes Bild prädiziert, bis eine prädizierte Position mit einer Position des greifbaren Bereichs übereinstimmt (siehe Abb. 3.3). Für diese Position wird errechnet, wieviel Zeit das Objekt benötigt, um diesen Ort zu erreichen (Greifzeit oder *time-to-contact*). Dann wird überprüft, ob der Greifarm diese Stelle (Greifort oder *time-to-contact-Ort*) innerhalb der Greifzeit anfahren kann. Ist die Überprüfung erfolgreich, d.h. der Greifarm kann den Greifort innerhalb der geforderten Greifzeit erreichen, wird der Greifvorgang eingeleitet. Ist die Überprüfung nicht erfolgreich, erfolgt eine Überprüfung, ob das Objekt im Folgebild im Greifbereich liegen wird. Anschließend wird analysiert, ob der Greifarm diese neue Position innerhalb der neuen Greifzeit erreichen kann. Dies wird solange fortgeführt, bis entweder das Objekt gegriffen werden kann oder bis das Objekt den greifbaren Bereich wieder verlassen hat. Im letzteren Fall ist ein Greifen nicht möglich.

### **Greifen erfolgt bei bestimmtem Abstand zum Roboterarm**

Diese Variante prädiziert für jedes Bild die Objektposition. Das Greifen soll erfolgen, wenn der Abstand zwischen Objekt und Roboter einen Minimalwert erreicht haben wird. Dies ist vorzugsweise ein Ort, der möglichst nahe am Roboter sein soll, damit die Entfernung, die der Roboterarm bewältigen muss, möglichst gering wird.

Für jedes Bild erfolgt eine Abstandsermittlung zwischen Objektposition und Greiferposition. Sobald die Abstände ansteigen, hat das Objekt den minimalen Abstand zum Greifer durchlaufen. Dies bedeutet, dass sich das Objekt vom Roboter entfernt. An diesem Punkt ergeben sich zwei Möglichkeiten, das Verfahren fortzusetzen. In der ersten Möglichkeit ist der minimale Abstand zum Manipulator der Greifort. In diesem Fall besteht die Gefahr, dass die prädizierte Objektposition mit dem Roboter kollidiert, wo ein Greifen nicht möglich ist. Hier befindet sich der Roboter selbst, stellt also ein Hindernis dar. Außerdem könnte der minimale Abstand so dicht am Roboter sein, dass das Greifen für den Roboter erschwert wird, was ein schnelles Greifen verhindert.

Aus diesem Grund bietet sich eine zweite Möglichkeit an, bei der der Abstand des Greifortes zum Manipulator zwar minimiert sein soll, der Greifort aber eine festgesetzte Mindestnähe zum Greifer haben muss („definierter Greifraum“) (siehe Abb. 3.4). Sobald eine prädizierte Objektposition mit dieser Mindestnähe übereinstimmt, wird diese Position als Greifort festgelegt. Wird die Mindestnähe unterschritten, gilt die letzte ermittelte prädizierte Position, die die Mindestnähe nicht unterschritten hat, als Greifort gewählt. Ist der Greifpunkt bestimmt, wird die Greifzeit geschätzt, also die Zeit, in der das Objekt den geschätzten Greifort vermutlich erreicht haben wird. Anschließend erfolgt eine Überprüfung, ob der Greifarm diese Position innerhalb der time-to-contact erreichen kann. Ist dies der Fall, wird der Greifvorgang eingeleitet. Im anderen Fall erfolgt eine Überprüfung, ob das Objekt für die nächste prädizierte Position innerhalb der Greifzeit für den Greifer erreichbar ist. Voraussetzung hierfür ist eine Greifposition, die im definierten Greifraum liegt. Diese Kontrolle passiert solange, bis das Objekt gegriffen werden kann oder bis das Objekt den greifbaren Bereich verlassen hat. Im letzteren Fall ist ein Greifen nicht möglich.

Der Vorteil der ersten Variante liegt darin, dass der Greifpunkt schneller bestimmt sein kann. Dadurch kann der Greifpunkt des nächsten zu greifenden Objekts früh bestimmt werden. Der Nachteil der ersten Variante ist, dass der Greifarm unter Umständen einen weiten Weg zurückzulegen hat. Dies hat eine lange Greifzeit zur Folge und nachfolgend zu greifende Objekte werden

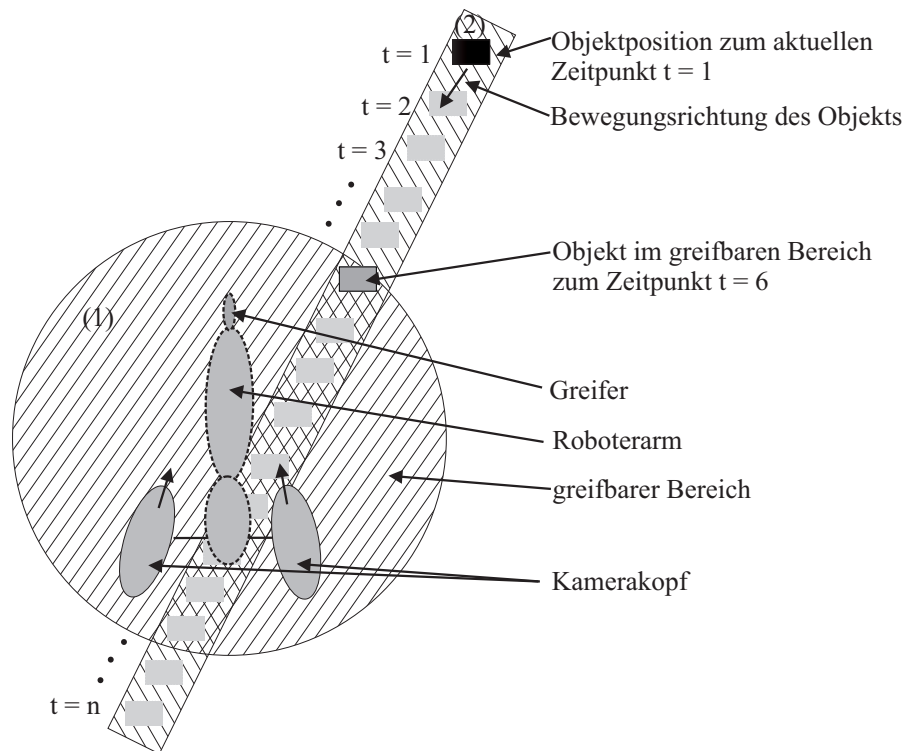


Abbildung 3.3: Schematische Ansicht der Greifszene von oben: Der Greifarm kann Objekte aufgrund seiner begrenzten Reichweite nur innerhalb des schraffierten Kreises greifen. Die Bahn des Objekts ist durch den rechteckigen schraffierten Bereich (2) gekennzeichnet. Das Objekt selbst ist durch das schwarze Rechteck oben markiert. Die späteren Positionen des Objekts zum Zeitpunkt  $t=2$ ,  $t=3$ , ...,  $t=n$  sind durch die darunter liegenden helleren Rechtecke angedeutet. In diesem Beispiel erreicht das Objekt den greifbaren Bereich zum Zeitpunkt  $t=6$  (dunkelgraues, schwarz umrandetes Rechteck). Der Ort, den das Objekt zu diesem Zeitpunkt erreicht haben wird, ist der Greifort, bzw. der „time-to-contact-Ort“.

eventuell nicht mehr greifbar.

Der Vorteil der zweiten Variante liegt in der Greifgeschwindigkeit, sofern sich der Greifort ausreichend entfernt vom Roboter befindet. Das Greifen ist meistens schnell abgeschlossen, wenn der Abstand vom ruhenden Manipulator zum Greifpunkt klein gehalten werden kann. Dadurch ist der Greifarm nach kurzer Zeit wieder einsatzbereit. Dies erhöht die Wahrscheinlichkeit des erfolgreichen Greifens auch von dicht beieinander liegenden Objekten und wird daher bei dieser Arbeit der ersten Variante vorgezogen.

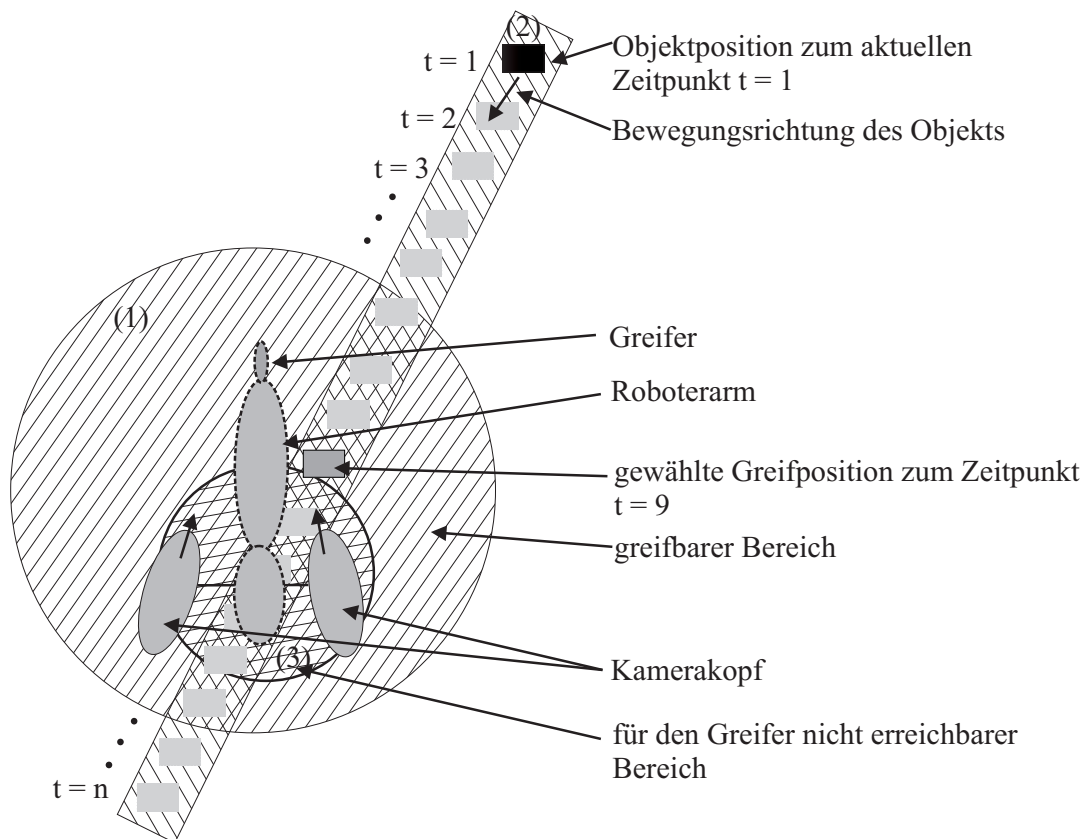


Abbildung 3.4: Schematische Ansicht der Greifszene von oben: Anordnung wie in Abb. 3.3 mit dem Unterschied, dass das Objekt dann gegriffen wird, wenn es eine Mindestnähe zum Roboterarm erreicht hat. Der Bereich, in dem das Objekt aufgrund zu großer Nähe zum Roboterarm nicht gegriffen werden kann bzw. nicht gegriffen werden soll, ist durch den schraffierten Kreis (3) innerhalb des greifbaren Bereichs (1) gekennzeichnet. In diesem Beispiel erreicht das Objekt den Bereich (3) zum Zeitpunkt  $t=9$ . Der Ort, den das Objekt zu diesem Zeitpunkt erreicht haben wird, ist der Greifort, bzw. der „time-to-contact-Ort“.



# Kapitel 4

## CoRA

### 4.1 Autonomes Assistenzsystem CoRA

Im Rahmen des bmb+f-Leitprojektes „Morpha“ wird an der Ruhr-Universität Bochum der autonome Assistenzroboter CoRA (siehe Abb. 4.1) entwickelt, der mit seinem Roboterarm interaktiv in einen Produktionsprozess mit menschlicher Beteiligung eingreifen kann. CoRA soll sich am Ende eines Förderbandes befinden und Gegenstände von diesem Förderband greifen, um diese interaktiv mit seinem menschlichem Partner weiter zu verwerten.

#### **CoRAs Aufbau**

Dem Roboter steht als Sensor ein schwenk- und neigbares Stereokamerasystem zur Verfügung. Es dient dazu, den Arbeitsbereich des Roboters zu überblicken und den Menschen und dessen Aktionen zu verfolgen. Die beiden Kameras bleiben bei Schwenk- oder Neigebewegungen parallel zueinander. CoRA ist mit einem redundanten Arm mit sieben Freiheitsgraden versehen, der sich an einem drehbaren Rumpf befindet. Da die Anordnung der Armgelenke der des menschlichen Armes nachempfunden ist, ermöglicht dies ein Greifen von Objekten unter Vermeidung von Kollisionen mit Hindernissen. Der Greifer ist mit einer taktilen Kraft-Momenten-Sensorik ausgestattet. Durch den drehbaren Rumpf wird der Greifraum erweitert. Dies gestattet eine einfache Zusammenarbeit mit dem Menschen. An dem Roboter ist eine 90x90 cm große Arbeitsfläche montiert.

#### **Kommunikation**

CoRA verfügt zur sprachlichen Kommunikation über einen Einzelworterkenner und eine Sprachausgabe. Der Roboter kann Objekte einer komplexen

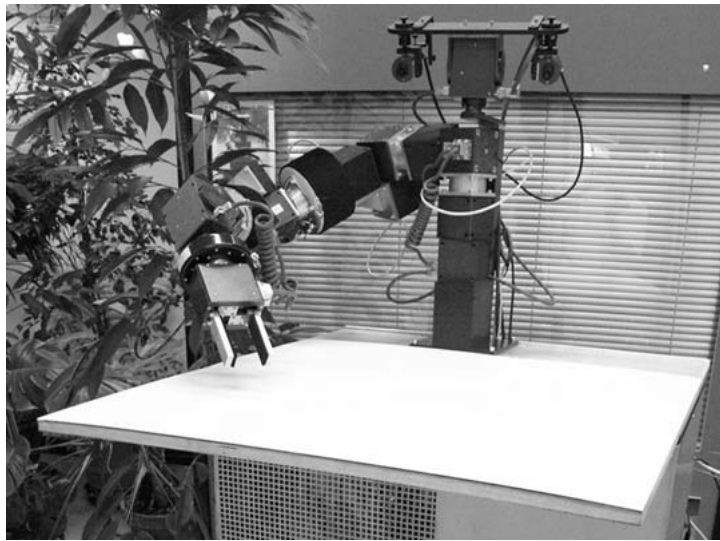


Abbildung 4.1: Das autonome Assistenzsystem CoRA.

Tischszene unterscheiden und gelernte Objekte wiedererkennen. Bei dieser Aufgabe werden der Ort und die Orientierung des Objekts erkannt. Dadurch ist es dem Roboter möglich, Objekte mit seinem Arm zu greifen. Über das Stereokamerapaar besitzt CoRA die Fähigkeit, die menschliche Hand zu finden und eine Zeigerichtung des Benutzers zu erkennen. Mittels einer Blickrichtungsanalyse kann der Roboter auf Bereiche schließen, denen der Benutzer seine Aufmerksamkeit zuwendet. Liegt ein Hindernis nicht im Sichtbereich des Stereokamerapaars, so wird eine Kollision mit dem Hindernis durch eine künstliche Haut vermieden (siehe Abb. 4.2). So kann CoRA Kontakt wahrnehmen, was auch dazu dient, die Bewegung des Arms durch menschliches Einwirken zu verbessern oder zu korrigieren.

### **Neuronale Strukturen zur Verhaltenserzeugung**

Um mit dem autonomen Assistenzsystem interagieren zu können, ist neben dem kommunikativen Informationsaustausch die Planung von Verhaltenssequenzen nötig, sowie deren Koordinierung und eine Adaption an veränderliche Randbedingungen.

Bei CoRA sind diese Maßnahmen durch neuronale dynamische Systeme verwirklicht, die in einer Feedbackstruktur aktive Verhaltensweisen erzeugen. Dabei werden sich ausschließende Verhalten durch Regeln kodiert und zu einer intelligenten Einbindung von Sprache verholfen. Kausale Relationen können von einer unterliegenden neuronalen Struktur gelernt werden, die

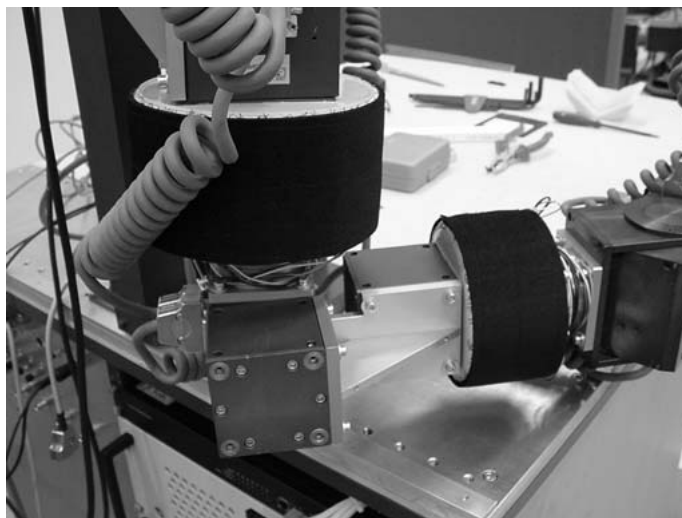


Abbildung 4.2: Die künstliche Haut besteht aus zwei schwarzen zylinderförmigen Manschetten.

außerdem Verhaltensplanungen adaptieren kann.

## 4.2 Bildaufnahme

Die in dieser Arbeit verwendeten Bilder besitzen eine Auflösung von  $700 \times 512$  Pixeln und wurden mit einem stationären 2-Freiheitsgrade-Stereokamerasystem aufgenommen, dessen Bildverarbeitung auf dem HSI-Farbraum basiert. Der HSI-Farbraum lässt sich mit Hilfe eines Kegels veranschaulichen, dessen Kegelhöhe die Intensität (engl. intensity) angibt, der Radius bzw. der Abstand zur Intensitätsachse beschreibt die Sättigung (engl. saturation) und der Farbton (engl. hue) wird durch einen Winkel um die Intensitätsachse bestimmt, der zwischen  $0^\circ$  und  $360^\circ$  liegt (siehe Abb. 4.3). Die Intensität kann wie die Sättigung Werte zwischen 0% und 100% annehmen. Im Falle der Intensität markiert 0% die Spitze des Kegels. Eine hohe Sättigung entspricht einem geringen Weißanteil und bewirkt kräftige Farben. Umgekehrt können bei einer geringen Sättigung die Farben aufgrund ihres hohen Weißanteils kaum noch unterschieden werden.

Das ppm-Format (*portable pixmap*) in abgewandelter Form dient hier als Bildeingabeformat der Farbbilder. Diese Farbbilder werden in Grauwertbilder umgewandelt, da eine Reduzierung der Datenmenge erforderlich ist und Grauwertbilder sich für die Verarbeitung gut eignen. Die Bildausgabe erfolgt im pgm-Format (*portable greymap*).

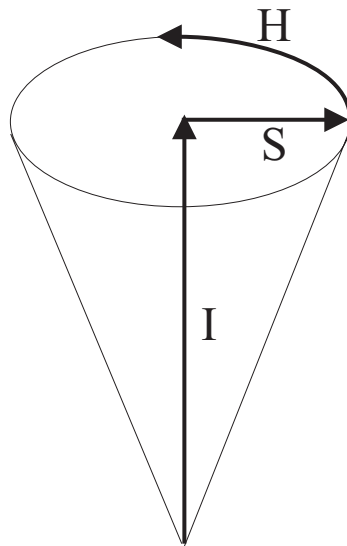


Abbildung 4.3: Der HSI-Farbraum. Der Farbraum ergibt sich aus der Intensität  $I$  (engl. intensity), der Sättigung  $S$  (engl. saturation) und dem Farbton  $H$  (engl. hue).

# Kapitel 5

## Ergebnisse

In diesem Kapitel wird ein Softwaremodul vorgestellt, das auf den in Kapitel 2 und 3 beschriebenen Verfahren basiert. Die schematische Darstellung des Softwaremoduls ist in Abb. 5.1 skizziert. Die Aufgabe des Softwaremoduls ist es, die Position eines Gegenstands zu präzisieren, um diesen aus der Bewegung zu greifen. Ferner wird ein *time-to-contact-Ort* bestimmt, zu dem ein Greifarm das Objekt fassen kann.

Zunächst liest das Softwaremodul die Bilddaten zweier Kameras ein. Anschließend erfolgt eine Umrechnung der Bilddaten in Weltkoordinaten, aus denen eine Bestimmung der Objektverschiebung per Hausdorffdistanz gewonnen wird. Hierüber lässt sich eine Objektposition präzisieren und eine Greifposition schätzen. Die für diese Verfahren verwendeten Algorithmen werden in der Behandlung von Beispielsequenzen angegeben.

Die Eignung dieses Verfahrens wird im Folgenden anhand von Beispielsequenzen mit unterschiedlichen Objekten diskutiert. Die Beispielsequenzen unterscheiden sich ferner dadurch, dass die Objekte in ihrer Orientierung, ihrer Bewegungsrichtung und ihrer Entfernung zum Roboter variieren.

Die verwendeten Szenen zeigen Sequenzen, in denen sich ein einzelnes Objekt gleichförmig, d.h. mit konstanter Geschwindigkeit, bewegt. Die zu verfolgenden Objekte in den Beispielsequenzen haben jeweils eine resultierende Geschwindigkeit von ca.  $20 \frac{mm}{Bild}$ . Der Hintergrund bleibt ebenso wie die Umgebung (Helligkeit etc.) unverändert.

### 5.1 Erste Beispielsequenz

Die Bestimmung der Bewegungsrichtung und der Geschwindigkeit des Objekts beginnt mit der Erfassung der Szene durch das Stereokamerapaar (siehe Abb. 5.2).

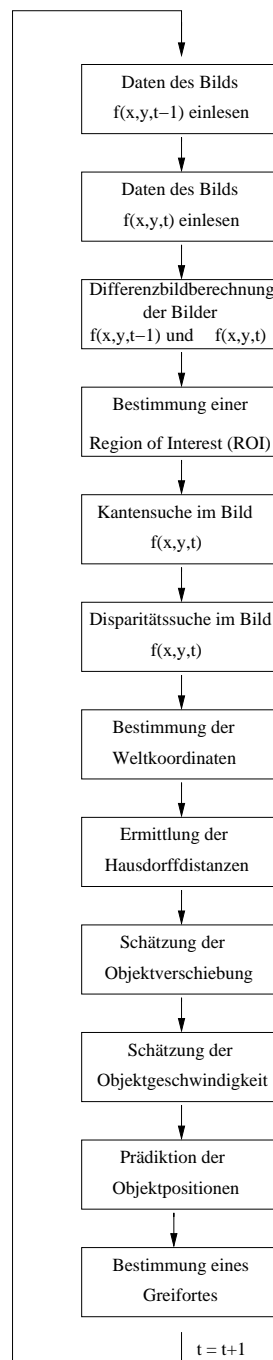
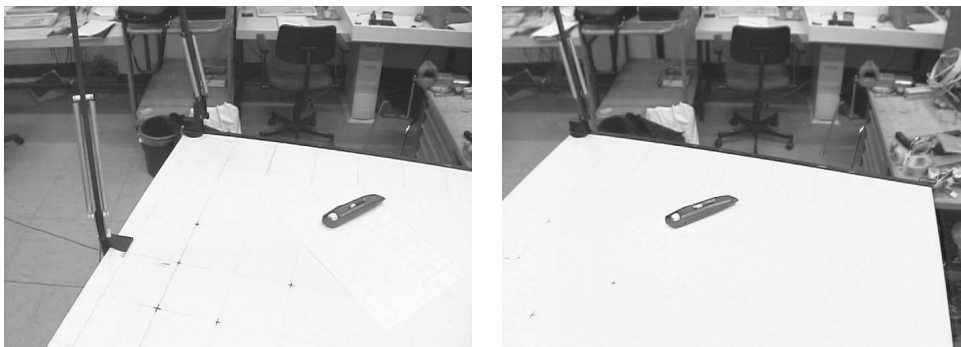


Abbildung 5.1: Schematische Darstellung der Objektverfolgung und der Greifpositionsbestimmung.

Abbildung 5.2: Linkes und rechtes Kamerabild zum Zeitpunkt  $t-1$ .Abbildung 5.3: Linkes und rechtes Kamerabild zum Zeitpunkt  $t$ .

Für die weitere Verarbeitung werden unwichtige Bereiche von den interessierenden Bereichen getrennt. Dies erfolgt über die Bildung einer *Region of Interest* (siehe Kapitel 2.2), die hier aus einem Differenzbild gewonnen wird (siehe Abb. 5.4). Das Differenzbild ergibt sich aus zwei Kamerabildern, die zu unterschiedlichen Zeitpunkten aufgenommen wurden und einen gleichen Bildbereich zeigen (vgl. Kapitel 2.1).

Das Clustern des Differenzbildes (siehe Kapitel 2.2) ermöglicht das Beseitigen von Bildstörungen (siehe Abb. 5.5), die die Bildung einer Region of Interest erschweren würden.

Durch die Bildung der Region of Interest verringert sich der zu betrachtende Bereich des Bildes, wodurch weniger Bildpunkte bearbeitet werden müssen.

Innerhalb der Region of Interest erfolgt eine Kantensuche (vgl. Kapitel 2.3). Um den Hintergrund von der Kantensuche auszuschließen, findet eine Kantensuche nur statt, wenn ein Pixel im Differenzbild gesetzt ist. Dies ge-



Abbildung 5.4: Die Differenzbilder aus Abb. 5.2 und Abb. 5.3.

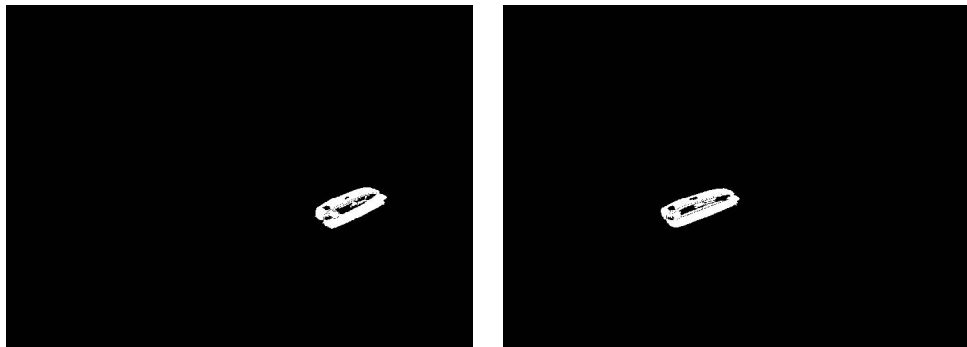


Abbildung 5.5: Die Differenzbilder nach der Entfernung von Störpixeln.

währleistet, dass nur Kanten, die zum Objekt gehören, gesucht werden (siehe Abb. 5.7).

In Abbildung 5.6 ist der Unterschied zwischen Kantenbildern zu erkennen, die mit dem eindimensionalen Differenzoperator aus Abschnitt 2.3.1 angefertigt wurden und von Kantenbildern, die sich bei Anwendung des in Abschnitt 2.3.2 beschriebenen Prewitt-Kantenoperators ergeben. Das Kantenbild, das bei Anwendung des Prewitt-Operators entsteht, weist weniger einzelne Störpixel auf. Die Anzahl der Störpixel in den Kantenbildern, die aus dem eindimensionalen Differenzoperator hervorgehen, sind allerdings recht gering, so dass der Einfluss von Störpixeln auf die Disparitätsbestimmung klein genug ist, aber in der weiteren Verarbeitung mit einbezogen werden muss. Daher bietet sich der einfachere eindimensionale Differenzoperator als Kantenoperator für das hier verwendete Softwaremodul an.

Die Korrespondenzsuche und die daran anschließende Disparitätsbestimmung erfolgt, wie in Abschnitt 2.4.2 beschrieben. Ein Disparitätsbild, das

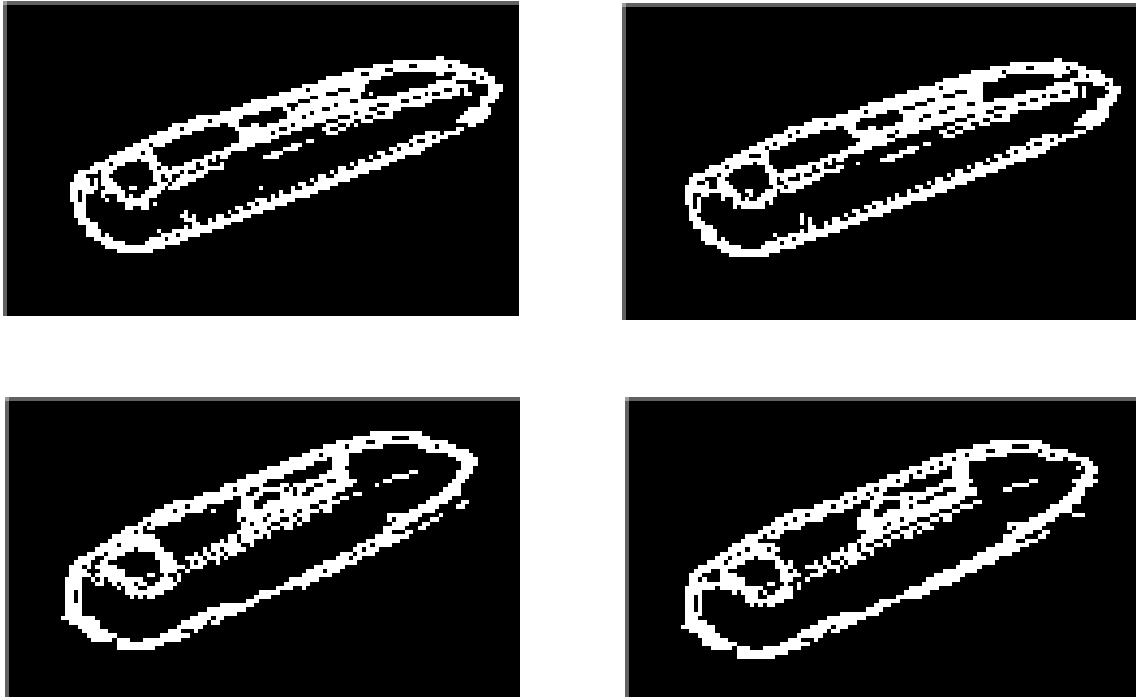


Abbildung 5.6: Ein Vergleich von Kantenbildern, die sich aus dem eindimensionalen Differenzoperator ergeben (jeweils linke Seite) mit Kantenbildern, die bei Anwendung des Prewitt-Kantenoperators entstehen (jeweils rechte Seite). Die oberen Bildern sind die Kantenbilder der linken Kamera und die unteren Bilder sind die Kantenbilder der rechten Kamera.

sich daraus für diese Sequenz ergibt, ist in Abbildung 5.8 zu sehen.

Aus den Disparitätswerten erhält man eine dreidimensionale Punktwolke. Diese Punktwolke entspricht einer dreidimensionalen Rekonstruktion des Objekts und gibt die Entfernung zum Roboter an.

Die Daten der Rekonstruktion werden gespeichert und es erfolgt die Aufnahme der nächsten zwei Kamerabilder (siehe Abb. 5.9). Aus diesen zwei Kamerabildern wird analog zum oben beschriebenen Verfahren eine Rekonstruktion der vorliegenden Szene ermittelt.

Aus dem Vergleich der Objektpositionen resultiert ein Flussfeld, aus dem die Geschwindigkeit und die Bewegungsrichtung des Objekts hervorgeht. Hierfür wird die Hausdorffdistanz verwendet. In diesem Fall wird die Objektverschiebung wie in Abschnitt 2.5.4 beschrieben ermittelt. Es wird die Punktwolke  $A(\vec{W}_A)$  aus dem letzten Bild über die Punktwolke  $B(\vec{W}_B)$  des aktuellen Bildes bewegt. Die Verschiebung von  $A(\vec{W}_A)$  erfolgt gemäß Glei-

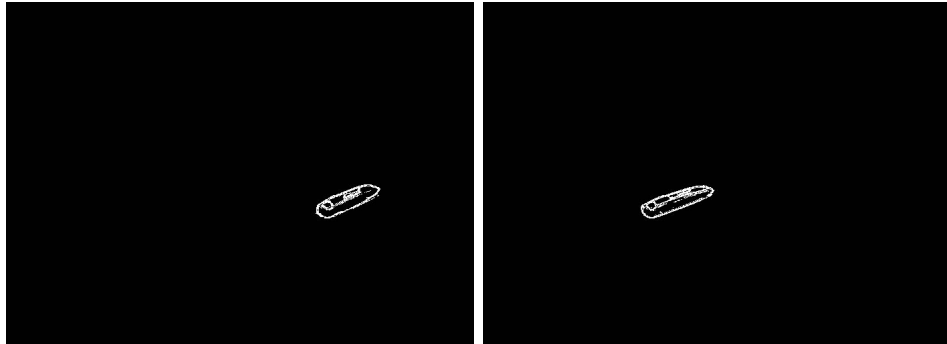


Abbildung 5.7: Die erzeugten Kantenbilder für die Grauwertbilder zum Zeitpunkt  $t-1$ .

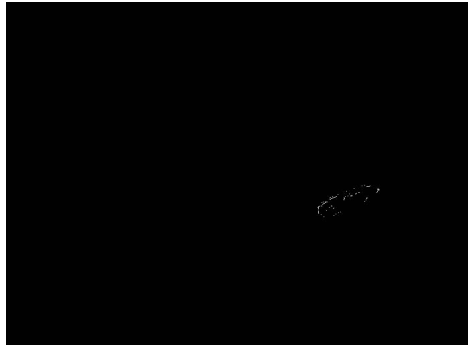


Abbildung 5.8: Das Disparitätsbild für die Abb. 5.3

chung 5.1

$$A'(\vec{W}_{A'}) = A(\vec{W}_A + \begin{pmatrix} 10 \\ 10 \\ 4 \end{pmatrix} \begin{pmatrix} i \\ j \\ k \end{pmatrix}), \quad (5.1)$$

mit

$$\begin{aligned} i &= -3, -2, \dots, 2, 3 \\ j &= -3, -2, \dots, 2, 3 \\ k &= -2.5, -1.5, \dots, 1.5, 2.5 \end{aligned} .$$

Diese Gleichung ist folgendermaßen zu interpretieren: nacheinander wird jeweils  $i$ , hieran  $j$  und zum Schluss  $k$  um eins erhöht.  $i$  wird von  $-3$  bis  $+3$



Abbildung 5.9: Auf Abb. 5.3 folgendes linkes und rechtes Kamerabild.

um eins erhöht und für jede Iteration die Hausdorffdistanz zur Punktwolke  $B(\vec{W}_B)$  bestimmt. Anschließend wird  $j$  um eins erhöht und  $i$  durchläuft erneut die Werte von  $-3$  bis  $+3$ , wobei für jede Iteration die Hausdorffdistanz bestimmt wird. Dies geschieht solange, bis  $j$  alle Werte von  $-3$  bis  $+3$  durchlaufen hat. Das nachfolgende Erhöhen von  $k$  verläuft analog.

Es ergeben sich hiernach  $7 \times 7 \times 6 = 294$  Positionen für die Punktwolke  $A(\vec{W}_A)$ . An jeder dieser Positionen findet eine Berechnung der Hausdorffdistanz zur Punktwolke  $B(\vec{W}_B)$  statt.

Analog erfolgt derselbe Prozess für die Punktwolke  $B(\vec{W}_B)$ , die dann in dem Raum aus dem vorherigen Bild verrückt wird. Die Bestimmung der Hausdorffdistanz erfolgt hierfür analog. Entsprechend Abschnitt 2.5.4 ergibt sich hieraus die Objektverschiebung.

Die Gleichung 5.1 hat eine Einschränkung der Bestimmungsgenauigkeit zur Folge. Die Objektverschiebung für die horizontale Bewegung und für die Bewegung in der Tiefe sind auf  $10$  mm begrenzt und auf  $5$  mm für die Höhe. Ferner bedeutet Gleichung 5.1, dass die Objektgeschwindigkeit im Intervall

$$\left[ (-30, -30, -10) \frac{\text{mm}}{\text{Bild}}, (30, 30, 10) \frac{\text{mm}}{\text{Bild}} \right]$$

liegen muss.

Für jedes Bild wird eine mittlere Verschiebung aus den vorhergegangenen Objektverschiebungen bestimmt, aus der eine mittlere Objektgeschwindigkeit und eine Bewegungsrichtung resultiert.

Aus der geschätzten Bewegungsrichtung und der geschätzten mittleren Geschwindigkeit lassen sich die Objektkoordinaten für jeden späteren Zeitpunkt präzisieren (siehe Abb. 5.10).

Aus diesen Daten wird ein Ort bestimmt, an dem ein Greifen des Objekts erfolgt. Das Bild 5.11 zeigt eine sehr gute Übereinstimmung des präzisierten



Abbildung 5.10: In dieser Abbildung ist die prädizierte Objektposition der real erreichten Objektposition anhand des Kantenbildes gegenübergestellt. Die prädizierten Pixel sind grau dargestellt, die Kantenpixel weiß. Es ist leicht zu erkennen, dass die beiden Orte gut übereinstimmen.

Greifortes mit dem real erreichten Greifort des Objekts. Der Großteil der prädizierten Bildpunkte liegt innerhalb des Gegenstands. Besonders deutlich wird, dass die prädizierten Bildpunkte sehr gut mit den äußeren Objektgrenzen übereinstimmen.



Abbildung 5.11: Gegenüberstellung des Kantenbildes der zehnten Bildaufnahme und der prädizierten Position für das Objekt.

Die Verfahren der folgenden Beispielsequenzen sind analog zu dem in diesem Abschnitt beschriebenen Verfahren.

## 5.2 Zweite Beispielsequenz: Objekt mit geringem Abstand zum Roboter

In dieser Sequenz bewegt sich ein Klebestift, zu erkennen in der Bildmitte auf der weißen Tischfläche, frontal auf den Roboter zu (siehe Abb. B.1). Der Mittelpunkt des Klebestifts befindet sich zu Beginn der Sequenz auf der weißen Arbeitsplatte des Roboters ca. 800 mm in  $y$ -Richtung und -300 mm in  $x$ -Richtung vom Koordinatenursprung des Roboters entfernt. Der Abstand des Objekts zum Roboter nach der zehnten Bildaufnahme beträgt 600 mm in  $y$ -Richtung und -300 mm in  $x$ -Richtung. Die Höhe des Klebestifts über der Arbeitsfläche beträgt ca. 50 mm. Die Geschwindigkeit des Objekts beträgt  $(0, 20, 0)^T \frac{\text{mm}}{\text{Bild}}$ .

Ein Ergebnis der Anwendung des eindimensionalen Differenzoperators auf die einzelnen Bilder dieser Sequenz ist in Abbildung 5.12 dargestellt. Man erkennt in den Bildern nur wenige Störpixel. Jedoch scheint der obere Teil des Klebestifts verformt zu sein. Am unteren Ende des Klebestifts ist weiter eine dickere Linie zu erkennen.

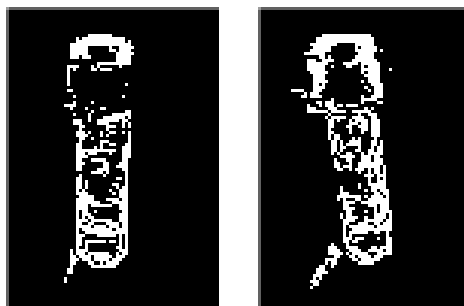


Abbildung 5.12: Ein vergrößertes Kantenbild der Sequenz, in dem unten links eine Linie zu sehen ist und die scheinbar zu breite Kappe des Klebestiftes.

Bei Vergrößerung des Ausschnitts des Grauwertbildes (siehe Abb. 5.13), sind Verwischungen im oberen Bereich des Klebestifts zu erkennen, die die Ursache für die scheinbare Verformung des Stiftes sind.

Eine Vergrößerung des Differenzbildes (siehe Abb. 5.14) und der zwei Grauwertbilder (siehe Abb. 5.15), aus denen das Differenzbild hervorgeht, liefert die Erklärung für die breite Linie am unteren Ende des Objekts. Der

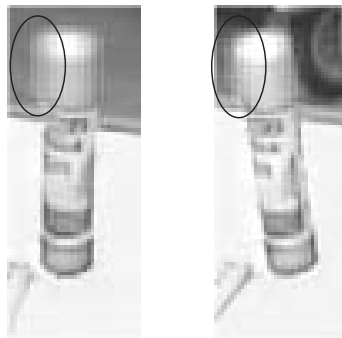


Abbildung 5.13: An den eingekreisten Stellen sind Verwischungen zu erkennen, die zu einer scheinbaren Verformung der Objektkonturen führen.

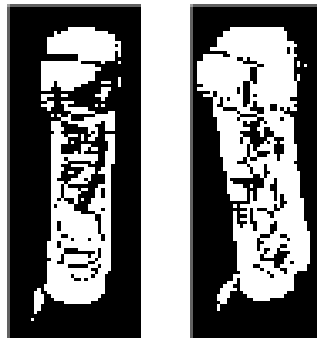


Abbildung 5.14: In dem Differenzbild ist zu erkennen, dass sich das Bild unten links am Klebestift verändert hat.

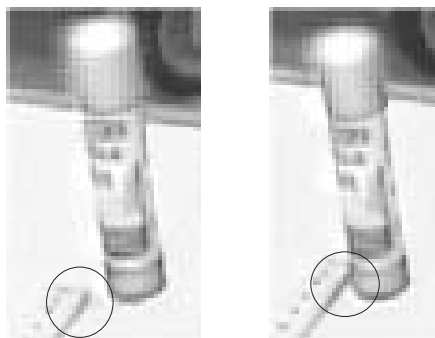


Abbildung 5.15: Eine Vergrößerung des Grauwertbildes zeigt, dass der Schatten des Klebestiftes die Veränderung im Bild hervorgerufen hat.

Schattenwurf des Klebestiftes auf den Maßstab hat eine Veränderung der Bildhelligkeit an dieser Stelle verursacht. Dadurch gibt es im Differenzbild eine Antwort an dieser Stelle. Der von dem Schatten getroffene Bereich ist zu groß, um als Störung aufgefasst zu werden und kann nicht durch Rauschunterdrückung (siehe Abschnitt 2.1) im Differenzbild entfernt werden. Folglich wird diese Stelle bei der Bestimmung der Objektkanten mit berücksichtigt.

Über die Hausdorffdistanz wird die Verschiebung des Objekts, wie in Abschnitt 2.5.4 beschrieben, ermittelt. Die Verschiebung liefert die Werte der einzelnen Geschwindigkeitskomponenten. Hierauf basierend werden die Objektpositionen in nachfolgenden Bildern geschätzt.

Trotz der Störungen im Kantenbild sieht man in Abbildung 5.16, dass sowohl der prädizierte Objektort gut mit der tatsächlich erreichten Position übereinstimmt als auch die prädizierte Greifposition mit der erreichten Greifposition. Durch den hohen Grad an Struktur auf dem Klebestift werden viele Kanten gefunden. Aus der reichhaltigen Struktur ergibt sich eine gute Disparitätsschätzung, was eine gute Bestimmung der jeweiligen Objektpositionen liefert.

Die Abweichung des prädizierten und des tatsächlich erreichten Ortes nach zehn Bildaufnahmen beträgt am Ende der Sequenz  $(5, -16, -8)mm$ . Die Ergebnisse der ermittelten Objektpositionen, der prädizierten Position für das nächste Bild und des *time-to-contact-Ortes* über die gesamte Sequenz sind in Tabelle A.1 festgehalten.

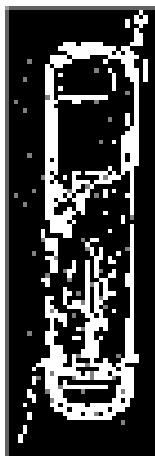


Abbildung 5.16: Diese Abbildung zeigt den Vergleich zwischen prädiziertem Objektort (graue Pixel) und des tatsächlich erreichten Objektorts (weiße Pixel). Die Orte stimmen gut überein. Nur wenige graue Pixel liegen neben den weißen Pixeln.

### 5.3 Dritte Beispielsequenz: Objekt mit geringem Abstand zum Roboter

Hier bewegt sich eine rote Kunststoffbox in der Bildmitte von links nach rechts über die Arbeitsfläche (siehe Abbildung B.2). Der Mittelpunkt der Kunststoffbox befindet sich zu Beginn der Sequenz  $-175$  mm in  $x$ -Richtung und  $760$  mm in  $y$ -Richtung vom Koordinatenursprung des Roboters entfernt. Am Sequenzzende befindet sich der Mittelpunkt der Box an den Koordinaten  $(25, 760, 15)^T$ . Der Mittelpunkt der Box liegt  $15$  mm über der Tischfläche. Die Geschwindigkeit der Box beträgt  $(20, 0, 0)^T \frac{\text{mm}}{\text{Bild}}$ .

Die hier verwendeten Verfahren sind analog zu den in der ersten und zweiten Beispielsequenz benutzten Verfahren (siehe Abschnitt 5.1).

Die Kanten der Kunststoffbox sind besser gefunden worden als die des Klebestiftes aus Abschnitt 5.2. Es sind fast ausschließlich Kanten des Objekts erkannt worden. Die Konturen der Plastikbox sind in dem Kantenbild deutlich erkennbar (siehe Abb. 5.17).



Abbildung 5.17: Die Kanten der Kunststoffbox wurden gut erkannt. Kanten, die nicht zum Objekt gehören, sind nicht zu sehen.

Die Plastikbox weist weniger Struktur auf als der Klebestift aus Abschnitt 5.2. Dementsprechend weniger Kanten werden gefunden. Da sich sehr viele Grauwerte ähnlich sind, erschwert dies eine erfolgreiche Disparitätssuche. Ein Vergleich eines Disparitätsbildes des Klebestiftes mit einem Disparitätsbild der Plastiksachtel zeigt, dass in dem Bild, in dem das Objekt mehr Struktur hat, bessere Disparitätswerte gefunden werden (siehe Abb. 5.18), obwohl der Klebestift und die Kunststoffschachtel ungefähr gleich groß sind.

Die Konsequenz ist, dass in dieser Sequenz die geschätzte Objektposition für das letzte Bild etwas neben dem tatsächlichen Objektort liegt (siehe Tabelle A.2 und Abb. 5.19). Die Abweichung des prädizierten Ortes und die real erreichte Objektposition nach zehn Bildaufnahmen beträgt  $(19, 65, -45)^T$  mm. Die Abweichung der  $y$ -Komponente von  $64$  mm und der  $z$ -Komponente von  $-45$  mm fällt recht hoch aus, ist aber aufgrund der genannten Gründe zu erwarten.



Abbildung 5.18: Das linke Bild zeigt ein Disparitätsbild der Sequenz aus Abschnitt 5.2. Das rechte Bild ist ein Disparitätsbild der in diesem Abschnitt betrachteten Sequenz. Obwohl die Objektkanten der Kunststoffbox eindeutig sind (siehe Abb. 5.17), fällt die Disparitätssuche aufgrund zu vieler ähnlicher Grauwerte schlechter aus als die Disparitätssuche in Abschnitt 5.2.

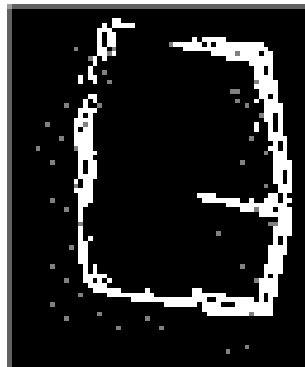


Abbildung 5.19: Der prädizierte Objektort (graue Pixel) für das zehnte Bild und der tatsächlich erreichte Ort (weiße Pixel). Der prädizierte Ort liegt deutlich neben dem realen Zielort.

## 5.4 Vierte Beispielsequenz: Objekt mit großem Abstand zum Roboter

In dieser Szene (siehe Abbildung B.3) hat das Objekt, ein Heftapparat, einen größeren Abstand zum Roboter als in den Sequenzen, die in 5.2 und in 5.3 besprochen wurden. Das Objekt befindet sich nun nicht auf der weißen Arbeitsfläche, sondern auf einem Rollwagen, der ungefähr 1,5 m vor dem Roboter steht. Der Heftapparat ist in der Bildmitte etwas oberhalb auf einem weißen Untergrund zu sehen. Die Bewegung verläuft von links nach rechts.

Anfangs befindet sich der Mittelpunkt des Heftapparats an den Koordinaten  $(370, 1970, -375)^T$ . In der letzten Bildaufnahme ist der Heftapparatmittelpunkt an den Koordinaten  $(545, 2020, -375)^T$  aufzufinden. Die Geschwindigkeitskomponenten betragen  $(19.4, 5.5, 0)^T \frac{\text{mm}}{\text{Bild}}$ .

Das Objekt selbst und die Kanten des Objekts werden trotz der kleinen Abbildung auf den Sensor gut gefunden (siehe Abb. 5.20).

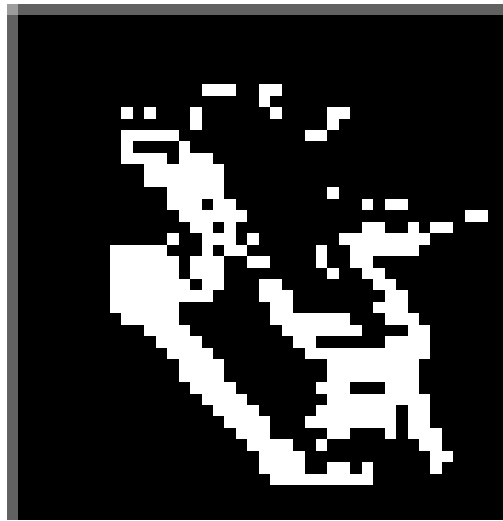


Abbildung 5.20: Das Kantenbild des Heftapparats. Rechts sind Kanten zu erkennen, die zur weißen Unterlage gehören. Die Ursache ist der Schattenwurf des Heftapparats, der eine Bildveränderung an der Stelle des Schattens hervorruft und bei der Differenzbildberechnung deutlich wird.

Die Kantenbilder und die Grauwertbilder enthalten genügend Informationen, um für die anschließende Disparitätssuche gute Werte zu liefern (siehe Abb. 5.21).

Für die darauf folgende Ermittlung der Objektverschiebung ergeben sich daher gute Werte. Daraus resultiert eine sehr gute Übereinstimmung für die Prädiktion und der Objektposition zum Zeitpunkt der zehnten Aufnahme (siehe Tabelle A.3 und Abb. 5.22). Die Abweichung des erreichten Ortes vom prädizierten Ort beträgt nur  $(-2, 0, 5)^T \text{mm}$  und ist damit als sehr gut einzuordnen.

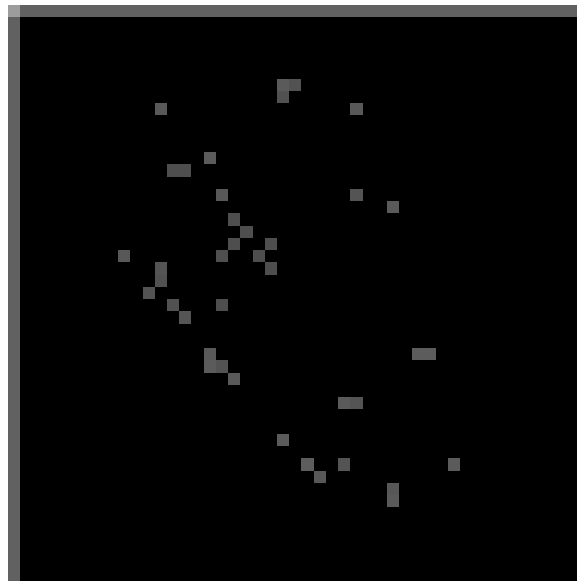


Abbildung 5.21: Ein Disparitätsbild aus der Sequenz.

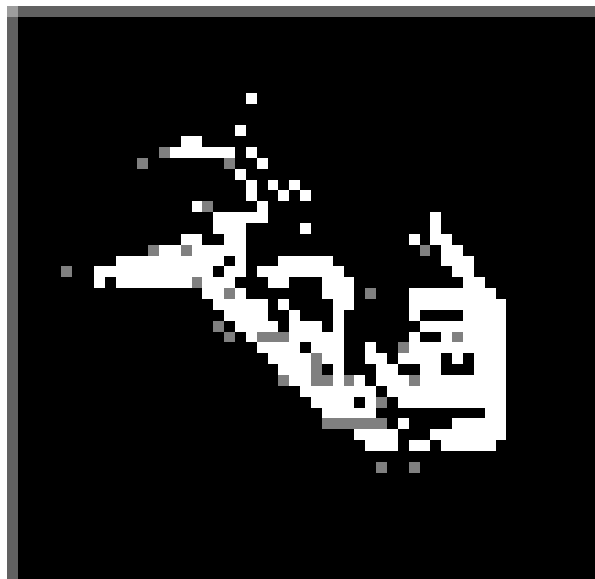


Abbildung 5.22: Der prädizierte Objektort (graue Pixel) für das zehnte Bild und der tatsächlich erreichte Ort (weiße Pixel). Der prädizierte Ort stimmt mit dem realen Ankunftsort gut überein.

## 5.5 Fünfte Beispielsequenz: Vergleich desselben Objekts bei großer Entfernung und bei geringer Entfernung zum Roboter

In diesem Abschnitt wird die Positionsprädiktion eines Objekts verglichen, das sich einmal nahe am Roboter bewegt und ein anderes mal eine größere Entfernung zum Roboter aufweist (siehe Abbildungen B.4 und B.5). Bei dem Objekt handelt es sich um den Heftapparat aus Abschnitt 5.4.

In der ersten Sequenz (Sequenz (A)) bewegt sich der Heftapparat, zu erkennen in der Bildmitte, diagonal auf der weißen Arbeitsplatte auf den Roboter zu. Zu Beginn liegt der Heftapparatmittelpunkt an den Koordinaten  $(-195, 720, 50)^T$ . Am Ende befindet sich der Mittelpunkt des Heftapparats an den Koordinaten  $(-51, 655, 20)^T$ . Für die Geschwindigkeitskomponenten erhält man  $(18.8, 7.2, 0)^T \frac{\text{mm}}{\text{Bild}}$ .

Die zweite Sequenz (Sequenz (B)) zeigt denselben Heftapparat, der sich jetzt weiter entfernt vom Roboter bewegt. Zu sehen ist der Heftapparat in der Mitte des oberen Bilddrittels auf einem weißen Untergrund. Der Heftapparat beschreibt eine frontale Bewegung auf den Roboter zu. Die Anfangskoordinaten des Heftapparatmittelpunkts sind  $(430, 2060, -375)^T$ . Am Ende der Sequenz liegt der Mittelpunkt bei den Koordinaten  $(460, 1885, -375)^T$ . Die Geschwindigkeitskomponenten lauten  $(3.3, 19.4, 0)^T \frac{\text{mm}}{\text{Bild}}$ .

Die Objektortprädiktion für das sich nahe am Roboter bewegende Objekt liefert ein gutes Ergebnis. Die Ortsprädiktion für das Objekt, das sich weiter vom Roboter entfernt befindet, fällt dagegen ungenauer aus.

Ein Grund hierfür ist darin zu finden, dass in der Sequenz (B) weniger Kantenpixel gefunden werden. Ein anderer Grund ist, dass die Entfernung mit zunehmender Disparität exponentiell abnimmt (siehe Abb. 5.25). In dem hier verwendeten Verfahren können die Disparitäten nur ganzzahlige Werte annehmen, da die Suche korrespondierender Pixel nicht auf Subpixelebene geschieht. Je kleiner die Disparitäten sind, desto größer ist die Entfernung. Dies bedeutet, dass der Unterschied zwischen zwei Entfernungen für zwei benachbarte kleine Disparitäten recht groß wird. Da die Objektposition aus der Disparität bestimmt wird, ergibt sich hieraus eine größere Ungenauigkeit für die Entfernungsbestimmung. Ferner hat das Objekt von der Sequenz (A) zur Sequenz (B) eine etwas andere Objektlage. In der Sequenz (B) ist vom Heftapparat hauptsächlich nur die Längsseite zum Stereokamerapaar orientiert, der hintere Teil des Heftapparats ist verdeckt. Dadurch verliert der Heftapparat etwas an Räumlichkeit und es sind an der Bestimmung der Hausdorffdistanz in der Mehrheit Raumpunkte einer Ebene beteiligt. Die Bestimmung für die  $y$ -Komponente fällt entsprechend unscharf aus. Die Fol-

ge ist eine ungenauere Bestimmung der Objektverschiebung im Raum (vgl. Kapitel 2.4.3 und Kapitel 2.5.4).

Hingegen ist der Heftapparat in der Sequenz (A) von mehreren Seiten zu erkennen, nämlich ein wenig von oben, ein wenig von hinten und ein wenig von der Seite. Die räumliche Information sollte ausreichend groß sein, um eine gute Positionsprädiktion zu liefern.

Vergleicht man die Kantenbilder aus Sequenz (A) und Sequenz (B) miteinander, so ist zu erkennen, dass die Kanten der Sequenz (A) eine bessere Qualität haben als die der Sequenz (B) (siehe Abb. 5.23). Dies ist aber auch zu erwarten, da der Heftapparat in der Sequenz (A) auf mehrere Sensorpixel abgebildet wird.

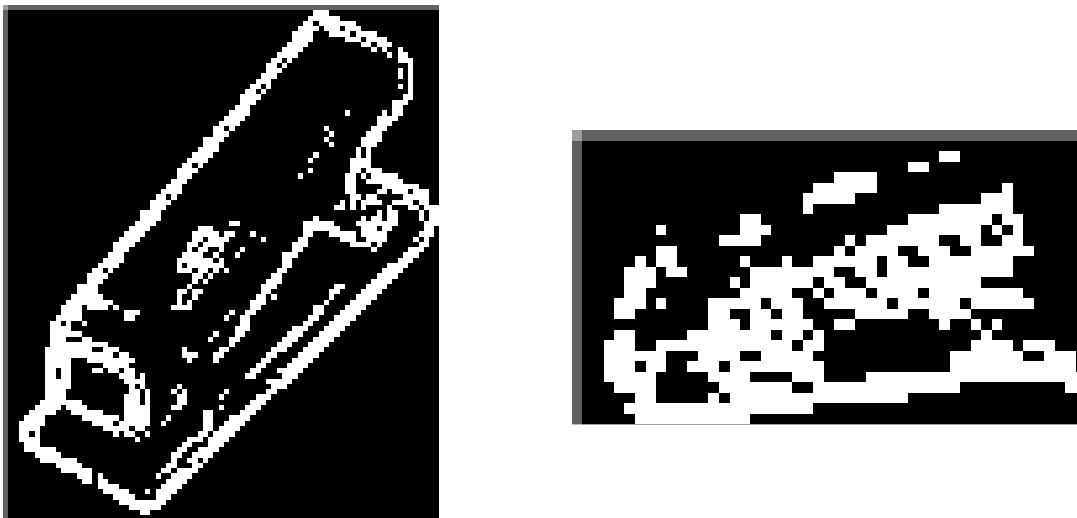


Abbildung 5.23: Die Kantenbilder der Sequenz (A) (linkes Bild) und Sequenz (B) (rechtes Bild). Aufgrund der Nähe zum Roboter werden in den Bildern der Sequenz (A) mehr Kanten gefunden. Die Struktur des Heftapparats ist in beiden Bildern erkennbar und in beiden Sequenzen werden vorwiegend nur Objektkanten gefunden.

Folglich werden in der Sequenz (A) mehr Disparitätswerte gefunden (siehe Abb. 5.24). Dadurch wird die Bestimmung der Objektposition für die Sequenz (A) genauer, was eine bessere Ermittlung der Objektverschiebung mit sich bringt.

Eine größere Fehlerquelle ist die Abnahme der Genauigkeit der Entfernung, wenn sich das Objekt weiter von den Kameras entfernt befindet. In Abbildung 5.25 ist ein Diagramm zu sehen, in dem die  $y$ -Komponenten der Weltkoordinaten in Abhängigkeit von der Disparität aufgetragen sind. Ist die

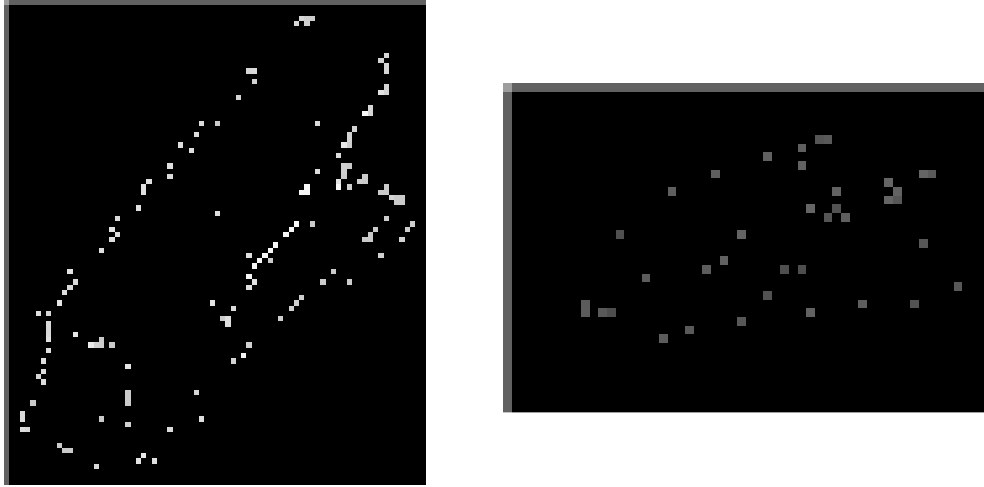


Abbildung 5.24: Disparitätsbilder aus beiden Sequenzen. Für beide Sequenzen wurden ausreichend viele und gute Disparitätswerte gefunden.

Disparität  $d$  des Pixels  $f(k_x, k_y)$  in einem Disparitätsbild  $f_d(d_x, d_y)$  bekannt, dann hängt die  $y$ -Komponente  $W_y$  der Weltkoordinate  $\vec{W}$  von der Disparität  $d$  in folgender Form ab:

$$W_y(d) = c_1 \left( \frac{c_2}{d} - 1 \right) + \frac{c_3 + c_4}{d}$$

mit

$$\begin{aligned} c_1 &= -\frac{b}{2} \sin \phi, \\ c_2 &= k_x - \frac{B}{2}, \\ c_3 &= \cos \phi \cos \theta \frac{bf}{p_v}, \\ c_4 &= \frac{2H}{p_v} b \left( k_y - \frac{H}{2} \right) \cos \phi \sin \theta, \end{aligned}$$

mit

$\phi$  = Azimutwinkel der Kamera,  $\theta$  = Neigungswinkel der Kamera,  
 $k_x$  =  $x$ -Koordinate des Pixels,  $k_y$  =  $y$ -Koordinate des Pixels,  
 $B$  = Bildbreite,  $H$  = Bildhöhe,  $p_v$  = Pixelbreite,  $p_h$  = Pixelhöhe,  
 $b$  = Basisbreite,  $f$  = Linsenbrennweite.

Eine Vergrößerung des Ausschnitts, der den Wertebereich der  $y$ -Weltkoordinaten für die Disparitätswerte von 61 pixel bis 101 pixel zeigt, ist in Abbildung 5.26 zu sehen. Diesen Wertebereich decken ungefähr die Disparitäten der Sequenz (B) ab. In Diagramm 5.26 kann man ablesen, dass sich  $W_y$  zwischen zwei

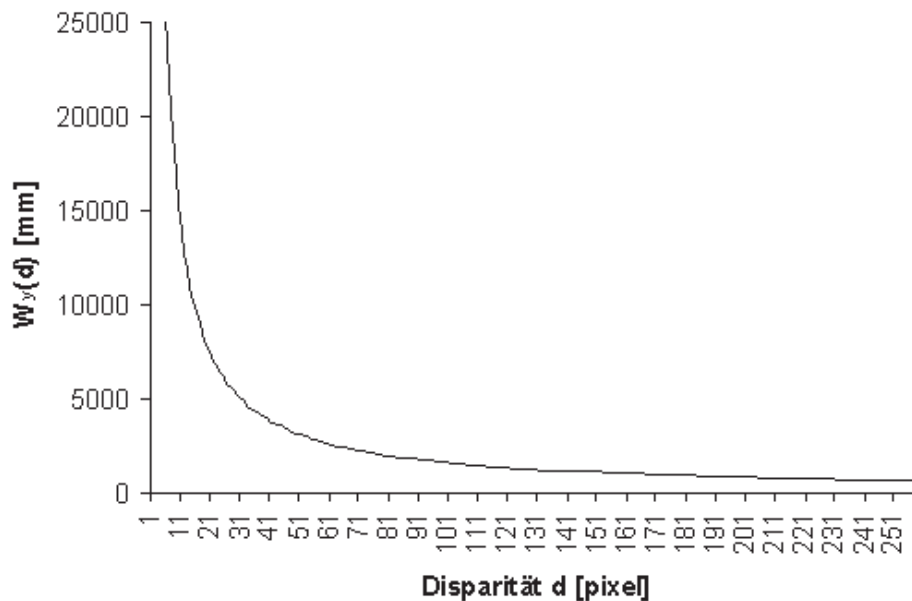


Abbildung 5.25: Diagramm der Entfernung über die Disparität. Die Entfernung nimmt exponentiell mit der Disparität ab.

Disparitätswerten um ca. 25 mm ändert. Dies bedeutet eine maximal mögliche Genauigkeit der Weltkoordinatenbestimmung von ca.  $W_y \pm 25$  mm, da die Disparitätswerte aufgrund des gewählten Verfahrens ganzzahlig sind.

In Diagramm 5.27 erkennt man den Wertebereich der  $y$ -Weltkoordinaten für die Disparitäten von 200 pixel bis 250 pixel. Dies entspricht in etwa den Disparitätswerten der Sequenz (A). Aus dem Diagramm lässt sich ablesen, dass die Änderung des  $y$ -Wertes zwischen zwei Disparitäten ca. 2,5 mm beträgt. Die Auflösung ist in diesem Fall also deutlich besser. Somit lassen sich die Weltkoordinaten für das Objekt in Sequenz (A) bis auf 2 mm Genauigkeit bestimmen.

Diese Toleranzen finden sich auch in den Abweichungen der Objektortsprädiktion und der tatsächlich erreichten Objektpositionen wieder, die sich für die Sequenz (A) und für die Sequenz (B) ergeben.

Am Ende beträgt die mittlere Abweichung des prädizierten Objektmittelpunktes vom tatsächlich erreichten Objektmittelpunkt für Sequenz (A)  $(5, 4, 11)^T$  mm und für die Sequenz (B)  $(7, 44, 22)^T$  mm. Die zu erwartende Abweichung des  $y$ -Wertes in der Sequenz (B) erklärt sich dadurch, dass der Heftapparat in dieser Sequenz eine ungünstige Orientierung zu den Kame-

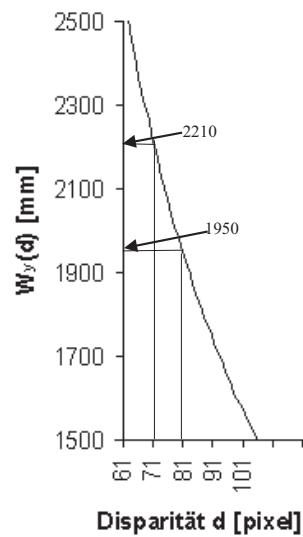


Abbildung 5.26: Vergrößerung eines Ausschnitts von Diagramm 5.25. Man kann ablesen, dass sich die Entfernung zwischen  $d=71$  pixel und  $d=81$  pixel um 260 mm ändert.

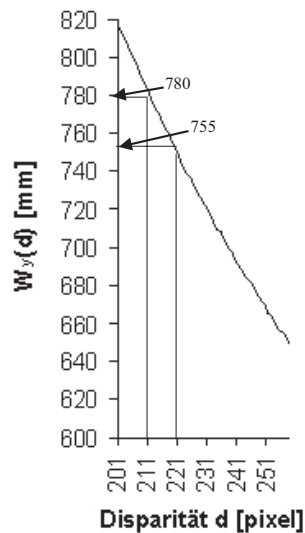


Abbildung 5.27: In diesem Diagramm ist zu erkennen, dass die Entfernung zwischen  $d=211$  pixel und  $d=221$  pixel um 25 mm abnimmt.

ras besitzt, da er hauptsächlich nur von einer Seite zu sehen ist. Die Un-

genauigkeit in der  $z$ -Komponente ist zum einen auf die Beschränkung der Bestimmungsgenauigkeit zurückzuführen und zum anderen darauf, dass der linke Teil des Heftapparats weniger Struktur aufweist als der rechte Teil. Dadurch ergeben sich für den rechten Teil des Heftapparats mehr Kanten und somit auch mehr Disparitätswerte für diese Hälfte des Heftapparats (siehe Abb. 5.28). Da der rechte Teil des Heftapparats einen höheren Schwerpunkt hat als der linke, verschiebt sich die prädizierte Höhe des Objektmittelpunktes nach oben. Am Mittelpunkt hat der Heftapparat die Höhe 50 mm. Der höchste Punkt des Heftapparats ist 70 mm hoch, also 20 mm höher als am Heftapparatmittelpunkt. Zusammen mit der durch die Auflösung begrenzten Genauigkeit erklärt dies die Abweichung der  $z$ -Komponente von 22 mm.

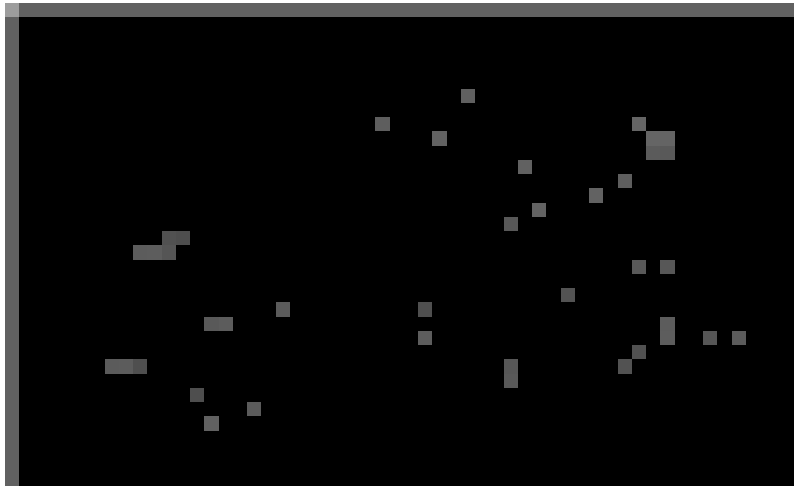


Abbildung 5.28: Disparitätsbild der Sequenz (B): die rechte Hälfte zeigt mehr Disparitätswerte als die linke. Zudem sind die Disparitätswerte der rechten Hälfte höher (zu erkennen an der Pixelhelligkeit) als in der linken Hälfte.

Wie in Tabelle A.4 und in Abbildung 5.29 zu erkennen, ist der Objektort für die Sequenz (A) gut geschätzt worden. In Tabelle A.5 erkennt man, dass sich der prädizierte Objektort und der tatsächliche Objektort nicht gleichen. In Abbildung 5.30 scheint es aber, als würde der Objektort gut vorhergesagt sein, da kaum ein graues Pixel weit entfernt von einem weißen Pixel liegt. Dies lässt sich dadurch erklären, dass lediglich die Entfernung (das entspricht der  $y$ -Komponente) ungenau geschätzt wurde, nicht aber die horizontale Lage (das entspricht der  $x$ -Komponente). Da in die Berechnung der Koordinaten der grauen Pixel aber nur die  $x$ - und die  $z$ -Komponente der Weltkoordinaten eingehen, deren Abweichung nicht so hoch ist, ergibt sich das vorliegende Prädiktionsbild.

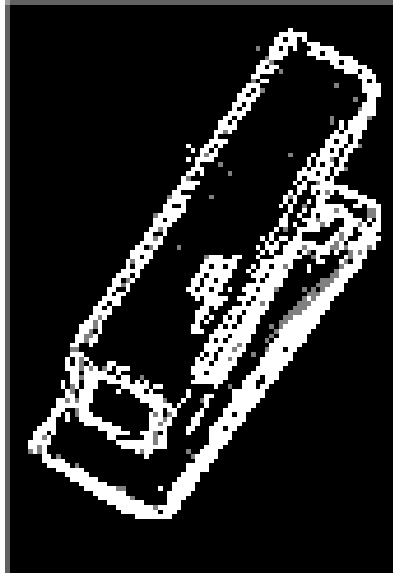


Abbildung 5.29: Der prädizierte Objektort (graue Pixel) für das zehnte Bild und der tatsächlich erreichte Ort (weiße Pixel). Der prädizierte Ort stimmt mit dem realen Ankunftsort sehr gut überein.

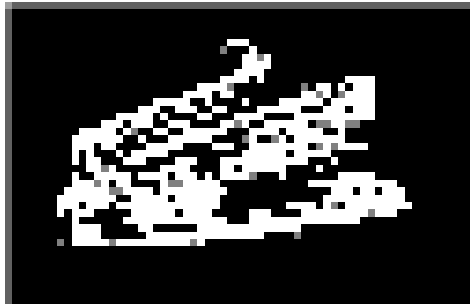


Abbildung 5.30: Der prädizierte Objektort (graue Pixel) für das zehnte Bild und der tatsächlich erreichte Ort (weiße Pixel). Da in die Berechnung des prädizierten Objektortes im Bild nur die prädizierte horizontale Lage und die Höhe des Heftapparats eingehen, scheint der prädizierte Ort mit dem realen Ankunftsort in dieser Abbildung gut übereinzustimmen. Allerdings ist in Tabelle A.5 zu erkennen, dass die Entfernung von der realen Objektposition abweicht.

# Kapitel 6

## Zusammenfassung und Ausblick

### 6.1 Allgemeine Bewertung der Ergebnisse

In dieser Arbeit wurde ein Softwaremodul zur ansichtenbasierten Erkennung und Verfolgung hochdynamischer Objekte für einen anthropomorphen Roboterassistenten implementiert. Der Roboter besitzt ein 2-Freiheitsgrade-Stereokamerasystem. Für die Objektverfolgung wird der Kamerakopf fixiert, das heißt, er bleibt während der Objektverfolgung unbewegt. Mit Hilfe des unbewegten Kamerakopfes kann durch ein Differenzbild die Objektbewegung lokalisiert werden, um die Objektverfolgung zu initialisieren. Das Differenzbild resultiert aus dem aktuell eingelesenen und dem vorangegangenen Bild. Um einen Bildbereich zu ermitteln, in dem die Objektbewegung stattfindet, erfolgt eine Segmentierung des Differenzbildes. Innerhalb dieses Bereichs erfolgen eine Kantenextraktion und eine Disparitätsanalyse. Aus diesem Wissen wird eine Raumposition des Gegenstands geschätzt. In die Schätzung fließt das Wissen über die definierte Objektgeometrie mit ein, was eine exaktere Schätzung der Raumposition ermöglicht. Die ermittelten Raumdaten werden gespeichert.

Ist das nächste Bild aufgenommen, erfolgt die Bestimmung der Raumdaten analog. Die dreidimensionale Verschiebung des Objekts wird über die partielle Hausdorffdistanz ermittelt. Diese erlaubt eine gute Schätzung der Objektverlagerung unter Anwesenheit von Störpunkten, die nicht zum Objekt gehören. Die Positionsänderung liefert die Geschwindigkeit, aus der die Raumpositionen für den Gegenstand prädictiert werden. Es wird die Position für das nächste Bild prädictiert, um im Folgebild die Differenz zwischen prädictierter Position von der berechneten Position zu ermitteln. Dadurch lassen sich die Geschwindigkeiten genauer schätzen, aus denen ein Mittelwert gebildet wird. Dies gewährleistet eine exaktere Bestimmung des Greifortes. Die

Ermittlung der Greifposition erfolgt, indem die prädizierte Objektposition mit einem Wert verglichen wird, der eine Mindestnähe zum Greifer darstellt. Stimmt eine prädizierte Position ungefähr mit der Mindestnähe überein, wird diese als Greifort gewählt, sofern der Greifer den Greifort innerhalb der *time-to-contact* erreichen kann.

Die Ergebnisse zeigen, dass dieses Verfahren zur Objektprädiktion für Gegenstände mit ausreichender Struktur gut geeignet ist. Die an die Umgebung gestellten Bedingungen sind die begrenzte Objektgeschwindigkeit, die definierte Objektgeometrie, eine konstante Helligkeit, unbewegte Kameras und die Bewegung eines einzelnen Objekts im Bild.

Die Tests wurden auf einer Sun Blade-100 durchgeführt. Die Prädiktion betrug für entfernte Gegenstände durchschnittlich 150 ms (im Maximalfall 300 ms) und für nahe Gegenstände im Durchschnitt 1000 ms (300 ms maximal und 2000 ms im minimalen Fall). Ein einzelnes Rechnermodul von CoRA ist wesentlich leistungsfähiger. Dadurch ist eine Echtzeitfähigkeit bei Einsatz dieses Softwaremoduls auf CoRA gegeben. In den meisten Beispielsequenzen wurden nach spätestens sechs Bildaufnahmen gute Übereinstimmungen zwischen prädiziertem und erreichtem Greifort erzielt.

## 6.2 Mögliche Verbesserungen und Weiterentwicklung

Um den Greifort schneller als nach sechs Bildaufnahmen zu präzisieren, lässt sich die Ermittlung der Positionsänderung des Gegenstands verfeinern. Zwar kann die Positionsverlagerung bei dem vorliegenden Softwaremodul beliebig genau bestimmt werden, dies schränkt aber die Echtzeitfähigkeit ein, da die Bestimmung der partiellen Hausdorffdistanzen aufwändig ist. Eine Vereinfachung der partiellen Hausdorffdistanz liefert eine lokale Approximation einer Distanzfunktion. Eine Distanzfunktion einer Menge  $B$  bezeichnet den minimalen Abstand zum Element einer Menge als Funktion des Ortes. Die Berechnung der partiellen Hausdorffdistanz kann durch die Auswertung der Distanzfunktion vereinfacht werden (siehe [17]). Die Distanzfunktion alleine liefert keine Verbesserung des Algorithmus. Diese ist erst durch eine lokale Approximation der Distanzfunktion möglich. Die lokale Approximation bedeutet die Berechnung der Distanzfunktion nur für Werte, die in der lokalen Nachbarschaft von Datenpunkten liegen. Dadurch wird der algorithmische Aufwand verringert, was eine Verfeinerung der Ermittlung der Positionsänderung unter Einhaltung der Echtzeitfähigkeit erlaubt.

## 6.2. MÖGLICHE VERBESSERUNGEN UND WEITERENTWICKLUNG 71

Um den Ort des Suchbereichs nach dem Gegenstand im Folgebild abzuschätzen, kann ein Kalmanfilter eingesetzt werden, sobald die Positionsprädiktion erfolgreich verläuft. Die aufwändige Bestimmung einer *Region of Interest* würde unter diesen Umständen nach einer Initialisierung des Kalmanfilters entfallen.

Um das Greifen endgültig einzuleiten, muss die Orientierung des Gegenstands im Raum bestimmt werden. Ein Verfahren, die Objektorientierung zu ermitteln, könnte ähnlich zu einer Arbeit in [18] erfolgen. Anstatt wie dort die Orientierung über Kantenbilder zu ermitteln, kann das Disparitätsbild herangezogen werden. Hierüber ist eine Orientierung des Gegenstands im dreidimensionalen Raum denkbar. Die Voraussetzung hierfür ist, dass das Disparitätsbild genügend Information besitzt. Die Disparitätsbilder in dieser Arbeit erfüllen diese Bedingung, so dass die Orientierung mit dieser Methode zu bestimmen sein müsste.



# Anhang A

## Tabellen

Bild	$\vec{x}_{real}[mm]$	$\vec{x}_{det}[mm]$	$\vec{x}_{pred}[mm]$	$\vec{x}_{pred,gr}[mm]$	$\delta_{pr_{det}}[mm]$	$\delta_{pr_{gr}}[mm]$	$\vec{v}_{det}[\frac{mm}{Bild}]$
1	$\begin{pmatrix} -300 \\ 780 \\ 50 \end{pmatrix}$	/	/	/	/	/	/
2	$\begin{pmatrix} -300 \\ 760 \\ 50 \end{pmatrix}$	/	/	/	/	/	$\begin{pmatrix} 0 \\ 0 \\ 2 \end{pmatrix}$
3	$\begin{pmatrix} -300 \\ 740 \\ 50 \end{pmatrix}$	$\begin{pmatrix} -301 \\ 723 \\ 54 \end{pmatrix}$	$\begin{pmatrix} -300 \\ 745 \\ 52 \end{pmatrix}$	$\begin{pmatrix} -261 \\ 643 \\ 102 \end{pmatrix}$	$\begin{pmatrix} 0 \\ 5 \\ 2 \end{pmatrix}$	$\begin{pmatrix} 39 \\ 43 \\ 52 \end{pmatrix}$	$\begin{pmatrix} 5 \\ -10 \\ 6 \end{pmatrix}$
4	$\begin{pmatrix} -300 \\ 720 \\ 50 \end{pmatrix}$	$\begin{pmatrix} -300 \\ 702 \\ 51 \end{pmatrix}$	$\begin{pmatrix} -296 \\ 713 \\ 60 \end{pmatrix}$	$\begin{pmatrix} -270 \\ 627 \\ 90 \end{pmatrix}$	$\begin{pmatrix} 4 \\ -7 \\ 10 \end{pmatrix}$	$\begin{pmatrix} 30 \\ 27 \\ 40 \end{pmatrix}$	$\begin{pmatrix} 3 \\ -13 \\ 4 \end{pmatrix}$
5	$\begin{pmatrix} -300 \\ 700 \\ 50 \end{pmatrix}$	$\begin{pmatrix} -300 \\ 678 \\ 52 \end{pmatrix}$	$\begin{pmatrix} -297 \\ 689 \\ 55 \end{pmatrix}$	$\begin{pmatrix} -270 \\ 601 \\ 77 \end{pmatrix}$	$\begin{pmatrix} 3 \\ -11 \\ 5 \end{pmatrix}$	$\begin{pmatrix} 30 \\ 1 \\ 27 \end{pmatrix}$	$\begin{pmatrix} 6 \\ -18 \\ 2 \end{pmatrix}$
6	$\begin{pmatrix} -300 \\ 680 \\ 50 \end{pmatrix}$	$\begin{pmatrix} -295 \\ 650 \\ 54 \end{pmatrix}$	$\begin{pmatrix} -295 \\ 661 \\ 54 \end{pmatrix}$	$\begin{pmatrix} -267 \\ 580 \\ 70 \end{pmatrix}$	$\begin{pmatrix} 5 \\ -19 \\ 4 \end{pmatrix}$	$\begin{pmatrix} 33 \\ -20 \\ 20 \end{pmatrix}$	$\begin{pmatrix} 6 \\ -18 \\ 2 \end{pmatrix}$
7	$\begin{pmatrix} -300 \\ 660 \\ 50 \end{pmatrix}$	$\begin{pmatrix} -304 \\ 653 \\ 40 \end{pmatrix}$	$\begin{pmatrix} -289 \\ 632 \\ 56 \end{pmatrix}$	$\begin{pmatrix} 276 \\ 591 \\ 59 \end{pmatrix}$	$\begin{pmatrix} 11 \\ -28 \\ 6 \end{pmatrix}$	$\begin{pmatrix} 24 \\ -9 \\ 9 \end{pmatrix}$	$\begin{pmatrix} 5 \\ -13 \\ 2 \end{pmatrix}$
8	$\begin{pmatrix} -300 \\ 640 \\ 50 \end{pmatrix}$	$\begin{pmatrix} -296 \\ 616 \\ 54 \end{pmatrix}$	$\begin{pmatrix} -299 \\ 640 \\ 42 \end{pmatrix}$	$\begin{pmatrix} -279 \\ 582 \\ 61 \end{pmatrix}$	$\begin{pmatrix} 1 \\ 0 \\ -8 \end{pmatrix}$	$\begin{pmatrix} 21 \\ -18 \\ 11 \end{pmatrix}$	$\begin{pmatrix} 5 \\ -14 \\ 3 \end{pmatrix}$
9	$\begin{pmatrix} -300 \\ 620 \\ 50 \end{pmatrix}$	$\begin{pmatrix} -304 \\ 606 \\ 38 \end{pmatrix}$	$\begin{pmatrix} -291 \\ 602 \\ 57 \end{pmatrix}$	$\begin{pmatrix} -289 \\ 583 \\ 51 \end{pmatrix}$	$\begin{pmatrix} 9 \\ -18 \\ 7 \end{pmatrix}$	$\begin{pmatrix} 11 \\ -17 \\ 1 \end{pmatrix}$	$\begin{pmatrix} 3 \\ -11 \\ 2 \end{pmatrix}$
10	$\begin{pmatrix} -300 \\ 600 \\ 50 \end{pmatrix}$	$\begin{pmatrix} -305 \\ 595 \\ 32 \end{pmatrix}$	$\begin{pmatrix} -301 \\ 595 \\ 32 \end{pmatrix}$	$\begin{pmatrix} -295 \\ 584 \\ 42 \end{pmatrix}$	$\begin{pmatrix} -1 \\ -5 \\ -18 \end{pmatrix}$	$\begin{pmatrix} 5 \\ -16 \\ -8 \end{pmatrix}$	$\begin{pmatrix} 4 \\ -11 \\ 2 \end{pmatrix}$

Tabelle A.1: Ergebnisse für die Objektortprädiktion der zweiten Beispielsequenz. Die Geschwindigkeitskomponenten betragen konstant  $(0, 20, 0) \frac{mm}{Bild}$ .

Bild	$\vec{x}_{real}[mm]$	$\vec{x}_{det}[mm]$	$\vec{x}_{pred}[mm]$	$\vec{x}_{pred,gr}[mm]$	$\delta_{pr_{det}}[mm]$	$\delta_{pr_{qr}}[mm]$	$\vec{v}_{det}[\frac{mm}{Bild}]$
1	$\begin{pmatrix} -155 \\ 760 \\ 35 \end{pmatrix}$	/	/	/	/	/	/
2	$\begin{pmatrix} -135 \\ 760 \\ 35 \end{pmatrix}$	/	/	/	/	/	$\begin{pmatrix} 20 \\ 0 \\ 6 \end{pmatrix}$
3	$\begin{pmatrix} -115 \\ 760 \\ 35 \end{pmatrix}$	$\begin{pmatrix} -103 \\ 691 \\ 60 \end{pmatrix}$	$\begin{pmatrix} -111 \\ 746 \\ 15 \end{pmatrix}$	$\begin{pmatrix} 57 \\ 691 \\ 108 \end{pmatrix}$	$\begin{pmatrix} 4 \\ -14 \\ -20 \end{pmatrix}$	$\begin{pmatrix} 32 \\ -69 \\ 93 \end{pmatrix}$	$\begin{pmatrix} 20 \\ 0 \\ 6 \end{pmatrix}$
4	$\begin{pmatrix} -95 \\ 760 \\ 35 \end{pmatrix}$	$\begin{pmatrix} -94 \\ 743 \\ 13 \end{pmatrix}$	$\begin{pmatrix} -83 \\ 691 \\ 66 \end{pmatrix}$	$\begin{pmatrix} 51 \\ 717 \\ 81 \end{pmatrix}$	$\begin{pmatrix} 12 \\ 69 \\ 31 \end{pmatrix}$	$\begin{pmatrix} 26 \\ -43 \\ 66 \end{pmatrix}$	$\begin{pmatrix} 20 \\ 0 \\ 6 \end{pmatrix}$
5	$\begin{pmatrix} -75 \\ 760 \\ 35 \end{pmatrix}$	$\begin{pmatrix} -70 \\ 748 \\ 10 \end{pmatrix}$	$\begin{pmatrix} -74 \\ 743 \\ 19 \end{pmatrix}$	$\begin{pmatrix} 57 \\ 733 \\ 60 \end{pmatrix}$	$\begin{pmatrix} 1 \\ -17 \\ 16 \end{pmatrix}$	$\begin{pmatrix} 32 \\ -27 \\ 45 \end{pmatrix}$	$\begin{pmatrix} 22 \\ 0 \\ 5 \end{pmatrix}$
6	$\begin{pmatrix} -55 \\ 760 \\ 35 \end{pmatrix}$	$\begin{pmatrix} -55 \\ 746 \\ 8 \end{pmatrix}$	$\begin{pmatrix} -48 \\ 748 \\ 15 \end{pmatrix}$	$\begin{pmatrix} 56 \\ 745 \\ 49 \end{pmatrix}$	$\begin{pmatrix} 7 \\ -12 \\ -20 \end{pmatrix}$	$\begin{pmatrix} 31 \\ -15 \\ 24 \end{pmatrix}$	$\begin{pmatrix} 22 \\ 2 \\ 6 \end{pmatrix}$
7	$\begin{pmatrix} -35 \\ 760 \\ 35 \end{pmatrix}$	$\begin{pmatrix} -34 \\ 748 \\ 9 \end{pmatrix}$	$\begin{pmatrix} -33 \\ 748 \\ 14 \end{pmatrix}$	$\begin{pmatrix} 53 \\ 757 \\ 41 \end{pmatrix}$	$\begin{pmatrix} 2 \\ -12 \\ -21 \end{pmatrix}$	$\begin{pmatrix} 28 \\ -3 \\ 16 \end{pmatrix}$	$\begin{pmatrix} 21 \\ 5 \\ 6 \end{pmatrix}$
8	$\begin{pmatrix} -15 \\ 760 \\ 35 \end{pmatrix}$	$\begin{pmatrix} -17 \\ 864 \\ -95 \end{pmatrix}$	$\begin{pmatrix} -13 \\ 753 \\ 15 \end{pmatrix}$	$\begin{pmatrix} 49 \\ 821 \\ -17 \end{pmatrix}$	$\begin{pmatrix} 2 \\ -7 \\ 20 \end{pmatrix}$	$\begin{pmatrix} 24 \\ 21 \\ -32 \end{pmatrix}$	$\begin{pmatrix} 21 \\ 7 \\ 7 \end{pmatrix}$
9	$\begin{pmatrix} 5 \\ 760 \\ 35 \end{pmatrix}$	$\begin{pmatrix} 923 \\ 924 \\ -121 \end{pmatrix}$	$\begin{pmatrix} 4 \\ 871 \\ -88 \end{pmatrix}$	$\begin{pmatrix} 48 \\ 883 \\ -64 \end{pmatrix}$	$\begin{pmatrix} -1 \\ 111 \\ -122 \end{pmatrix}$	$\begin{pmatrix} 23 \\ 123 \\ -69 \end{pmatrix}$	$\begin{pmatrix} 22 \\ 10 \\ 5 \end{pmatrix}$
10	$\begin{pmatrix} 25 \\ 760 \\ 35 \end{pmatrix}$	$\begin{pmatrix} 17 \\ 755 \\ 1 \end{pmatrix}$	$\begin{pmatrix} 25 \\ 934 \\ -116 \end{pmatrix}$	$\begin{pmatrix} 44 \\ 825 \\ -30 \end{pmatrix}$	$\begin{pmatrix} 0 \\ 174 \\ -151 \end{pmatrix}$	$\begin{pmatrix} 19 \\ 65 \\ -45 \end{pmatrix}$	$\begin{pmatrix} 23 \\ 12 \\ 3 \end{pmatrix}$

Tabelle A.2: Ergebnisse für die Objektortprädiktion der dritten Beispielsequenz. Die Geschwindigkeitskomponenten betragen konstant  $(20, 0, 0) \frac{mm}{Bild}$ .

Bild	$\vec{x}_{real}[mm]$	$\vec{x}_{det}[mm]$	$\vec{x}_{pred}[mm]$	$\vec{x}_{pred,gr}[mm]$	$\delta_{pr_{det}}[mm]$	$\delta_{pr_{gr}}[mm]$	$\vec{v}_{det}[\frac{mm}{Bild}]$
1	$\begin{pmatrix} 370 \\ 1970 \\ -375 \end{pmatrix}$	/	/	/	/	/	/
2	$\begin{pmatrix} 389 \\ 1975 \\ -375 \end{pmatrix}$	/	/	/	/	/	$\begin{pmatrix} 30 \\ 0 \\ 6 \end{pmatrix}$
3	$\begin{pmatrix} 409 \\ 1981 \\ -375 \end{pmatrix}$	$\begin{pmatrix} 374 \\ 1993 \\ -377 \end{pmatrix}$	$\begin{pmatrix} 374 \\ 1995 \\ -382 \end{pmatrix}$	$\begin{pmatrix} 614 \\ 1993 \\ -329 \end{pmatrix}$	$\begin{pmatrix} -35 \\ 14 \\ -5 \end{pmatrix}$	$\begin{pmatrix} 69 \\ -27 \\ 46 \end{pmatrix}$	$\begin{pmatrix} 30 \\ 0 \\ 6 \end{pmatrix}$
4	$\begin{pmatrix} 428 \\ 1986 \\ -375 \end{pmatrix}$	$\begin{pmatrix} 383 \\ 2009 \\ -383 \end{pmatrix}$	$\begin{pmatrix} 404 \\ 1993 \\ -371 \end{pmatrix}$	$\begin{pmatrix} 589 \\ 2036 \\ -346 \end{pmatrix}$	$\begin{pmatrix} -24 \\ 7 \\ 4 \end{pmatrix}$	$\begin{pmatrix} 44 \\ 16 \\ 29 \end{pmatrix}$	$\begin{pmatrix} 26 \\ 10 \\ 3 \end{pmatrix}$
5	$\begin{pmatrix} 448 \\ 1992 \\ -375 \end{pmatrix}$	$\begin{pmatrix} 401 \\ 1994 \\ -381 \end{pmatrix}$	$\begin{pmatrix} 409 \\ 2019 \\ -380 \end{pmatrix}$	$\begin{pmatrix} 576 \\ 2036 \\ -355 \end{pmatrix}$	$\begin{pmatrix} -39 \\ 27 \\ -5 \end{pmatrix}$	$\begin{pmatrix} 31 \\ 16 \\ 20 \end{pmatrix}$	$\begin{pmatrix} 27 \\ 7 \\ 3 \end{pmatrix}$
6	$\begin{pmatrix} 467 \\ 1997 \\ -375 \end{pmatrix}$	$\begin{pmatrix} 422 \\ 1988 \\ -367 \end{pmatrix}$	$\begin{pmatrix} 428 \\ 2001 \\ -378 \end{pmatrix}$	$\begin{pmatrix} 569 \\ 2012 \\ -351 \end{pmatrix}$	$\begin{pmatrix} -39 \\ 4 \\ -3 \end{pmatrix}$	$\begin{pmatrix} 24 \\ -8 \\ 24 \end{pmatrix}$	$\begin{pmatrix} 28 \\ 0 \\ 4 \end{pmatrix}$
7	$\begin{pmatrix} 487 \\ 2003 \\ -375 \end{pmatrix}$	$\begin{pmatrix} 447 \\ 2081 \\ -408 \end{pmatrix}$	$\begin{pmatrix} 450 \\ 1988 \\ -363 \end{pmatrix}$	$\begin{pmatrix} 564 \\ 2045 \\ -376 \end{pmatrix}$	$\begin{pmatrix} -37 \\ 78 \\ 12 \end{pmatrix}$	$\begin{pmatrix} 19 \\ 25 \\ -1 \end{pmatrix}$	$\begin{pmatrix} 28 \\ -1 \\ 2 \end{pmatrix}$
8	$\begin{pmatrix} 506 \\ 2008 \\ -375 \end{pmatrix}$	$\begin{pmatrix} 469 \\ 1998 \\ -376 \end{pmatrix}$	$\begin{pmatrix} 475 \\ 2080 \\ -406 \end{pmatrix}$	$\begin{pmatrix} 558 \\ 2018 \\ -374 \end{pmatrix}$	$\begin{pmatrix} -37 \\ 10 \\ -31 \end{pmatrix}$	$\begin{pmatrix} 13 \\ -2 \\ 1 \end{pmatrix}$	$\begin{pmatrix} 28 \\ -2 \\ 1 \end{pmatrix}$
9	$\begin{pmatrix} 526 \\ 2014 \\ -375 \end{pmatrix}$	$\begin{pmatrix} 495 \\ 2032 \\ -393 \end{pmatrix}$	$\begin{pmatrix} 497 \\ 1996 \\ -375 \end{pmatrix}$	$\begin{pmatrix} 554 \\ 2026 \\ -383 \end{pmatrix}$	$\begin{pmatrix} -31 \\ 18 \\ 0 \end{pmatrix}$	$\begin{pmatrix} 9 \\ 6 \\ -8 \end{pmatrix}$	$\begin{pmatrix} 28 \\ 1 \\ 1 \end{pmatrix}$
10	$\begin{pmatrix} 545 \\ 2020 \\ -375 \end{pmatrix}$	$\begin{pmatrix} 506 \\ 2015 \\ -379 \end{pmatrix}$	$\begin{pmatrix} 523 \\ 2033 \\ -392 \end{pmatrix}$	$\begin{pmatrix} 543 \\ 2020 \\ -380 \end{pmatrix}$	$\begin{pmatrix} 22 \\ 13 \\ -17 \end{pmatrix}$	$\begin{pmatrix} -2 \\ 0 \\ -5 \end{pmatrix}$	$\begin{pmatrix} 27 \\ -1 \\ 2 \end{pmatrix}$

Tabelle A.3: Ergebnisse für die Objektortprädiktion der vierten Beispielsequenz. Die Geschwindigkeitskomponenten betragen konstant  $(19.4, 5.5, 0) \frac{mm}{Bild}$ .

Bild	$\vec{x}_{real}[mm]$	$\vec{x}_{det}[mm]$	$\vec{x}_{pred}[mm]$	$\vec{x}_{pred,gr}[mm]$	$\delta_{pr,det}[mm]$	$\delta_{pr,gr}[mm]$	$\vec{v}_{det}[\frac{mm}{Bild}]$
0	$\begin{pmatrix} -195 \\ 720 \\ 50 \end{pmatrix}$	/	/	/	/	/	/
1	$\begin{pmatrix} -176 \\ 713 \\ 50 \end{pmatrix}$	/	/	/		/	$\begin{pmatrix} 20 \\ 0 \\ 6 \end{pmatrix}$
2	$\begin{pmatrix} -157 \\ 706 \\ 50 \end{pmatrix}$	$\begin{pmatrix} -174 \\ 696 \\ 7 \end{pmatrix}$	$\begin{pmatrix} -175 \\ 699 \\ 23 \end{pmatrix}$	$\begin{pmatrix} -14 \\ 696 \\ 55 \end{pmatrix}$	$\begin{pmatrix} -18 \\ -7 \\ -27 \end{pmatrix}$	$\begin{pmatrix} 11 \\ 41 \\ 5 \end{pmatrix}$	$\begin{pmatrix} 20 \\ 0 \\ 6 \end{pmatrix}$
3	$\begin{pmatrix} -138 \\ 699 \\ 50 \end{pmatrix}$	$\begin{pmatrix} -163 \\ 688 \\ 9 \end{pmatrix}$	$\begin{pmatrix} -154 \\ 696 \\ 13 \end{pmatrix}$	$\begin{pmatrix} -19 \\ 692 \\ 46 \end{pmatrix}$	$\begin{pmatrix} -16 \\ -3 \\ -37 \end{pmatrix}$	$\begin{pmatrix} 6 \\ 37 \\ -4 \end{pmatrix}$	$\begin{pmatrix} 20 \\ 0 \\ 4 \end{pmatrix}$
4	$\begin{pmatrix} -120 \\ 692 \\ 50 \end{pmatrix}$	$\begin{pmatrix} -148 \\ 677 \\ 4 \end{pmatrix}$	$\begin{pmatrix} -143 \\ 688 \\ 13 \end{pmatrix}$	$\begin{pmatrix} -17 \\ 684 \\ 40 \end{pmatrix}$	$\begin{pmatrix} -23 \\ -4 \\ -37 \end{pmatrix}$	$\begin{pmatrix} 8 \\ 29 \\ -10 \end{pmatrix}$	$\begin{pmatrix} 22 \\ 0 \\ 5 \end{pmatrix}$
5	$\begin{pmatrix} -101 \\ 684 \\ 50 \end{pmatrix}$	$\begin{pmatrix} -120 \\ 673 \\ 6 \end{pmatrix}$	$\begin{pmatrix} -126 \\ 677 \\ 6 \end{pmatrix}$	$\begin{pmatrix} -13 \\ 678 \\ 33 \end{pmatrix}$	$\begin{pmatrix} -25 \\ -7 \\ -41 \end{pmatrix}$	$\begin{pmatrix} 12 \\ 23 \\ -17 \end{pmatrix}$	$\begin{pmatrix} 22 \\ 0 \\ 4 \end{pmatrix}$
6	$\begin{pmatrix} -82 \\ 677 \\ 50 \end{pmatrix}$	$\begin{pmatrix} -86 \\ 695 \\ 43 \end{pmatrix}$	$\begin{pmatrix} -98 \\ 673 \\ 10 \end{pmatrix}$	$\begin{pmatrix} -7 \\ 687 \\ 46 \end{pmatrix}$	$\begin{pmatrix} -16 \\ -4 \\ -40 \end{pmatrix}$	$\begin{pmatrix} 18 \\ 32 \\ -4 \end{pmatrix}$	$\begin{pmatrix} 21 \\ 0 \\ 4 \end{pmatrix}$
7	$\begin{pmatrix} -63 \\ 670 \\ 50 \end{pmatrix}$	$\begin{pmatrix} -67 \\ 675 \\ 21 \end{pmatrix}$	$\begin{pmatrix} -65 \\ 695 \\ 47 \end{pmatrix}$	$\begin{pmatrix} -16 \\ 675 \\ 41 \end{pmatrix}$	$\begin{pmatrix} -2 \\ 25 \\ -3 \end{pmatrix}$	$\begin{pmatrix} 9 \\ 20 \\ -9 \end{pmatrix}$	$\begin{pmatrix} 14 \\ -4 \\ 5 \end{pmatrix}$
8	$\begin{pmatrix} -44 \\ 663 \\ 50 \end{pmatrix}$	$\begin{pmatrix} -55 \\ 674 \\ 32 \end{pmatrix}$	$\begin{pmatrix} -53 \\ 671 \\ 26 \end{pmatrix}$	$\begin{pmatrix} -28 \\ 667 \\ 43 \end{pmatrix}$	$\begin{pmatrix} -11 \\ 8 \\ -24 \end{pmatrix}$	$\begin{pmatrix} -3 \\ 12 \\ -7 \end{pmatrix}$	$\begin{pmatrix} 8 \\ -7 \\ 6 \end{pmatrix}$
9	$\begin{pmatrix} -25 \\ 655 \\ 50 \end{pmatrix}$	$\begin{pmatrix} -35 \\ 661 \\ 24 \end{pmatrix}$	$\begin{pmatrix} -47 \\ 667 \\ 38 \end{pmatrix}$	$\begin{pmatrix} -30 \\ 659 \\ 36 \end{pmatrix}$	$\begin{pmatrix} -22 \\ 12 \\ -12 \end{pmatrix}$	$\begin{pmatrix} -5 \\ 5 \\ -14 \end{pmatrix}$	$\begin{pmatrix} 4 \\ -10 \\ 6 \end{pmatrix}$

Tabelle A.4: Ergebnisse für die Objektortprädiktion der fünften Beispielsequenz für das nähere Objekt. Die Geschwindigkeitskomponenten betragen konstant  $(18.8, 7.2, 0) \frac{mm}{Bild}$ .

Bild	$\vec{x}_{real}[mm]$	$\vec{x}_{det}[mm]$	$\vec{x}_{pred}[mm]$	$\vec{x}_{pred,gr}[mm]$	$\delta_{pr_{det}}[mm]$	$\delta_{pr_{gr}}[mm]$	$\vec{v}_{det}[\frac{mm}{Bild}]$
1	$\begin{pmatrix} 430 \\ 2060 \\ -375 \end{pmatrix}$	/	/	/	/	/	/
2	$\begin{pmatrix} 433 \\ 2041 \\ -375 \end{pmatrix}$	/	/	/	/	/	$\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$
3	$\begin{pmatrix} 436 \\ 2021 \\ -375 \end{pmatrix}$	$\begin{pmatrix} 430 \\ 2084 \\ -371 \end{pmatrix}$	$\begin{pmatrix} 450 \\ 2362 \\ -461 \end{pmatrix}$	$\begin{pmatrix} 430 \\ 2084 \\ -371 \end{pmatrix}$	$\begin{pmatrix} -6 \\ 63 \\ 4 \end{pmatrix}$	$\begin{pmatrix} 30 \\ 101 \\ 4 \end{pmatrix}$	$\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$
4	$\begin{pmatrix} 440 \\ 2002 \\ -375 \end{pmatrix}$	$\begin{pmatrix} 419 \\ 1967 \\ -364 \end{pmatrix}$	$\begin{pmatrix} 430 \\ 2084 \\ -371 \end{pmatrix}$	$\begin{pmatrix} 435 \\ 2025 \\ -371 \end{pmatrix}$	$\begin{pmatrix} -5 \\ 23 \\ 17 \end{pmatrix}$	$\begin{pmatrix} 25 \\ 140 \\ 17 \end{pmatrix}$	$\begin{pmatrix} 3 \\ 0 \\ 0 \end{pmatrix}$
5	$\begin{pmatrix} 443 \\ 1982 \\ -375 \end{pmatrix}$	$\begin{pmatrix} 417 \\ 1981 \\ -342 \end{pmatrix}$	$\begin{pmatrix} 422 \\ 1967 \\ -346 \end{pmatrix}$	$\begin{pmatrix} 441 \\ 2009 \\ -347 \end{pmatrix}$	$\begin{pmatrix} -2 \\ 27 \\ 28 \end{pmatrix}$	$\begin{pmatrix} 19 \\ 124 \\ 28 \end{pmatrix}$	$\begin{pmatrix} 5 \\ 2 \\ 1 \end{pmatrix}$
6	$\begin{pmatrix} 446 \\ 1963 \\ -375 \end{pmatrix}$	$\begin{pmatrix} 432 \\ 1961 \\ -331 \end{pmatrix}$	$\begin{pmatrix} 422 \\ 1983 \\ -341 \end{pmatrix}$	$\begin{pmatrix} 452 \\ 1995 \\ -336 \end{pmatrix}$	$\begin{pmatrix} 6 \\ 32 \\ 39 \end{pmatrix}$	$\begin{pmatrix} 8 \\ 110 \\ 39 \end{pmatrix}$	$\begin{pmatrix} 6 \\ 4 \\ 1 \end{pmatrix}$
7	$\begin{pmatrix} 450 \\ 1943 \\ -375 \end{pmatrix}$	$\begin{pmatrix} 444 \\ 1973 \\ -391 \end{pmatrix}$	$\begin{pmatrix} 438 \\ 1965 \\ -330 \end{pmatrix}$	$\begin{pmatrix} 460 \\ 1990 \\ -364 \end{pmatrix}$	$\begin{pmatrix} 10 \\ 47 \\ 11 \end{pmatrix}$	$\begin{pmatrix} 0 \\ 105 \\ 11 \end{pmatrix}$	$\begin{pmatrix} 6 \\ 3 \\ 0 \end{pmatrix}$
8	$\begin{pmatrix} 453 \\ 1924 \\ -375 \end{pmatrix}$	$\begin{pmatrix} 441 \\ 1882 \\ -349 \end{pmatrix}$	$\begin{pmatrix} 450 \\ 1976 \\ -391 \end{pmatrix}$	$\begin{pmatrix} 461 \\ 1939 \\ -356 \end{pmatrix}$	$\begin{pmatrix} 8 \\ 15 \\ 19 \end{pmatrix}$	$\begin{pmatrix} 1 \\ 54 \\ 19 \end{pmatrix}$	$\begin{pmatrix} 7 \\ 2 \\ 0 \end{pmatrix}$
9	$\begin{pmatrix} 456 \\ 1904 \\ -375 \end{pmatrix}$	$\begin{pmatrix} 449 \\ 1962 \\ -354 \end{pmatrix}$	$\begin{pmatrix} 448 \\ 1884 \\ -349 \end{pmatrix}$	$\begin{pmatrix} 461 \\ 1951 \\ -355 \end{pmatrix}$	$\begin{pmatrix} 5 \\ 47 \\ 20 \end{pmatrix}$	$\begin{pmatrix} 1 \\ 66 \\ 20 \end{pmatrix}$	$\begin{pmatrix} 6 \\ 0 \\ 0 \end{pmatrix}$
10	$\begin{pmatrix} 460 \\ 1885 \\ -375 \end{pmatrix}$	$\begin{pmatrix} 456 \\ 1897 \\ -350 \end{pmatrix}$	$\begin{pmatrix} 455 \\ 1962 \\ -354 \end{pmatrix}$	$\begin{pmatrix} 462 \\ 1924 \\ -352 \end{pmatrix}$	$\begin{pmatrix} 5 \\ 77 \\ 21 \end{pmatrix}$	$\begin{pmatrix} 2 \\ 39 \\ 23 \end{pmatrix}$	$\begin{pmatrix} 6 \\ 0 \\ 0 \end{pmatrix}$

Tabelle A.5: Ergebnisse für die Objektortprädiktion der fünften Beispielsequenz für das entferntere Objekt. Die Geschwindigkeitskomponenten betragen konstant  $(3.3, 19.4, 0) \frac{mm}{Bild}$ .



Anhang B

Sequenzen

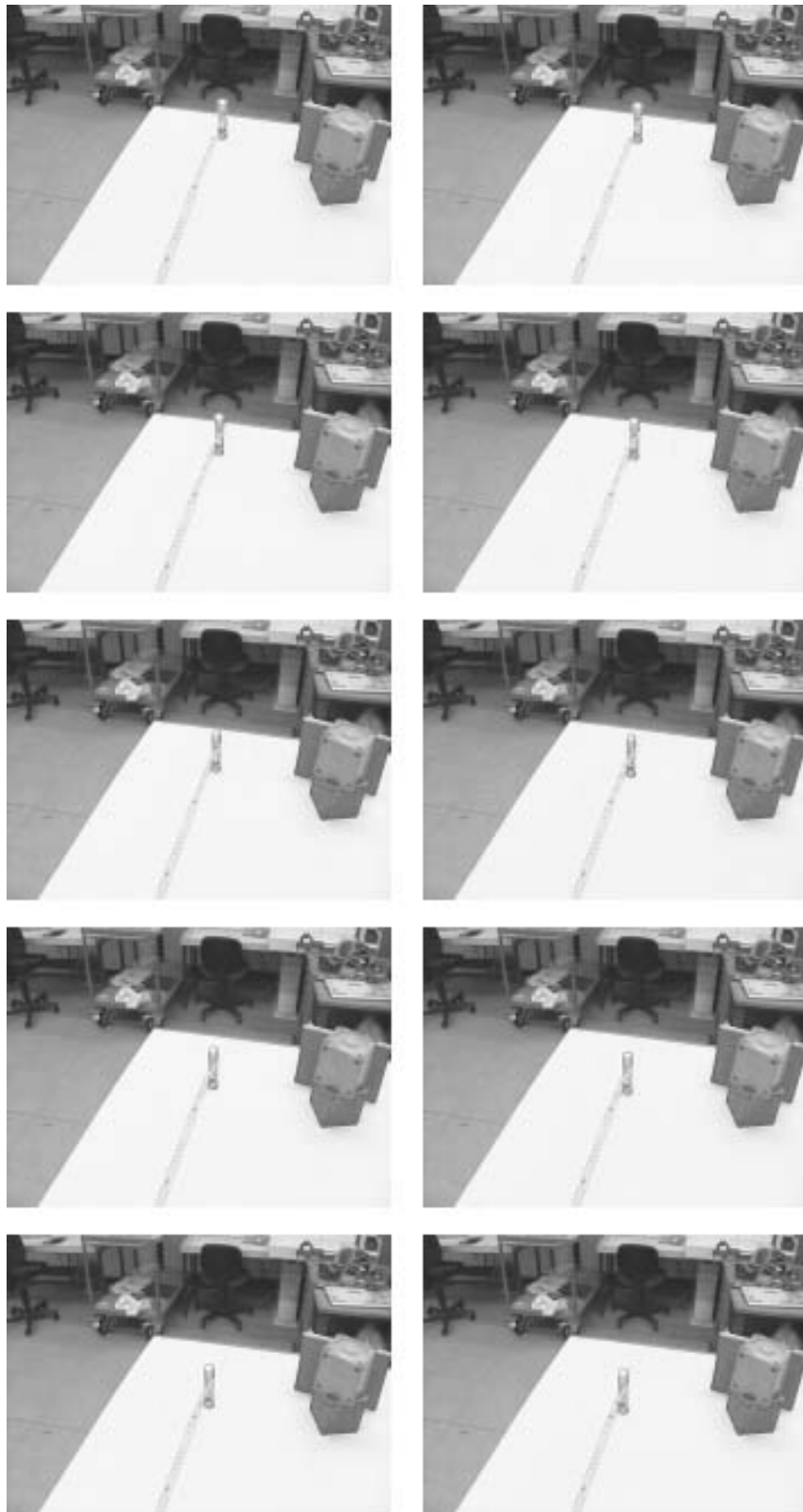


Abbildung B.1: Kamerabilder der gesamten Sequenz aus Abschnitt 5.2

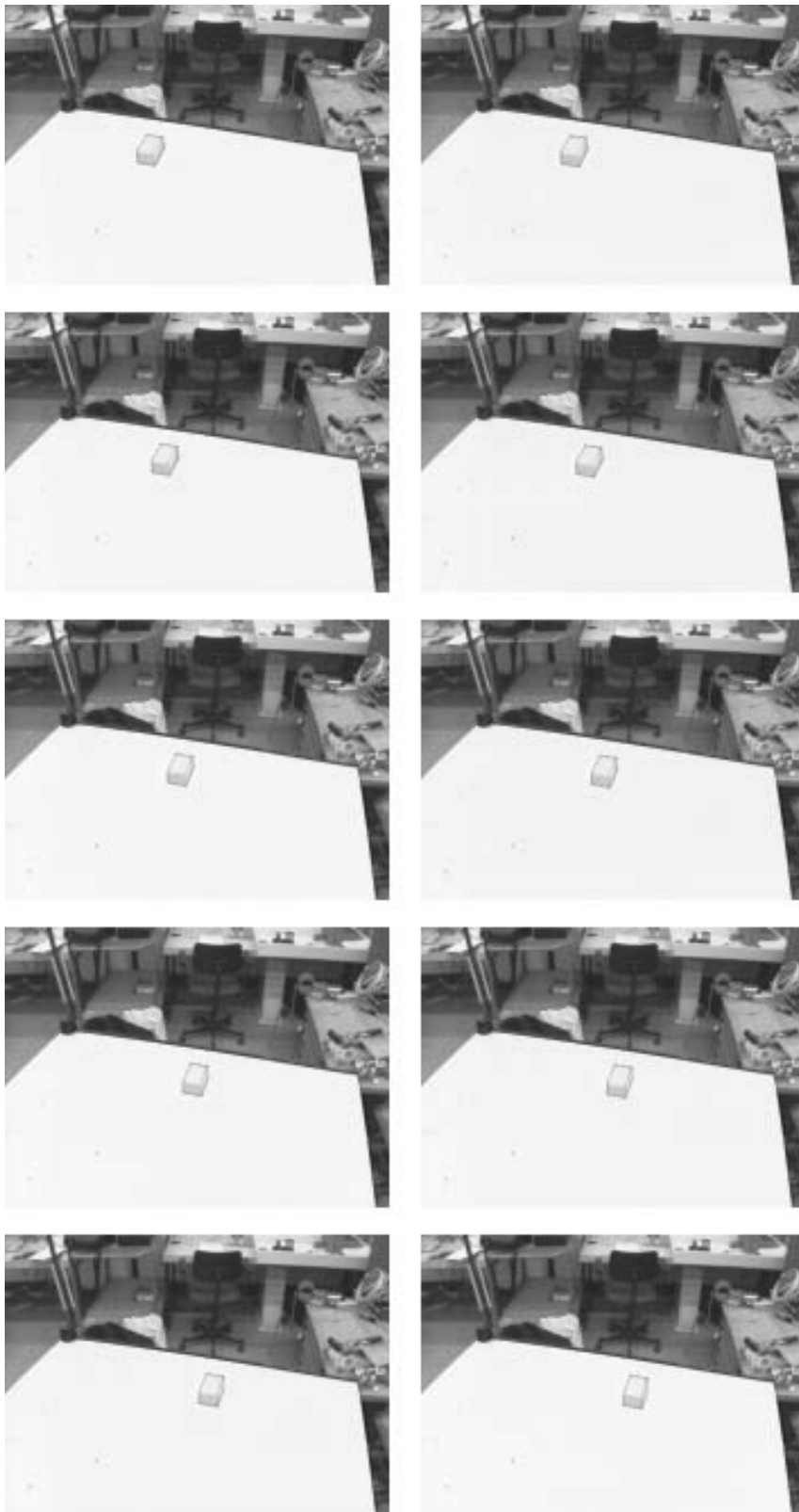


Abbildung B.2: Kamerabilder der gesamten Sequenz aus Abschnitt 5.3

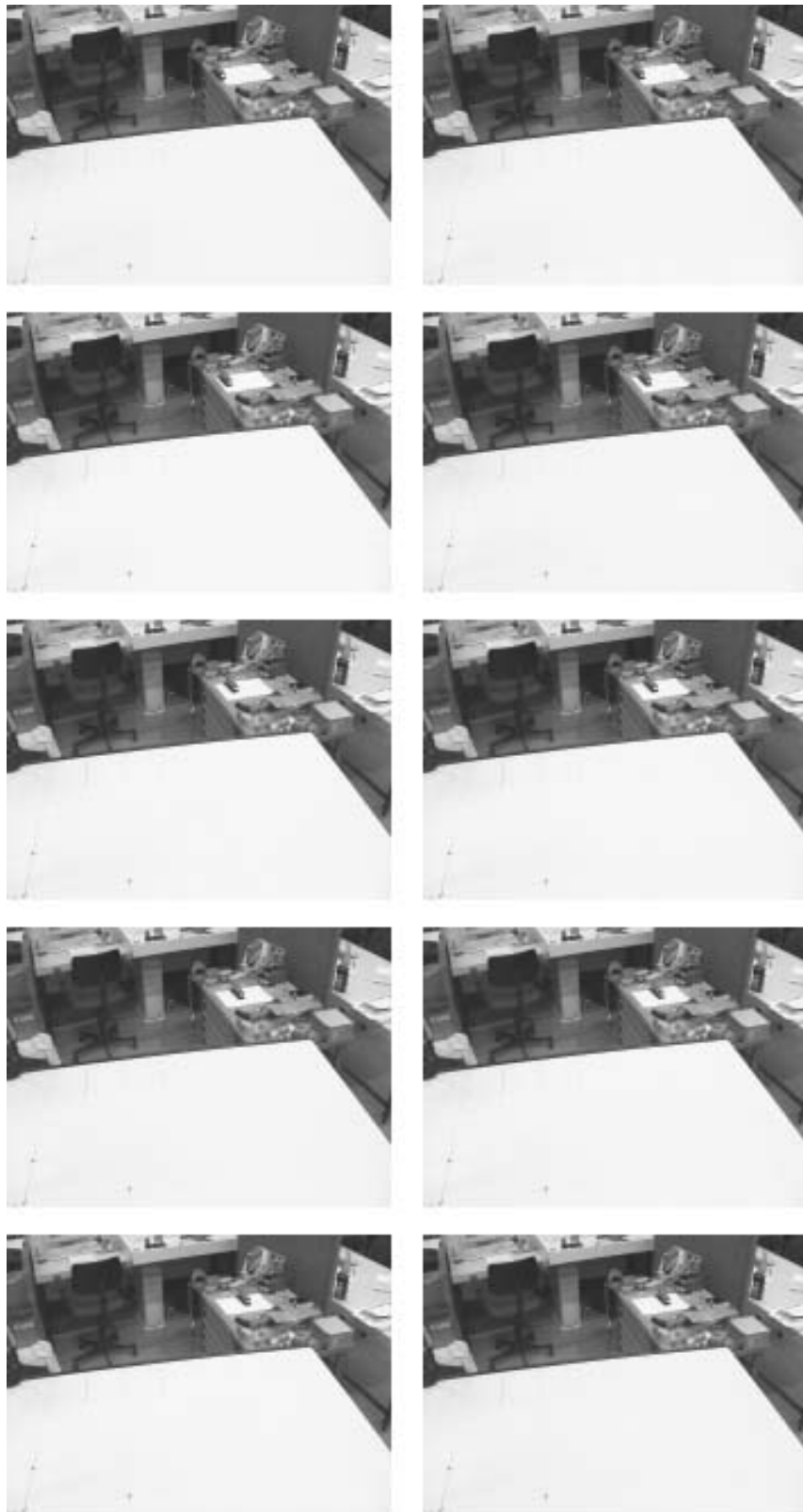


Abbildung B.3: Kamerabilder der gesamten Sequenz aus Abschnitt 5.4

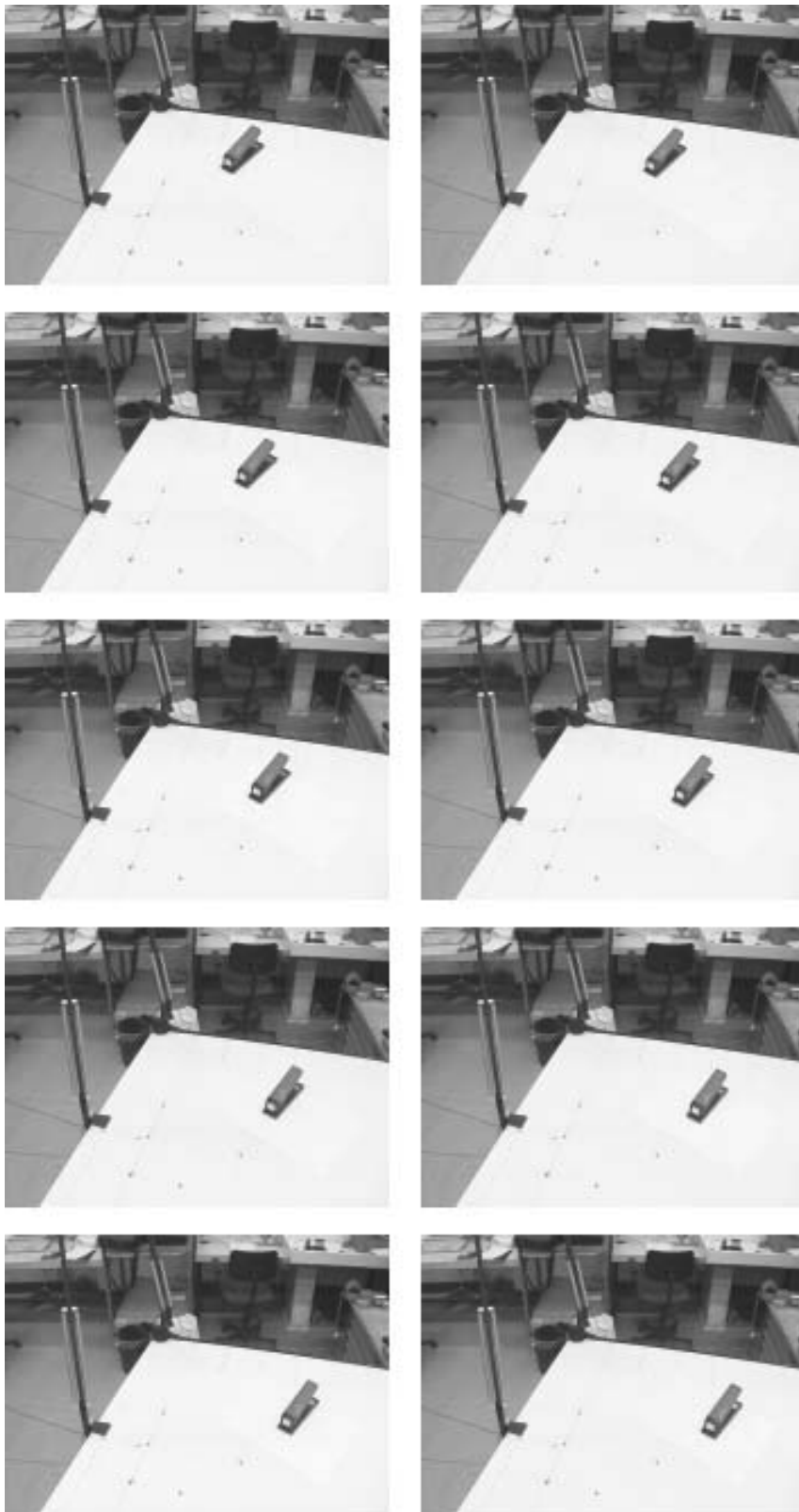


Abbildung B.4: Alle Bilder der Sequenz (A) aus Abschnitt 5.5.



Abbildung B.5: Alle Bilder der Sequenz (B) aus Abschnitt 5.5.

# Anhang C

## Technische Daten der Rechner und der Kameras

Rechnername	max
CPU	AMD Athlon 1700+ (1466MHz)
physikalischer Speicher	512MB DDR-RAM 266MHz
Betriebssystem	Suse Linux 8.0

Tabelle C.1: Rechner CoRA

Rechnername	mottek
Model	Sun Blade-100
CPU	UltraSPARC IIe 500 MHz
physikalischer Speicher	348 MB
Virtueller Speicher	1,3 GB
Betriebssystem	SunOS Release 5.8

Tabelle C.2: Rechner für Testläufe

Typ	Sony XC-999P (Pal)
Bildelemente	752 (H) x 582 (V)
Sensorgröße	6,4 x 4,8 mm
Minimale Beleuchtung	4,5 lux, F1,2 AGC: On
Sensitivität	2000 lux F5,6 AGC: Off (0dB)
Video-Ausgang	Y/C
Verschlusszeit	1/50s
Farbtemperatur	3200K (Innenraum)
Signal-Rausch-Verhältnis	> 46 dB, AGC Off
Arbeitstemperatur	0°C bis 40 °C

Tabelle C.3: Tabelle mit den vom Hersteller gegebenen Grunddaten

Kamera	rechts	links
Breite in Pixeln (Silizium)	752	752
Höhe in Pixeln (Silizium)	582	582
Breite in Pixeln (Bild)	700	700
Höhe in Pixeln (Bild)	512	512
Pixelbreite (Silizium)	0,0085106 mm	0,0085106 mm
Pixelhöhe (Silizium)	0,0082474 mm	0,0082474 mm
Pixelbreite (Bild)	0,0091429 mm	0,0091429 mm
Pixelhöhe (Bild)	0,0082474 mm	0,0082474 mm
Brennweite	6,0318443 mm	6,0373019 mm
Bildhauptpunkt x	347,64523	350,11509
Bildhauptpunkt y	274,67398	259,15769
radiale Verzerrung	0,0056723	0,0055524
Skalierung x	1,0256634	1,0229612
Fehlerbeurteilung der Kalibrierung		
Normalisierter Kalibrierungsfehler		
Mittelwert	1,16794 mm	1,07155 mm
Standardabweichung	0,779694 mm	0,621063 mm
Maximale Abweichung	7,54163 mm	4,42282 mm
Objektraum Kalibrierungsfehler		
Mittelwert	0,302013 mm	0,269844 mm
Standardabweichung	0,190145 mm	0,167753 mm
Maximale Abweichung	1,70749 mm	2,10732 mm

Tabelle C.4: Durch Kalibrierung ermittelte Kameraparameter (siehe [19])

# Literaturverzeichnis

- [1] A.G. Bors und I. Pitas. Prediction and tracking of moving objects in image sequences. *IEEE Transactions on Image Processing*, 9(8):1441–1445, August 2000.
- [2] A. Elnagar. Prediction of moving objects in dynamic environments using Kalman-Filters. *IEEE International symposium on computational intelligence in robotics and automation*, Juli/August 2001.
- [3] C.E. Smith und N.P. Papanikolopoulos. Issues and experimental results in vision-guided robotic grasping of static or moving objects. *Industrial Robot*, 25(2):134–140, 1998.
- [4] U. Frese, B. Bäuml, S. Haidacher, G. Schreiber, I. Schaefer, M. Hähle und G.Hirzinger. Off-the-shelf vision for a robotic ball catcher. *IEEE Proceedings International on intelligent robots and systems (IROS)*, Seiten 1623–1629, 2001.
- [5] G. Hommel, F. Hofmann und J. Henz. The TU Berlin high-precision sensor glove. *Forth International Scientific Conference*, 1994.
- [6] P. Smith, T. Drummond und R. Cipolla. Segmentation of multiple motions by edge tracking between two frames. In *British Machine Vision Conference*, September 2000.
- [7] C. Born und B. Völpel. Grouping bits to objects. *IRINI*, März 1995.
- [8] J. Shi und C. Tomasi. Good features to track. In *1994 IEEE Conference on Computer Vision and Pattern Recognition (CVPR'94)*, Seiten 593 – 600, 1994.
- [9] V.S. Nalwa und T.O. Binford. On detection edges. *IEEE Transactions on pattern analysis and machine intelligence*, 8(6), November 1986.
- [10] R.M. Haralick und L.G. Shapiro. *Computer and Robot Vision*, volume 1. Addison-Wesley Publishing Company, 1992.

- [11] L.G. Roberts. Machine perception of three dimensional solids. *Optical and Electrooptical Information Processing*, Seiten 159–197, 1965.
- [12] B. Jähne. *Digitale Bildverarbeitung*. Springer-Verlag, 4. Ausgabe, 1997.
- [13] J. Canny. A computational approach to edge detection. *IEEE Transactions on pattern analysis and machine intelligence*, 8(6), November 1986.
- [14] S.D. Cochran. *Surface Description from Binocular Stereo*. Doktorarbeit, University of Southern California, November 1990.
- [15] H.A. Mallot. *Sehen und die Verarbeitung visueller Information*. Vieweg-Verlag, 2. Auflage, 2000.
- [16] D. Huttenlocher, G. Klandermann und W. Rucklidge. Comparing images using the Hausdorff-distance. *IEEE Transactions on pattern analysis and machine intelligence*, 15(9):850–863, 1993.
- [17] M. Werner. *Objektverfolgung und Objekterkennung mittels der partiellen Hausdorffdistanz*. Doktorarbeit, Institut für Neuroinformatik, Ruhr-Universität Bochum, Juni 1998.
- [18] M. Khalid. Entwicklung eines Verfahrens zum visuellen Führen eines Roboterarms. Diplomarbeit, Institut für Neuroinformatik, Ruhr-Universität Bochum, 2000.
- [19] R. Tsai. A versatile camera calibration technique for high-accuracy 3d machine vision metrology using off-the-shelf TV-cameras and lenses. *IEEE Journal Robotics and Automation*, Seiten 323–344, 1987.

# Tabellenverzeichnis

A.1	Ergebnisse für die Objektortprädiktion der zweiten Beispielsequenz. . . . .	73
A.2	Ergebnisse für die Objektortprädiktion der dritten Beispielsequenz. . . . .	74
A.3	Ergebnisse für die Objektortprädiktion der vierten Beispielsequenz. . . . .	75
A.4	Ergebnisse für die Objektortprädiktion der fünften Beispielsequenz für das nähere Objekt. . . . .	76
A.5	Ergebnisse für die Objektortprädiktion der fünften Beispielsequenz für das entferntere Objekt. . . . .	77
C.1	Rechner CoRA . . . . .	85
C.2	Rechner für Testläufe . . . . .	85
C.3	Tabelle mit den vom Hersteller gegebenen Grunddaten . . . . .	86
C.4	Durch Kalibrierung ermittelte Kameraparameter . . . . .	86



# Danksagung

Zunächst möchte ich mich bei Herrn Prof. von Seelen bedanken, der mir ermöglicht hat, diese Arbeit am Institut für Neuroinformatik zu schreiben. Weiter gilt mein Dank Herrn Prof. Wieck, der sich bereit erklärte, diese Arbeit als Zweitgutachter zu bewerten.

Mein Dank gilt außerdem allen Mitarbeitern des Instituts, die mir in dem vergangenen Jahr mit Rat und Tat zur Seite standen. Besonders bedanken möchte ich mich bei dem Betreuer der Arbeit Dipl. phys. Ioannis Iossifidis, sowie bei Dr. Carsten Bruckhoff und Dipl. Ing. Christoph Theis, die mir bei der Erstellung der Arbeit in zahlreichen Diskussionen hilfreich waren.

Zu guter Letzt möchte ich mich bei meiner Lebensgefährtin, meiner Familie und meinen Freunden bedanken, die gerade in letzter Zeit sehr viel Verständnis und Geduld für mich und meine Arbeit aufgebracht haben.