Akkumulieren von Objektrepräsentationen im Wahrnehmungs-Handlungs-Zyklus

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

vorgelegt von
Marcus Ackermann

Kiel, im Januar 2000
Prof. Dr. Gerald Sommer
Betreuer: Dr. Norbert Krüger

Inhaltsverzeichnis

1 Einführung

1.1 Problemstellung und Motivation

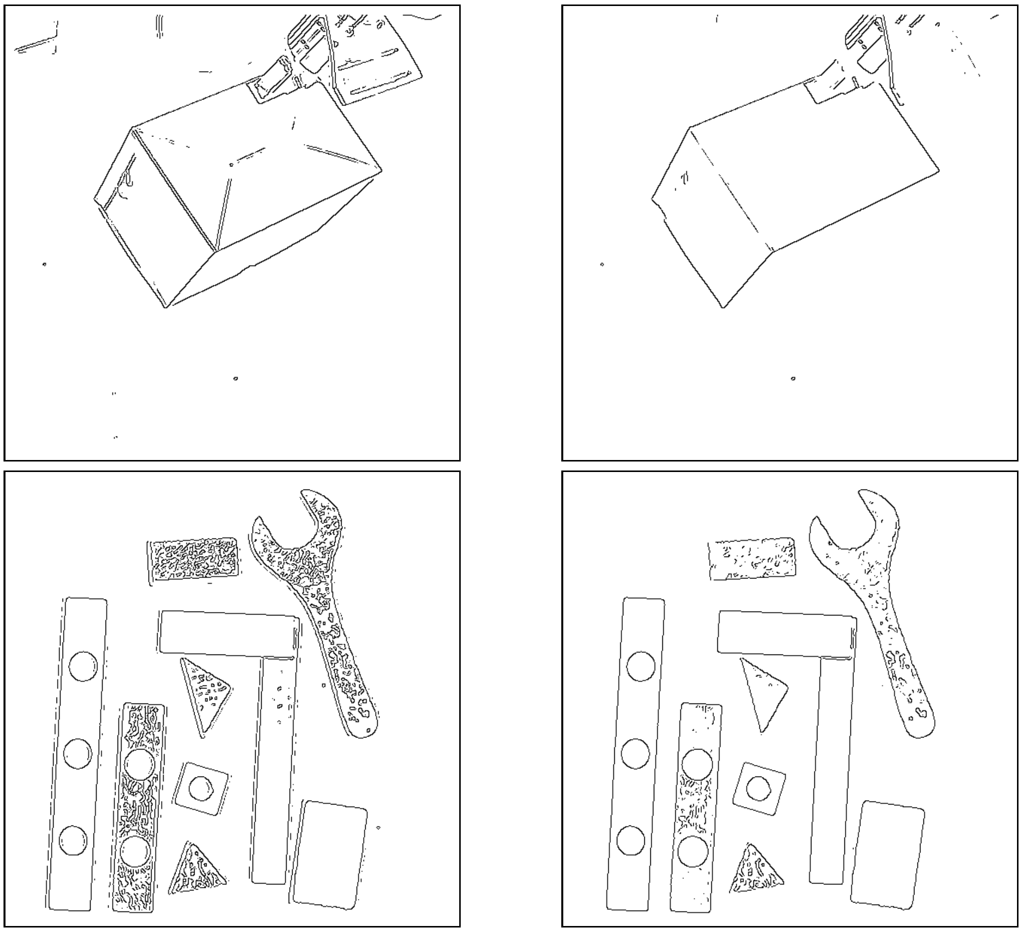
Aus Bildern gewonnene Information ist – selbst unter vereinfachten Bedingungen – in der Regel unsicher. Ein sehr einfaches Beispiel ist die Kantenextraktion basierend auf der Bestimmung des Gradientenbetrags und abschließender Anwendung eines Schwellwerts. Abbildung 1.1 zeigt zwei verschiedene Aufnahmen mit je zwei verschiedenen Schwellwerten. Es ist deutlich zu sehen, daß die Wahl eines vernünftigen Schwellwerts vom Bild abhängig und nicht allgemeingültig zu treffen ist.

In dieser Arbeit wird das Objekt mit Hilfe dreidimensionaler Liniensegmente dargestellt. Die dreidimensionale Struktur wird (kalibrierte Kameras vorausgesetzt) mit Hilfe von 2D-Korrespondenzen in Stereobildern extrahiert. Diese Tiefenschätzung ist ungenau, und trotz konstanter Beleuchtung hat auch die Orientierung des Objekts einen entscheidenden Einfluß auf die Objektrepräsentation (siehe Abbildung 1.2).

Eine auf dieser unsicheren Information basierende Entscheidung ist problematisch, und die Funktionalität eines solchen Systems ist sehr von den äußeren Gegebenheiten abhängig (Beleuchtung, Kameras, Objekt usw.). Möchte man eine Objektrepräsentation aus Bildern bzw. Bildsequenzen gewinnen, so ist es sinnvoll, dieses Objekt mit Hilfe von zuverlässiger Information zu beschreiben, die nicht sehr von den Zufälligkeiten der Aufnahmesituation oder von Schwellwerteinstellungen abhängt.

Eine Möglichkeit zur Extraktion zuverlässiger Information ist die Berücksichtigung globalerer Zusammenhänge. Natürliche Bilder und Bildsequenzen stellen hochgradig redundante Information dar, die sich aufgrund ihrer Redundanz gut komprimieren lassen [ES98]. Redundanz kann auch genutzt werden, um die Extraktion von Information aus Bildern zum Beispiel aufgrund von Konsensforderungen durch ihre Einbettung in globalere Zusammenhänge zu verbessern [El38, HPB84, Pa98]. Ein wichtiger Zusammenhang ist die raum-zeitliche Veränderung von starren Körpern bei Bewegungen. Dieser ist analytisch durch sechs Parameter explizit beschreibbar.

Das menschliche visuelle System nutzt Zusammenhänge der Objekte während einer Bewegung und akkumuliert Informationen über die Zeit. Daher ist es ihm möglich, die Wahrnehmung von Objekten durch seine Handlung zu beeinflussen und die dabei gewonnene Information für eine zukünftige Handlung nutzbar zu



niedriger Schwellwert | hoher Schwellwert

Abbildung 1.1: Einfluß des Schwellwerts auf die Kantenextraktion

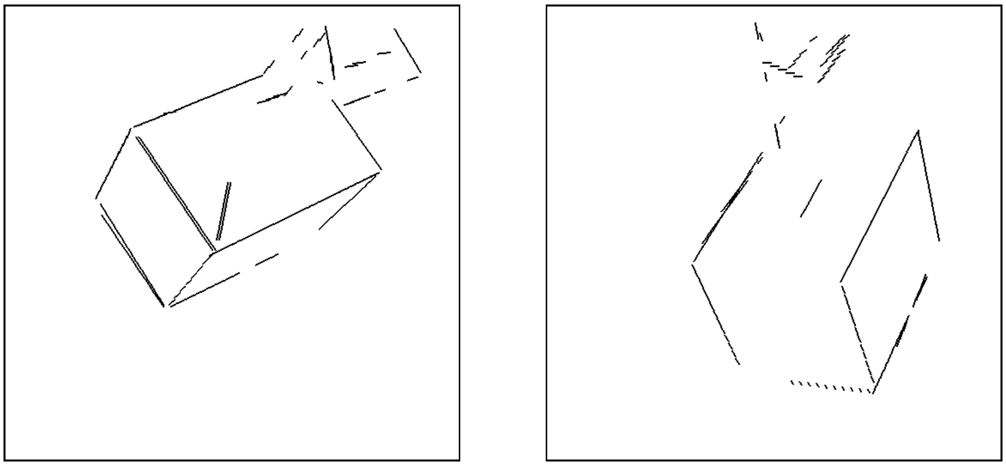


Abbildung 1.2: Objektrepräsentationen verschiedener Ansichten

machen: Wahrnehmung und Handlung sind also eng miteinander verknüpft und unterstützen sich gegenseitig. Beides findet in einem Wahrnehmungs-Handlungs-Zyklus statt [Ko92, So97]. Die Simulation dieser menschlichen Fähigkeit ist von entscheidender Bedeutung für die Formalisierung von Systemen der visuell basierten Robotik.

Gegenstand dieser Arbeit ist es, den raum-zeitlichen Zusammenhang zu formalisieren, um aus einer Sequenz unsicherer Information in Form verschiedener Objektansichten eine stabilere Objektrepräsentation zu erhalten. Information, die in einer Ansicht nicht extrahierbar ist, kann in einer anderen Ansicht extrahierbar sein. Durch Akkumulation der Repräsentationen im Laufe einer Bewegung wird Stabilität erreicht, da die Entitäten der Repräsentationen während der Bewegung bestätigt oder nicht bestätigt werden. Dabei ermöglicht die aktive Veränderung des Objekts und das Wissen um die Parameter der Veränderung die Lösung des für das erfolgreiche Akkumulieren wesentlichen Korrespondenzproblems. Im Unterschied zu modellbasierten Verfahren (z.B. [Yu91, LTC97, Wi97, HH89, Ri97]) wird es so möglich, Objektrepräsentationen aus realen Szenen zu extrahieren, anstatt sie manuell vorzugeben.

1.2 Aufbau und Ziel der Arbeit

Das aus der vorliegenden Arbeit hervorgegangene System basiert auf einem Algorithmus, der auf eine Sequenz von Stereobildpaaren angewendet wird und die jeweis aus den Bildpaaren gewonnenen Repräsentationen akkumuliert. Ein Objekt wird von einem Roboter gegriffen und dem System in verschiedenen Positionen und Orientierungen gezeigt. Die Unsicherheit der aus einzelnen Ansichten extrahierten Information wird mittels Konfidenzbewertungen modelliert und über mehrere Bilder akkumuliert. Die in den einzelnen Repräsentationen enthalten Fehler sollen also durch die Akkumulation eliminiert werden.

In einem Wahrnehmungs-Handlungs-Zyklus [Ko92, So97] wird abwechselnd eine Repräsentation extrahiert und anschließend eine Bewegung des Objekts ausgeführt, um eine neue Ansicht zu gewinnen. Die Objektrepräsentation wird nun iterativ mit der gerade extrahierten Repräsentation aktualisiert. Eine Objektrepräsentation besteht aus Liniensegmenten und wird mit kalibrierten Kameras unter Nutzung der Epipolargeometrie extrahiert. Das Korrespondenzproblem im Aktualisierungsschritt wird durch die Kenntnis der Bewegung gelöst.

Mittels der orientierungsselektiven Hough-Transformation [Pr90] werden in den Einzelbildern Linien extrahiert, die teilweise mit den Objektkanten zusammenfallen. Diese Linien werden diskretisiert und die entstehenden Liniensegmente anhand lokaler Information mit einer Konfidenz bewertet. An dieser Stelle wird das Gestalt-Prinzip Kollinearität [El38] berücksichtigt, ein Liniensegment kann nur existieren, wenn es globalen Support (durch die Hough-Transformation) sowie lokalen Support (durch die Konfidenzbewertung) hat.

Die Arbeit ist wie folgt aufgebaut: In Kapitel 2 werden die mathematischen Grundlagen erläutert, die Systemumgebung beschrieben und einige Vorüberlegungen zur Objektrepräsentation ausgeführt. Die orientierungsselektive Hough-Transformation wird in Kapitel 3 beschrieben. In einem ausführlichen Exkurs werden einige Verfahren zur Suche nach lokalen Maxima im Akkumulatorarray der Hough-Transformation untersucht und schließlich ein neues Verfahren vorgestellt. Kapitel 4 beschäftigt sich mit der Diskretisierung der Hough-Geraden und der Extraktion einer 3D-Repräsentation daraus. In Kapitel 5 wird schließlich der Akkumulationsalgorithmus vorgestellt und in Kapitel 6 ausführlich getestet. Zusammenfassung und Ausblick in Kapitel 7 beschließen diese Arbeit.

Die prinzipielle Vorgehensweise bei der Akkumulation mehrerer Objektrepräsentationen in einem Wahrnehmungs-Handlungs-Zyklus verdeutlicht das Ablaufdiagramm in Abbildung 1.3. Auf dieses Ablaufdiagramm wird im folgenden mehrfach Bezug genommen. Die Schritte Gewinnung einer Bildrepräsentation sowie Akkumulation werden in den Kapiteln 4 und 5 ausführlich beschrieben und dort mit Hilfe weiterer Ablaufdiagramme verdeutlicht.

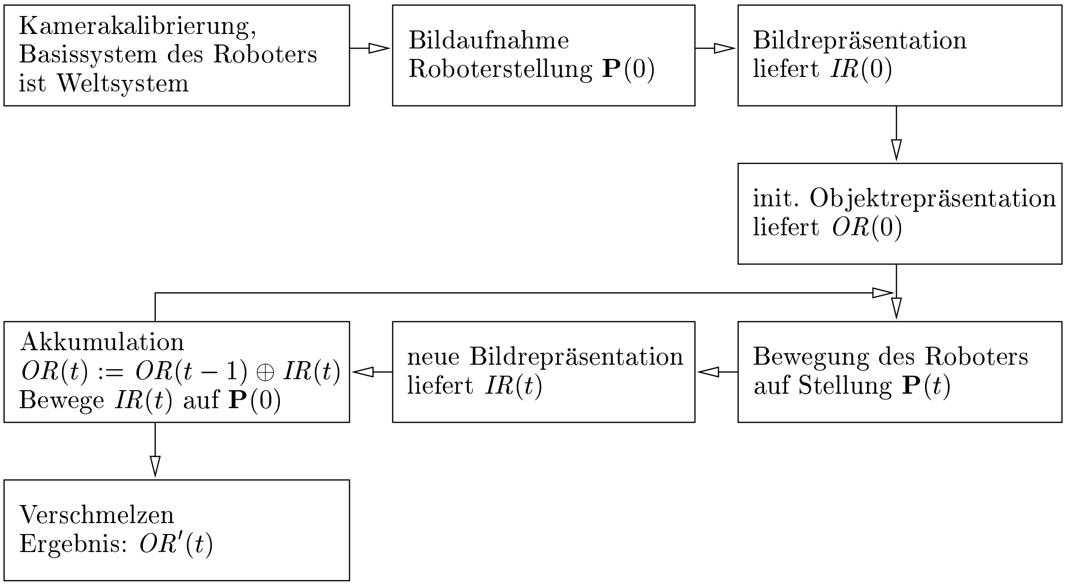
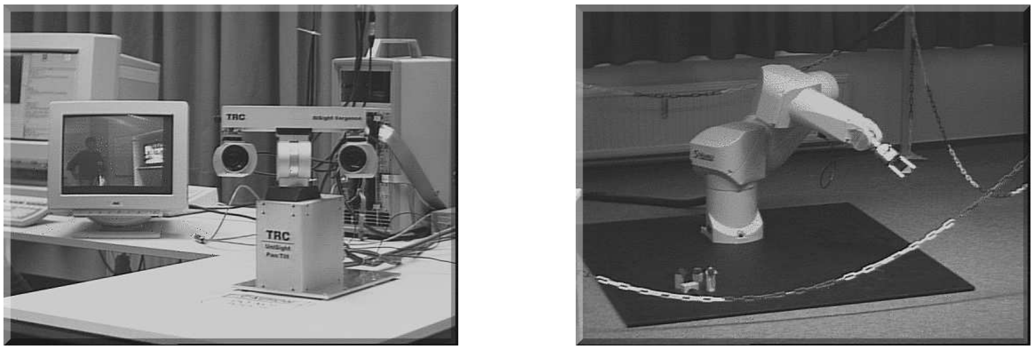


Abbildung 1.3: Vorgehen bei der Gewinnung der Objektrepräsentation

2 Systemumgebung und Grundlagen

Das Sehgreifsystem, auf dem diese Arbeit beruht und mit dem alle Versuche durchgeführt werden, besteht aus einem Roboterarm (siehe Abschnitt 2.1) und einem Stereokamerakopf mit zwei Schwarz-Weiß-CCD-Kameras (siehe Abschnitt 2.2). Die Kameras werden mit einem Standardverfahren kalibriert, um aus einem Bildpaar Tiefeninformation zu erhalten. Aufgrund der suboptimalen Wahl der Lage des Ursprungs des Weltkoordinatensystems, definiert durch ein Kalibrierungsobjekt, muß eine weitere Koordinatentransformation nachträglich bestimmt werden (Abschnitte 2.2.1 und 2.2.2). Abschnitt 2.2.3 beschreibt die Mathematik der Roboterbewegungen, um diese später auf Objektrepräsentationen anzuwenden. Um die Roboterbewegung effizient anwenden zu können, enthält Abschnitt 2.3 einige Vorüberlegungen zur Darstellung von Lage und Orientierung von Objekten im Rechner.

Insgesamt liefert die in diesem Kapitel beschriebene Hardware eine Sequenz von Stereobildpaaren. Ein Objekt, das sich im Greifer des Roboterarms befindet, wird bewegt (siehe Abbildung 2.1), die Sequenz zeigt damit das Objekt in verschiedenen Ansichten. Die zu den Ansichten gehörenden Roboterstellungen sind bekannt, so daß die starren Körperbewegungen des Objekts zwischen jeweils zwei Ansichten einfach zu bestimmen sind. Eine Beispielsequenz findet sich in Abbildung 2.2.



Der TRC-Kamerakopf — Der Stäubli RX90

Abbildung 2.1: Die verwendete Hardware

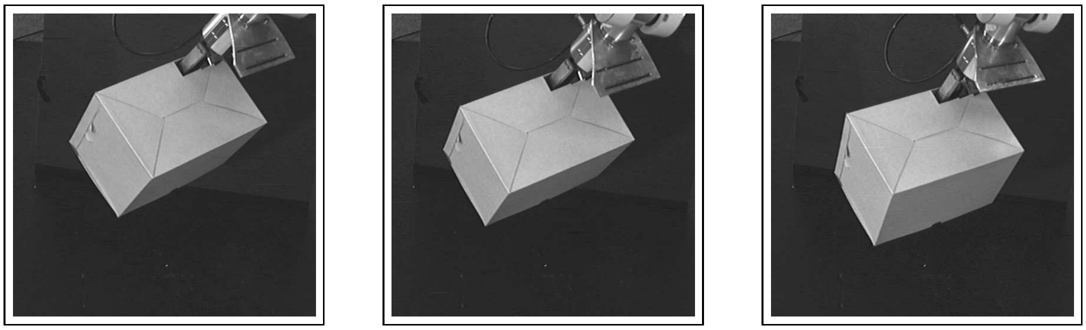
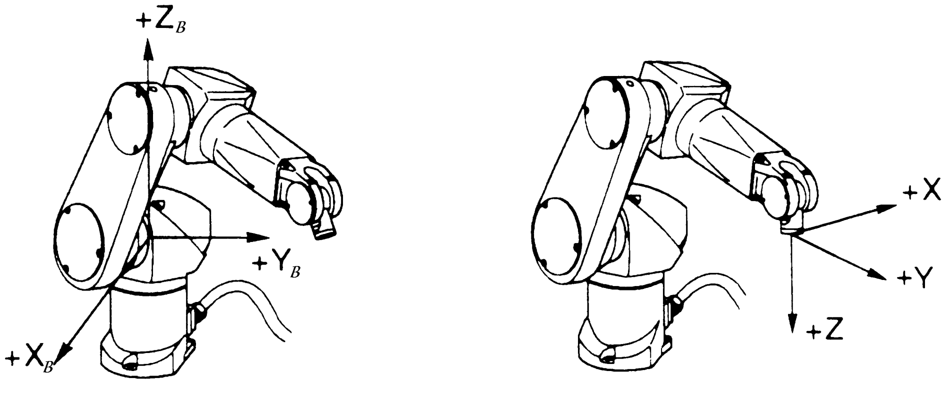


Abbildung 2.2: Beispielsequenz, jeweils linke Kamerabilder

2.1 Der Roboterarm

Der verwendete Roboterarm Typ RX90 der Firma Stäubli besitzt sechs Gelenke und damit sechs Freiheitsgrade sowie einen Greifer mit zwei Fingern, die sich öffnen und schließen lassen. Die Position des Greifers wird jedoch in dieser Arbeit nicht durch die Stellungen der Gelenke, also durch die Winkel der beiden durch ein Gelenk verbundenen Gliedmaßen, beschrieben, sondern durch Lage und Orientierung des Greifers in einem äußeren Bezugskoordinatensystem. Dieses Bezugssystem ist das Basiskoordinatensystem $\mathcal{B}$, dessen Ursprung in der Mitte des untersten Gelenks liegt (Abbildung 2.3 links). Die Position des Greifers ist nun gegeben durch die Koordinaten (X, Y, Z), gemessen in mm. Das Toolkoordinatensystem $\mathcal{T}$ (Abbildung 2.3 rechts), das die Ausrichtung des Greifers bestimmt, hat seinen Ursprung im Greifergelenk, also im Punkt (X, Y, Z) des Basiskoordinatensystems $\mathcal{B}$.



Basiskoordinatensystem $\mathcal{B}$ Toolkoordinatensystem $\mathcal{T}$

Abbildung 2.3: Koordinatensysteme des Stäubli RX90 (aus [St94])

Die Orientierung des Toolkoordinatensystems ist gegeben durch drei Winkel (α, β, γ) als ZYZ-Euler-Winkel, auch *yaw*, *pitch* und *roll* genannt. Der Steuerschrank des Roboterarms löst die Probleme der inversen Kinematik, d.h. bestimmt die Gelenk-stellungen aus der in Form des 6-Tupels $(X, Y, Z, \alpha, \beta, \gamma)$ gegebenen Greiferposition. Die Greiferposition beschreibt also den Zustand des Roboters in Form von den sechs Parametern der Koordinatentransformation $\mathbf{T}_{\mathcal{B}}^{\mathcal{T}}$, die Basis- in Toolkoordinaten transformiert.

ZYZ-Euler-Winkel bedeutet hierbei, daß das Basiskoordinatensystem durch sukzessive Rotation um die z-Achse durch den Winkel α (yaw), dann um die y-Achse durch den Winkel β (pitch) und schließlich erneut um die z-Achse durch den Winkel γ (roll) mit anschließender Translation um (X, Y, Z) in das Toolkoordinatensystem übergeht. Während die Rotationsachsen bei sukzessiven Rotationen um feste Winkel unverändert bleiben, werden bei Darstellung in Euler-Winkeln jedoch stets die Rotationsachsen für die folgenden Rotationen mitgedreht.

In festen Winkeln werden die Rotationsquaternionen aufeinanderfolgender Rotationen von links multipliziert. Die gesamte Rotation $\mathbf{q}$ in Euler-Winkeln wird nach einem Satz von Euler durch Umkehrung der Multiplikationsreihenfolge der zugehörigen Rotationsquaternionen berechnet. Eine Rotation in Euler-Winkeln entspricht also einer Rotation in festen Winkeln mit umgekehrter Reihenfolge der Rotationsachsen. Zur erschöpfenden Behandlung dieser Thematik sei auf [Cr89] verwiesen.

Seien also $\mathbf{q}_z(\alpha'), \mathbf{q}_y(\beta'), \mathbf{q}_z(\gamma')$ die zu einer Rotation $(\alpha', \beta', \gamma')$ gehörigen Rotationsquaternionen, dann ergibt sich die Rotationsquaternion $\mathbf{q}$ in ZYZ-Euler-Winkeln zu

$$\mathbf{q} = \mathbf{q}_z(\alpha') \cdot \mathbf{q}_y(\beta') \cdot \mathbf{q}_z(\gamma') \quad \text{(Multiplikation von rechts)}, \tag{2.1}$$

was einer Rotation in ZYZ-festen Winkeln um $(\gamma', \beta', \alpha')$ entspricht. Einige weitere Bemerkungen zur Verwendung von Quaternionen finden sich in Abschnitt 2.3.

2.2 Der Kamerakopf

Zur Gewinnung von 3D-Information eines Objekts werden zwei verschiedene Ansichten benötigt. Jede Ansicht ist dabei eine Projektion auf die Bildebene der jeweiligen Kamera, die durch das Lochkameramodell beschrieben wird. Sind die beiden zugehörigen Projektionsmatrizen bekannt, so kann aus zwei in beiden Projektionen korrespondierenden Punkten Tiefeninformation gewonnen werden. Die Bestimmung der Projektionsmatrix einer Kamera nennt man Kalibrierung und wird in Abschnitt 2.2.1 beschrieben.

Die in dieser Arbeit verwendeten Kameras sind zwei Schwarz-Weiß-CCD-Kameras gleichen Typs, die Bestandteil des TRC-Stereokamerakopfs sind. Dieser Kamerakopf besitzt vier mechanische Freiheitsgrade, nämlich einen Schwenkwinkel (*pan*), einen Neigungswinkel (*tilt*) sowie die Vergenzwinkel der Kameras, die die Stellung der optischen Achse der jeweiligen Kamera zur Normalposition beschreiben. Ferner läßt sich für jede Kamera die Brennweite von $11,5\ mm$ bis $69\ mm$ stufenlos einstellen sowie die Fokussierung als auch die Blendeneinstellung regeln. Da die Projektionsmatrizen abhängig von all diesen Parametern sind, werden beide Kameras zu Versuchsbeginn fest auf den Greifer des Roboterams fokussiert und bleiben von da an unverändert.

2.2.1 Kalibrierung

Die Projektionsmatrix, die die Abbildung der dreidimensionalen Welt auf ein zweidimensionales Kamerabild bestimmt, enhält sowohl die extrinsischen als auch die intrinsischen Parameter der Kamera sowie die Brennweite. Die intrinsischen Kameraparameter beschreiben den Aufbau der Kamera. Dies sind Informationen über Breite und Höhe eines Pixels (das i.a. nicht quadratisch ist) und den Schnittpunkt der optischen Achse mit der Bildebene. Die extrinsischen Parameter beschreiben Lage und Orientierung der Kamera, d.h. des optischen Zentrums, in einem wählbaren Weltkoordinatensystem $\mathcal{W}$.

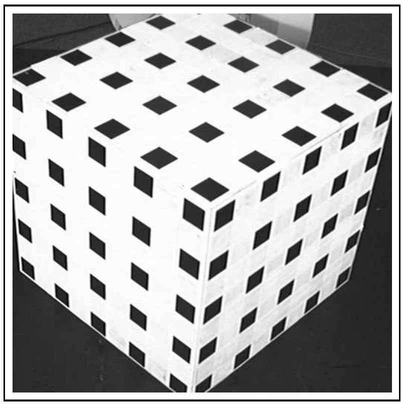


Abbildung 2.4: Kalibrierungswürfel

Faugeras [Fa93] beschreibt ein Verfahren zur Schätzung dieser Parameter mittels eines Kalibrierungsobjekts. An dieser Stelle wurde ein Würfel verwendet, auf dessen weißer Oberfläche schwarze Kästchen mit konstantem Abstand aufgeklebt sind (Abbildung 2.4). Die vordere Ecke des Würfels – etwa in der Bildmitte – ist der Ursprung des Weltkoordinatensystems $\mathcal{W}$, die drei Kanten, die in dieser Ecke zusammentreffen, stellen die Koordinatenachsen dar. Die Eckpunkte der Kästchen auf der Würfeloberfläche können mit einfachen Verfahren detektiert werden, ihre Weltkoordinaten sind aufgrund der regelmäßigen Struktur der Würfeloberfläche bekannt.

Aus mindestens vier korrespondierenden Paaren von Welt- und Bildkoordinaten können die Parameter der Projektionsmatrix durch Lösen eines linearen Gleichungssystems bestimmt werden. Da die Bilder verrauscht sind, ist es ratsam, eine wesentlich höhere Zahl an Korrespondenzpaaren zu ermitteln und die Projektionsmatrix mittels eines linearen Ausgleichsverfahrens zu bestimmen. Die Implementierung dieses Verfahrens wurde von Hahn [Ha99] vorgenommen, in seiner Arbeit finden sich ausführlichere Erläuterungen sowie die mathematischen Hintergründe der Kamerakalibrierung.

2.2.2 Bestimmung eines neuen Weltkoordinatensystems

Die Wahl des Weltkoordinatensystems während des Kalibrierungsvorgangs ist für die hier beschriebene Problemstellung suboptimal, da die Bewegung des Roboters von Position $\mathbf{P} = (X, Y, Z, \alpha, \beta, \gamma)$ zu Position $\mathbf{P}' = (X', Y', Z', \alpha', \beta', \gamma')$ relativ zum Basiskoordinatensystem des Roboters (vgl. Abschnitt 2.1) angegeben wird. Daher muß die Transformation $\mathbf{T}_{\mathcal{W}}^{\mathcal{B}}$ bestimmt werden, die Weltkoordinaten in Basiskoordinaten transformiert. Eine solche Transformation ist durch sechs Parameter eindeutig bestimmt, drei Rotations- und drei Translationsparameter: $\mathbf{T}_{\mathcal{W}}^{\mathcal{B}} = \mathbf{TR}$, wobei $\mathbf{R}$ ein Rotationsoperator und $\mathbf{T}$ ein Translationsoperator sind. Abbildung 2.5 zeigt die Beziehungen der verschiedenen Koordinatensysteme. Im Rest dieses Abschnitts wird ein einfaches Verfahrung zur Bestimmung von $\mathbf{T}_{\mathcal{B}}^{\mathcal{W}}$ abgeleitet. Der Leser, der nicht an den Einzelheiten interessiert ist, kann direkt zu 2.2.3 übergehen.

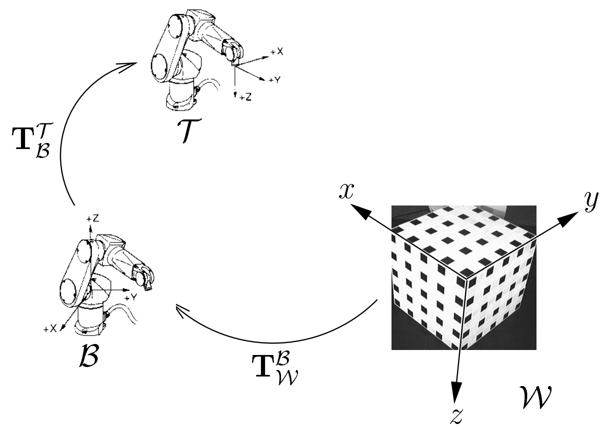


Abbildung 2.5: Die drei verwendeten Koordinatensysteme und ihre Transformationen

Die sechs Parameter können bestimmt werden, indem aus n Stereobildpaaren die Koordinaten von m beliebigen markanten Punkten manuell bestimmt und deren Weltkoordinaten berechnet werden (siehe dazu Abschnitt 4.3). Man erhält Punkte $\mathbf{p}_i^j$, wobei i die Nummer des Stereobildpaares und j die Nummer des Punktes darstellen. Für zwei Bildpaare i, i' sei $\mathbf{M}_{i,i'}$ der Operator der zugehörigen Bewegung des Greifers von $\mathbf{P}_i$ nach $\mathbf{P}_{i'}$ bezüglich $\mathcal{B}$. Dann gilt für jedes j:

$$\mathbf{p}_{i'}^j = \mathbf{T}_{\mathcal{B}}^{\mathcal{W}} \mathbf{M}_{i,i'} \mathbf{T}_{\mathcal{W}}^{\mathcal{B}} \mathbf{p}_i^j, \tag{2.2}$$

also Transformation des Punktes in Basiskoordinaten, Anwendung der Bewegung (siehe Abschnitt 2.2.3) und Rücktransformation in Weltkoordinaten. Bekannt sind hierbei $\mathbf{p}_i^j$, $\mathbf{p}_{i'}^j$ und $\mathbf{M}_{i,i'}$ für alle $i, i' \in \{1, \ldots, n\}$ und $j \in \{1, \ldots, m\}$, ferner gilt $\mathbf{T}_{\mathcal{B}}^{\mathcal{W}} = (\mathbf{T}_{\mathcal{W}}^{\mathcal{B}})^{-1}$. Die sechs Transformationsparameter können bestimmt werden, indem das folgende Minimierungsproblem bezüglich dieser sechs Transformationsparameter iterativ gelöst wird:

$$\text{Minimiere} \sum_{j=1}^{m} \sum_{i=1}^{n-1} \sum_{i'=i+1}^{n} ||\mathbf{p}_{i'}^{j} - \mathbf{T}_{\mathcal{B}}^{\mathcal{W}} \mathbf{M}_{i,i'} \mathbf{T}_{\mathcal{W}}^{\mathcal{B}} \mathbf{p}_{i}^{j}||_2 \tag{2.3}$$

Auf diese Weise wird die Summe der Fehler aller Punkte j und aller Bewegungen von i nach i', $i \neq i'$ minimiert. Der Fehler wird dabei durch den euklidischen Abstand von berechnetem und tatsächlichem Weltpunkt repräsentiert.

2.2.3 Roboterbewegungen

Die hier erwähnten Koordinatensysteme sind das durch die Kalibrierung bestimmte Weltkoordinatensystem $\mathcal{W}$ sowie die beiden Roboterkoordinatensysteme $\mathcal{B}$ und $\mathcal{T}$, das Basis- und das Toolkoordinatensystem (siehe Abbildung 2.5).

Im folgenden wird das Basiskoordinatensystem $\mathcal{B}$ als Weltkoordinatensystem verwendet. Die aus der Tiefenschätzung berechneten Koordinaten werden sofort mit der in Abschnitt 2.2.2 bestimmten Koordinatentransformation $\mathbf{T}_{\mathcal{W}}^{\mathcal{B}}$ in Basiskoordinaten transformiert. Damit enfällt die bei jeder Roboterbewegung durchzuführende Transformation $\mathbf{T}_{\mathcal{W}}^{\mathcal{B}}$ und die anschließende Rücktransformation $\mathbf{T}_{\mathcal{B}}^{\mathcal{W}}$. Auf die explizite Erwähnung dieser Transformationen wird im folgenden verzichtet.

Eine Roboterbewegung $\mathbf{M}_{\mathbf{P}}^{\mathbf{P}'}$ von $\mathbf{P}$ nach $\mathbf{P}'$ läßt sich nun beschreiben als Rücktransformation von $\mathbf{P}$ in Nullstellung $\mathbf{P}_N = (0,0,0,0,0,0)$ und anschließende Transformation nach $\mathbf{P}'$. Sei $\mathbf{M}_{\mathbf{P}}$ die Transformation, die Koordinaten von der Nullstellung des Roboters nach $\mathbf{P}$ transformiert, dann ist $\mathbf{M}_{\mathbf{P}}$ für $\mathbf{P} = (X, Y, Z, \alpha, \beta, \gamma)$ bestimmt durch

$$\mathbf{M}_{\mathbf{P}} := \mathbf{T}_{\mathbf{P}} \mathbf{R}_{\mathbf{P}} = \mathbf{T}_{\mathbf{P}} \mathbf{R}_z(\alpha) \mathbf{R}_y(\beta) \mathbf{R}_z(\gamma), \tag{2.4}$$

wobei $\mathbf{R}_{\mathbf{P}}$ der Rotationsoperator ist, der die Rotation (α, β, γ) in ZYZ-Euler-Winkeln ausführt, und $\mathbf{T}_{\mathbf{P}}$ der Translationsoperator.

Somit ergibt sich die gesamte Roboterbewegung von $\mathbf{P}$ nach $\mathbf{P}'$

$$\mathbf{M}_{\mathbf{P}}^{\mathbf{P}'} := \mathbf{M}_{\mathbf{P}'} \mathbf{M}_{\mathbf{P}}^{-1} = \mathbf{T}_{\mathbf{P}'} \mathbf{R}_{\mathbf{P}'} \mathbf{R}_{\mathbf{P}}^{-1} \mathbf{T}_{\mathbf{P}}^{-1}. \tag{2.5}$$

Der Operator $\mathbf{M}_{\mathbf{P}}$ ist in Quaternionenform gegeben durch

$$\mathbf{p}' = \mathbf{q}_{\mathbf{P}} \mathbf{p} \mathbf{q}_{\mathbf{P}}^{-1} + \mathbf{t}_{\mathbf{P}} = \mathbf{q}_z(\alpha) \mathbf{q}_y(\beta) \mathbf{q}_z(\gamma) \mathbf{p} \mathbf{q}_z^{-1}(\gamma) \mathbf{q}_y^{-1}(\beta) \mathbf{q}_z^{-1}(\alpha) + \mathbf{t}_{\mathbf{P}}, \tag{2.6}$$

wobei $\mathbf{p}$ ein zu transformierender Punkt, $\mathbf{q}$ Rotationsquaternionen und $\mathbf{t}$ eine Translationsquaternion sind. Für den Operator $\mathbf{M}_{\mathbf{P}}^{\mathbf{P}'}$ ergibt sich somit

$$\mathbf{p}'' = \mathbf{q}_{\mathbf{P}'} \mathbf{q}_{\mathbf{P}}^{-1} (\mathbf{p} - \mathbf{t}_{\mathbf{P}}) \mathbf{q}_{\mathbf{P}} \mathbf{q}_{\mathbf{P}'}^{-1} + \mathbf{t}_{\mathbf{P}'}. \tag{2.7}$$

2.3 Vorüberlegungen zur Objektrepräsentation

Die bisher in diesem Kapitel beschriebenen Transformationen liefern die formale Beschreibung der Objektbewegung durch den Roboter. Diese Beschreibung von Bewegungen wird benötigt, um die Lage und Orientierung einer Objektrepräsentation in eine Darstellung bezüglich einer initialen Lage und Orientierung zu überführen.

Da die Darstellung der Roboterbewegung in Matrizenform jedoch aus der Multiplikation von acht Matrizen besteht und dieses Problem schlecht konditioniert ist, wurden bei der Implementation Quaternionenmultiplikationen zur Rotation und Quaternionenadditionen zur Translation verwendet. Die Quaternionendarstellung hat bessere numerische Eigenschaften und bietet auch Geschwindigkeitsvorteile. [Se90] bietet eine ausführliche Einführung in die Mathematik der Quaternionen und stellt auch ein Verfahren zur Umrechnung von Rotationsmatrizen in Rotationsquaternionen und umgekehrt vor, das numerisch gut konditioniert ist. Auf die Äquivalenz der beiden Darstellungen und die numerischen Vorteile der Quaternionendarstellung sei an dieser Stelle besonders hingewiesen.

Über die Struktur der Repräsentation von Objekten im Rechner wurde noch keine Entscheidung getroffen, dies geschieht in Abschnitt 4.1. Da die Kanten von Objekten repräsentiert werden sollen, werden für die Repräsentation Informationen über die Position im Raum und die Orientierung einer Kante benötigt. Die folgende Diskussion soll die Entscheidung der Repräsentation von Position und Orientierung erleichtern, indem die durch die Hardware festgelegte Repräsentation von Bewegungen durch Rotationen in ZYZ-Euler-Winkeln und anschließende Translation berücksichtigt wird.

Die Position im Raum wird durch die 3D-Koordinaten eines Punktes repräsentiert. Wird ein Punkt $\mathbf{p}$ bewegt, so erhält man in Quaternionendarstellung

$$\mathbf{p}' = \mathbf{q}\mathbf{p}\mathbf{q}^{-1} + \mathbf{t}, \qquad (2.8)$$

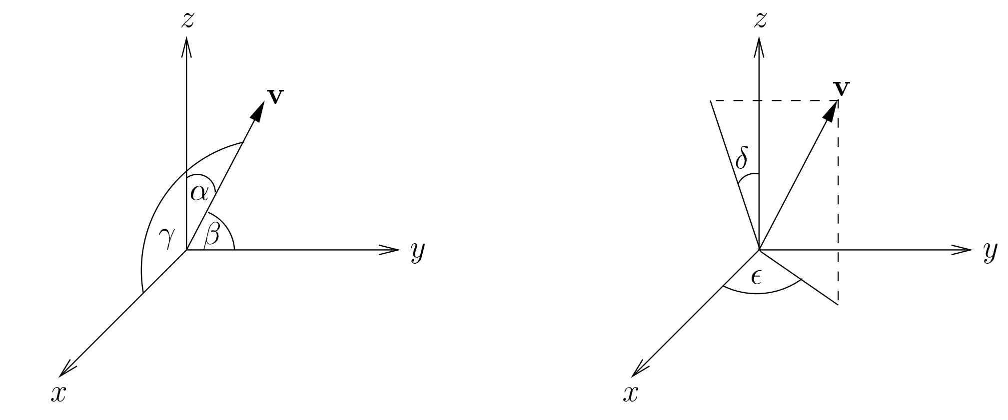
wobei $\mathbf{q}$ die entsprechende Rotationsquaternion und $\mathbf{t}$ der Translationsvektor in Form einer rein vektoriellen Quaternion sind. Bei der Repräsentation von Orientierungen, die üblicherweise durch mehrere Winkel erfolgt, kann jedoch vermutet werden, daß durch geschickte Wahl der Repräsentation auf Quaternionenmultiplikationen zur Berechnung der transformierten Orientierung verzichtet werden kann. Die Hoffnung ist, daß sich die Winkel lediglich additiv verändern.

Eine einfache Beurteilung von drei möglichen Darstellungen für Orientierungen ergibt:

1. Darstellung durch die Winkel der Orientierungsachse zu den drei Koordinatenachsen, dabei ist ein Winkel redundant (Abbildung 2.6 links). Bei jeder Rotation in ZYZ-Euler-Winkeln verändern sich alle Winkel, jedoch nicht additiv. Dies läßt vermuten, daß die Darstellung der rotierten Orientierung für alle Winkel kompliziert ist und somit eine Vereinfachung gegenüber der Durchführung von Quaternionenmultiplikationen nicht erzielt wird.

2. Darstellung durch die Winkel der Projektionen der Orientierungsachse auf zwei Ebenen, gegeben durch die Koordinatenachsen (Abbildung 2.6 rechts). Die Eigenschaften dieser Darstellung sind denen aus 1. ähnlich.

3. Darstellung in sphärischen Koordinaten (Abbildung 2.7) durch den Winkel ψ der Projektion der Orientierungsachse in die xy-Ebene zur x-Achse (Azimut) und den Winkel ϕ der Orientierungsachse zur z-Achse (Polarwinkel). Bei Drehung um die z-Achse bleibt ϕ konstant, ψ verändert sich additiv, bei Drehung um die y-Achse verändern sich beide Winkel. Da es bei der ZYZ-Darstellung nur eine Drehung um die y-Achse gibt, aber zwei um die z-Achse, könnte die gesamte Rotation eine einfache Darstellung haben.



Drei Winkel zu den Koordinatenachsen Projektionen auf zwei Ebenen

Abbildung 2.6: Mögliche Parametrisierungen der Orientierung

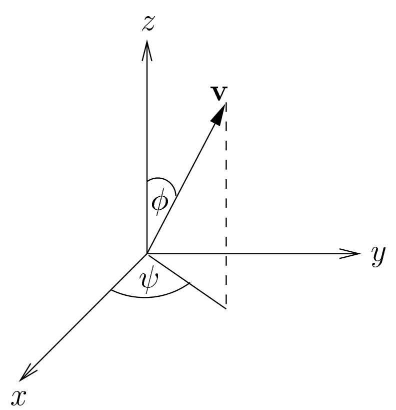


Abbildung 2.7: Sphärische Koordinaten

Die Darstellung der Orientierung $\mathbf{v}$ in sphärischen Koordinaten (ϕ, ψ) in Form eines Einheitsvektors ergibt ([ZK95]):

$$\mathbf{v} = \begin{pmatrix} \sin\phi\cos\psi \\ \sin\phi\sin\psi \\ \cos\phi \end{pmatrix} \tag{2.9}$$

Ferner gilt $0 \leq \phi < \pi$, $0 \leq \psi < 2\pi$. Die folgende Rechnung bestimmt die Winkel ϕ' und ψ' von $\mathbf{v}'$ nach Anwendung einer Rotation in ZYZ-Euler-Winkeln durch die Winkel (α, β, γ).

Durch sukzessive Rotation von $\mathbf{v}$ um die z-Achse durch den Winkel γ, des Ergebnisvektors $\mathbf{v}'$ um die y-Achse durch den Winkel β und schließlich des Ergebnisvektors $\mathbf{v}''$ um die z-Achse durch den Winkel α erhält man den rotierten Vektor $\mathbf{v}'''$, aus dem sich ϕ' und ψ' analytisch bestimmen lassen zu

$$\phi' = \arccos(\cos\phi\cos\beta - \sin\phi\cos(\psi+\gamma)\sin\beta) \tag{2.10}$$

$$\psi' = \begin{cases} \arcsin\left(\frac{\sin\phi\sin(\psi+\gamma)}{\sin\phi''}\right) + \alpha & \text{falls Gleichung 2.18 erfüllt} \\ \pi - \arcsin\left(\frac{\sin\phi\sin(\psi+\gamma)}{\sin\phi''}\right) + \alpha & \text{sonst} \end{cases} \tag{2.11}$$

Diese Darstellung erfordert tatsächlich weniger Rechenaufwand als die Berechnung des Vektors $\mathbf{v}$ aus den Winkeln ϕ und ψ, die Rotation dieses Vektors und zuletzt die Berechnung der Winkel ϕ' und ψ' aus dem rotierten Vektor $\mathbf{v}'$.

Schrittweise Drehung des Vektors $\mathbf{v}$ nacheinander um γ, β und α ergibt:

$$\mathbf{v}' := \mathbf{R}_z(\gamma)\cdot\mathbf{v} = \begin{pmatrix} \cos\gamma & -\sin\gamma & 0 \\ \sin\gamma & \cos\gamma & 0 \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} \sin\phi\cos\psi \\ \sin\phi\sin\psi \\ \cos\phi \end{pmatrix} \tag{2.12}$$

$$= \begin{pmatrix} \sin\phi(\cos\psi\cos\gamma - \sin\psi\sin\gamma) \\ \sin\phi(\cos\psi\sin\gamma + \sin\psi\cos\gamma) \\ \cos\phi \end{pmatrix} \tag{2.13}$$

$$= \begin{pmatrix} \sin\phi\cos(\psi+\gamma) \\ \sin\phi\sin(\psi+\gamma) \\ \cos\phi \end{pmatrix} \tag{2.14}$$

Es folgt die Drehung um β:

$$\mathbf{v}'' := \mathbf{R}_y(\beta)\cdot\mathbf{v}' = \begin{pmatrix} \cos\beta & 0 & \sin\beta \\ 0 & 1 & 0 \\ -\sin\beta & 0 & \cos\beta \end{pmatrix} \cdot \begin{pmatrix} \sin\phi\cos(\psi+\gamma) \\ \sin\phi\sin(\psi+\gamma) \\ \cos\phi \end{pmatrix} \tag{2.15}$$

$$= \begin{pmatrix} \sin\phi\cos(\psi+\gamma)\cos\beta + \cos\phi\sin\beta \\ \sin\phi\sin(\psi+\gamma) \\ \cos\phi\cos\beta - \sin\phi\cos(\psi+\gamma)\sin\beta \end{pmatrix} \tag{2.16}$$

Zur Vereinfachung lassen sich an dieser Stelle Winkel ϕ'', ψ'' berechnen, so daß gilt

$$\mathbf{v}'' = \begin{pmatrix} \sin\phi''\cos\psi'' \\ \sin\phi''\sin\psi'' \\ \cos\phi'' \end{pmatrix} \tag{2.17}$$

Man erhält ein Gleichungssystem mit den beiden Unbekannten ϕ'', ψ'':

$$\begin{aligned}
\sin\phi''\cos\psi'' &= \sin\phi\cos(\psi+\gamma)\cos\beta + \cos\phi\sin\beta && (2.18)\\
\sin\phi''\sin\psi'' &= \sin\phi\sin(\psi+\gamma) && (2.19)\\
\cos\phi'' &= \cos\phi\cos\beta - \sin\phi\cos(\psi+\gamma)\sin\beta && (2.20)
\end{aligned}$$

Aus Gleichung (2.20) läßt sich ϕ'' bestimmen zu

$$\phi'' = \arccos(\cos\phi\cos\beta - \sin\phi\cos(\psi+\gamma)\sin\beta), \tag{2.21}$$

wegen $0 \leq \phi'' < \pi$ gibt es keine zweite Lösung $2\pi - \phi''$.

Falls $\phi'' = 0$, dann ist ψ'' beliebig, denn es handelt sich hierbei um eine Darstellung der z-Achse. Ansonsten ergibt sich aus Gleichung (2.19):

$$\begin{aligned}
\psi''_1 &= \arcsin\left(\frac{\sin\phi\sin(\psi+\gamma)}{\sin\phi''}\right) && (2.22)\\
\psi''_2 &= \pi - \arcsin\left(\frac{\sin\phi\sin(\psi+\gamma)}{\sin\phi''}\right) && (2.23)
\end{aligned}$$

Anhand Gleichung (2.18) läßt sich bestimmen, ob $\psi'' = \psi''_1$ oder $\psi'' = \psi''_2$.

Die letzte Drehung um α ergibt:

$$\begin{aligned}
\mathbf{v}''' := \mathbf{R}_z(\alpha)\cdot\mathbf{v}'' &= \begin{pmatrix} \cos\gamma & -\sin\gamma & 0 \\ \sin\gamma & \cos\gamma & 0 \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} \sin\phi''\cos\psi'' \\ \sin\phi''\sin\psi'' \\ \cos\phi'' \end{pmatrix} && (2.24)\\
&= \begin{pmatrix} \sin\phi''(\cos\psi''\cos\alpha - \sin\psi''\sin\alpha) \\ \sin\phi''(\cos\psi''\sin\alpha + \sin\psi''\cos\alpha) \\ \cos\phi'' \end{pmatrix} && (2.25)\\
&= \begin{pmatrix} \sin\phi''\cos(\psi''+\alpha) \\ \sin\phi''\sin(\psi''+\alpha) \\ \cos\phi'' \end{pmatrix} && (2.26)
\end{aligned}$$

Hieraus erhält man die Darstellungen (2.10) und (2.11) für ϕ' und ψ'.

3 Extraktion von Bildprimitiva

Eine Entscheidung, durch welche Entitäten ein Objekt repräsentiert wird, wurde noch nicht getroffen. In Abschnitt 2.3 wurde jedoch bereits eine Darstellung durch Lage und Orientierung im Raum untersucht. Eine solche Darstellung läßt eine einfache Bewegung einer Objektrepräsentation zu und ermöglicht die Akkumulation mehrerer Objektrepräsentationen im Wahrnehmungs-Handlungs-Zyklus, da zwei Repräsentationen so bewegt werden können, daß sie das Objekt in derselben Ansicht darstellen.

Welches Verfahren zur Extraktion von Bildprimitiva genutzt wird, wird in Abschnitt 3.1 diskutiert. Ab Abschnitt 3.2 wird das gewählte Verfahren vorgestellt, verbessert und untersucht, Experimente in Abschnitt 3.7 und eine Zusammenfassung in Abschnitt 3.8 beschließen das Kapitel.

3.1 Wahl des Verfahrens

Grundlage jeglicher Verfahren zur Extraktion von Objektberandungen ist die Annahme, daß 3D-Oberflächendiskontinuitäten korrespondierende Grauwertkanten im Bild haben. Es stellt sich nun allerdings die Frage, welche Bildprimitiva für eine Objektrepräsentation zugrunde gelegt werden. Beschränkt man sich auf polyedrische Objekte, so liegt es nahe, die Objektberandung mit Liniensegmenten, die die Eckpunkte verbinden, zu beschreiben. Für diese Arbeit werden als Elemente der Objektrepräsentation stückweise Objektkanten beschreibende Liniensegmente verwendet. Benötigt wird dazu ein Verfahren, das Objektkanten als Bildprimitiva extrahiert.

Lokale Verfahren sind jedoch sensitiv gegenüber Rauschen, so daß die Festlegung von Anfangs- und Endpunkt eines Liniensegments aufgrund der Lokalität des Verfahrens sehr problematisch ist. Beim Matching in einem Stereobildpaar ist aber eine hohe Genauigkeit bei der Extraktion in beiden Bildern unabdingbar.

Allgemein geeigneter ist die Nutzung eines globalen Verfahrens wie z.B. der Hough-Transformation [Ho62], die diejenigen Geraden bestimmt, die mit im Bild vorhandenen Kanten übereinstimmen. Die Hough-Transformation ist sehr robust und weder sensitiv gegenüber Rauschen noch gegenüber Okklusion. Der Nachteil dieses globalen Verfahrens ist der Verlust der Lokalität: Ergebnis der Transformation ist eine

Menge von Geraden, auf denen die Berandung polyedrischer Objekte liegt. Wo aber genau auf der Geraden die Berandungen liegen, ist nicht bekannt. Dieses Problem läßt sich allerdings lösen, indem man die gewonnene Geradendarstellung mit dem Bild vergleicht und die signifikanten Segmente der Geraden durch Vergleich mit lokaler Information bestimmt. Diese Mischung aus einem globalen und einem lokalen Verfahren ist also eine vielversprechende Variante, um aus einem Bild die Berandung eines Objekts zu extrahieren.

Die Hough-Transformation (Abschnitt 3.2) transformiert das Problem der Suche nach Geraden im Bild auf das einfacher zu lösende Problem der Suche nach lokalen Maxima in einem zweidimensionalen Array. In Abschnitt 3.3 wird die Kantenextraktion beschrieben. Die an sich vor der Hough-Transformation durchgeführte Binarisierung des Grauwertbildes wird erst im Anschluß an die Hough-Transformation beschrieben, denn die gewählte Form der Binarisierung hängt von den Erfordernissen der Transformation ab. Abschnitt 3.4 und 3.5 enthalten wesentliche Verbesserungen der Hough-Transformation, Abschnitt 3.6 beschreibt ausführlich die Suche nach lokalen Maxima im Akkumulatorarray.

3.2 Die Hough-Transformation

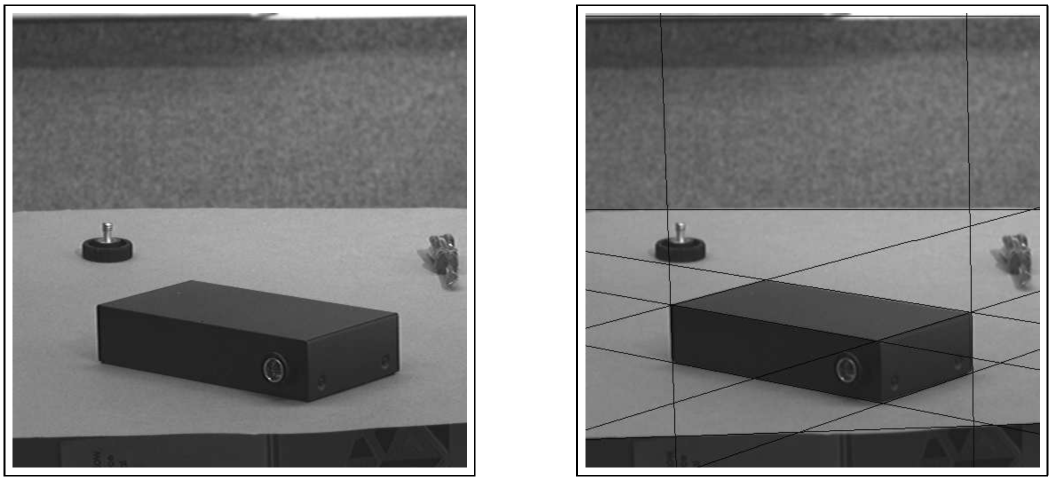
Eines der meistbenutzten Verfahren zur Geradendetektion ist die Hough-Transformation, die schon 1962 von P.V.C. Hough in den USA zum Patent angemeldet wurde [Ho62]. Die Hough-Transformation gehört zu den sogenannten Votierungsverfahren und dient der Erkennung von Strukturen in Binärbildern, die durch eine implizite Funktion beschreibbar sind. Sie stellt eine Beziehung zwischen einem binarisierten Grauwertbild und dem Raum aller Parameter der impliziten Funktion her und schätzt somit die Parameter der Struktur. Die Hough-Transformation ist also nicht nur geeignet, Geraden zu finden, sondern auch beispielsweise Kreise oder Ellipsen. Da es in dieser Arbeit nur darum geht, Geraden aus Bildern zu extrahieren, werden die Ausführungen auf diesen Fall beschränkt. Eine ausführlichere Abhandlung des allgemeinen Falls findet sich in [DH72, Ja95]. Abbildung 3.1 zeigt das Ergebnis der Hough-Transformation, angewendet auf ein Bild, das auch Ausgangsbild für viele der folgenden Abbildungen ist.

Zuerst muß also eine geeignete Parameterdarstellung $\mathbf{f_p}$ in impliziter Form von Geraden bezüglich eines Bildpunktes $\mathbf{p} = (x, y)$ gefunden werden. Die allgemein übliche Darstellung

$$\mathbf{f_p}(m, c) = mx + c - y = 0 \tag{3.1}$$

mit den Parametern Steigung m und y-Achsenabschnitt c ist ungeeignet denn eine senkrechte Gerade läßt sich nicht beschreiben. Es bietet sich vielmehr die Polarform einer Geraden an:

$$\mathbf{f_p}(r, \theta) = x \cos\theta + y \sin\theta - r = 0. \tag{3.2}$$

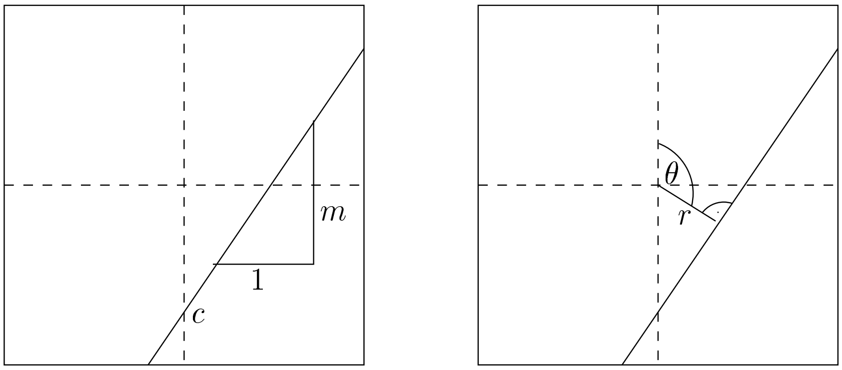


Ausgangsbild

mit der Hough-Transformation gefundene Geraden

Abbildung 3.1: Funktion der Hough-Transformation

Die Polarform stellt eine Gerade durch deren kürzesten Abstand r zum Ursprung (Bildmitte) und den Winkel θ der Normalen zur y-Achse dar (Abbildung 3.2).



Darstellung durch Steigung m und y-Achsenabschnitt c

Darstellung in (r, θ)-Polarform

Abbildung 3.2: Parameterdarstellungen von Geraden

Die Hough-Transformation arbeitet nun wie folgt: Jeder Punkt $\mathbf{p} = (x, y)$ im Binärbild $B := [0, \ldots, X-1] \times [0, \ldots, Y-1]$ votiert für alle Parametersätze (r, θ), die ihn erzeugt haben können, also für alle Geraden, die durch $\mathbf{p}$ gehen. Derjenige Parametersatz, für den die meisten Bildpunkte votiert haben, ist der *Sieger* und somit der die Gerade am besten beschreibende Parametersatz (siehe Abbildung 3.3).

Formal läßt sich die Hough-Transformation wie folgt beschreiben: Für jeden Bildpunkt $\mathbf{p} = (x, y)$ und für jeden Parametersatz (r, θ) erhöhe $A[r, \theta]$ um eins, falls gilt

$\mathbf{f_p}(r, \theta) = 0$ (die Gerade (r, θ) geht durch $\mathbf{p}$). Es gilt also

$$A[r, \theta] := \sum_{\mathbf{p} \in B} f(\mathbf{p}, r, \theta) \tag{3.3}$$

mit

$$f(\mathbf{p}, r, \theta) := \begin{cases} 1 & \text{falls } \mathbf{f_p}(r, \theta) = 0 \\ 0 & \text{sonst} \end{cases} \tag{3.4}$$

Die lokalen Maxima in A liefern nun die Parametersätze für die im Bild gefundenen Geraden (zur Maximumsuche siehe Kapitel 3.6).

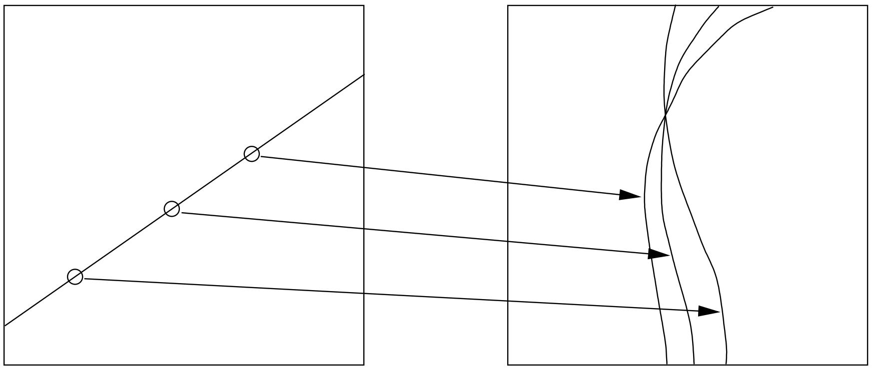


Abbildung 3.3: Verdeutlichung der Votierung. Jeder Bildpunkt (links) votiert für alle möglichen Parametersätze, die ihn erzeugt haben können (Sinusodialkurven im rechten Bild), im Idealfall einer exakten Geraden schneiden sich alle diese Kurven in genau einem Punkt, nämlich dem Punkt der Parameter der Geraden, für den dann auch die meisten Punkte votiert haben.

In der Praxis werden die Votierungen in einem Akkumulatorarray gespeichert, dessen Ausmaße sich an den Wertebereichen für r und θ orientieren. Es gilt $0° \leq \theta < 180°$ und $-\frac{e}{2} \leq r \leq +\frac{e}{2}$ mit $e := \sqrt{X^2 + Y^2}$ als Länge der Bilddiagonalen, also der maximale mögliche Abstand einer Geraden zum Ursprung. Das Akkumulatorarray A wird nun diskretisiert, wobei zwei Faktoren s_r und s_θ gewählt werden können, die die Abtastung festlegen, so daß sich die Ausmaße von A mit $n_\theta = s_\theta \cdot 180$ und $n_r = s_r \cdot e$ ergeben. Je größer diese Faktoren sind, desto genauer werden die Parameter abgetastet, desto größer wird aber auch der Rechenaufwand. Mit $s_r = s_\theta = 1$ erreicht man einen guten Kompromiß zwischen Genauigkeit und Rechenaufwand.

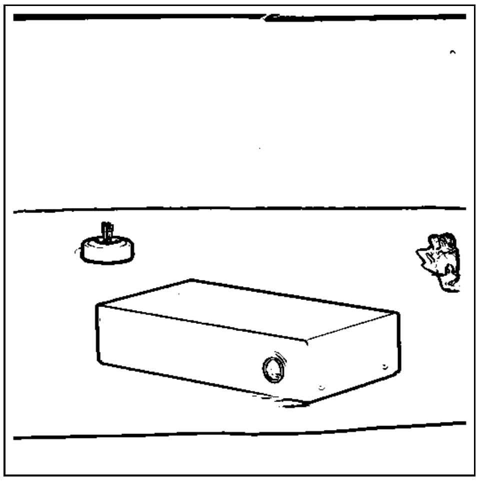
Bei der Implementation dieses Verfahrens muß beachtet werden, daß auch wirklich jede Akkumulatorzelle $A[r, \theta]$ um eins erhöht wird, für die gilt $\mathbf{f_p}(r, \theta) = 0$. Die durch den Parametersatz (r, θ) beschriebene Gerade geht also durch $\mathbf{p}$. Ein einfaches Verfahren ist es, für jeden Bildpunkt $\mathbf{p}$ und für jedes $0° \leq \theta < 180°$ das zugehörige r zu berechnen, die Werte (r, θ) gemäß den Faktoren s_r und s_θ zu runden und in das Akkumulatorarray einzutragen. Hat nun eine Sinusodialkurve lokal eine Steigung > 1, so gibt es Zellen im Akkumulatorarray, die inkrementiert werden müßen, auf diese Weise allerdings ausgelassen werden, da die Abtastung in θ-Richtung zu grob ist. Formal geschieht folgendes: Seien $\theta_1 = \theta$ und $\theta_2 = \theta + \frac{1}{s_\theta}$ im Akkumulatorarray benachbarte Werte, dann lassen sich die zugehörigen r_1 und r_2 berechnen mit $|r_2 - r_1| > \frac{1}{s_r}$, die Akkumulatorzellen $A[r_1, \theta_1]$ und $A[r_2, \theta_2]$ sind also nicht benachbart. Diese Lücke ist jedoch sehr klein und kann durch lineare Interpolation geschlossen werden.

Der Vorteil der Hough-Transformation liegt nun darin, daß sie sehr robust ist gegenüber Rauschen, Okklusion oder Lücken in der Kante durch schlechte Kantenextraktion. Dies begründet sich damit, daß die Hough-Transformation die gesuchten Strukturen nicht lokal im Bild durch Betrachtung einzelner Pixel und deren Nachbarn sucht, sondern vielmehr global die Votierungen akkumuliert. Fehlen also Punkte in einer Kante, so spiegelt sich dies lediglich in der Höhe des lokalen Maximums im Akkumulatorarray wider, während ein lokales Verfahren, das nur Bildpunkte und deren Umgebung betrachtet, Probleme hätte, diese Lücke zu schließen. Allerdings geht bei der Hough-Transformation die Lokalität verloren. Es ist im Nachhinein aus dem Akkumulatorarray nicht mehr feststellbar, welche Bereiche im Binärbild für eine Gerade votiert haben. Damit ist also nicht feststellbar, welche Segmente der Geraden eine Kante repräsentieren und welche nicht.

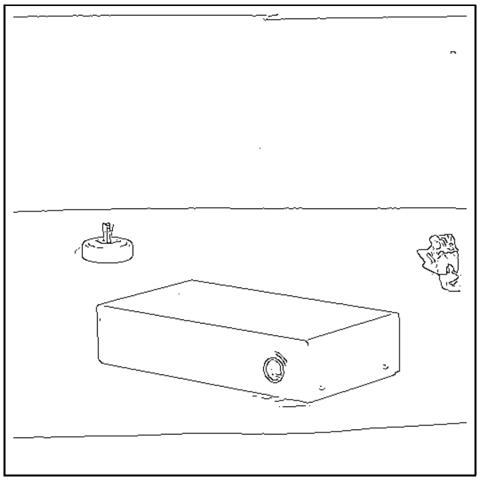
Lösen läßt sich dieses Problem aber einfach durch Projektion der Geraden auf das Ausgangsbild bzw. das daraus erzeugte Kantenbild, das der Hough-Transformation als Eingabe diente. Außerdem erschwert der Verlust der Lokalität noch die Maximumsuche, denn es müssen signifikante Maxima von nicht signifikanten, durch Rauschen verursachte Maxima unterschieden werden. In Abschnitt 3.6 wird dieses Problem ausführlich diskutiert.

3.3 Kantenextraktion

Die Eingabe der Hough-Transformation ist ein binarisiertes Grauwertbild, welches die Grauwertkanten enthält. Eine solche Binarisierung läßt sich einfach durch Berechnung des Gradientenbetrags für jeden Bildpunkt und anschließende Schwellwertbildung erreichen, allerdings hat dies den Nachteil, daß die so bestimmten Objektkanten meist eine Breite von mehr als nur einem Pixel haben, denn Grauwertkanten weisen einen glatten und keinen sprunghaften Grauwertanstieg auf. Dies wirkt sich nachteilig auf die Hough-Transformation aus, denn eine Kante von beispielsweise drei Pixel Breite entspricht drei parallelen Geradensegmenten, die jeweils ein eigenes Maximum im Parameterraum erzeugen. Ergebnis sind drei benachbart liegende Maxima, so daß eine exakte Lokalisierung des signifikanten Maximums nicht möglich ist. Aus diesem Grund wird an die Schwellwertbildung noch die Unterdrückung von Nichtmaxima [Ja95] angeschlossen, die nur diejenigen Kantenpunkte übrig läßt, die in Richtung des ansteigenden Gradienten den höchsten Betrag haben, also der eigentlichen signifikanten Kante entsprechen (siehe Abbildung 3.4).



ohne Unterdrückung von Nichtmaxima



mit Unterdrückung von Nichtmaxima

Abbildung 3.4: Wirkung der Unterdrückung von Nichtmaxima

Zuerst wird also für jeden Bildpunkt $\mathbf{p}$ des Ausgangsbildes der Gradient in x- und y-Richtung berechnet. Dafür haben sich die Sobel-Operatoren

$$S_x := \frac{1}{8}\begin{pmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{pmatrix} \quad \text{und} \quad S_y := \frac{1}{8}\begin{pmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{pmatrix} \tag{3.5}$$

bewährt, jedoch beträgt der Richtungsfehler des Sobel-Operators – also die Differenz zwischen der aus den Antworten der Sobel-Operatoren berechneten Richtungsinformation und der tatsächlichen Richtung des Gradienten – bis zu 5° [Sch96]. Scharr hat in [Sch96] den Sobel-Operator leicht modifiziert und bewiesen, daß die modifizierten Sobel-Operatoren

$$S'_x := \frac{1}{32}\begin{pmatrix} 3 & 0 & -3 \\ 10 & 0 & -10 \\ 3 & 0 & -3 \end{pmatrix} \quad \text{und} \quad S'_y := \frac{1}{32}\begin{pmatrix} -3 & -10 & -3 \\ 0 & 0 & 0 \\ 3 & 10 & 3 \end{pmatrix} \tag{3.6}$$

optimal zur Bestimmung der Gradientenorientierung sind. Der Winkelfehler ist deutlich niedriger, die Standardabweichung beträgt 0.161° im Gegensatz zu 2.20° beim Sobel-Operator. Diese Verbesserung macht sich allerdings nicht bei der Kantenextraktion bemerkbar, sondern vielmehr bei den folgenden Erweiterungen der Hough-Transformation, die sehr von der Genauigkeit der Bestimmung des Gradientenwinkels abhängen.

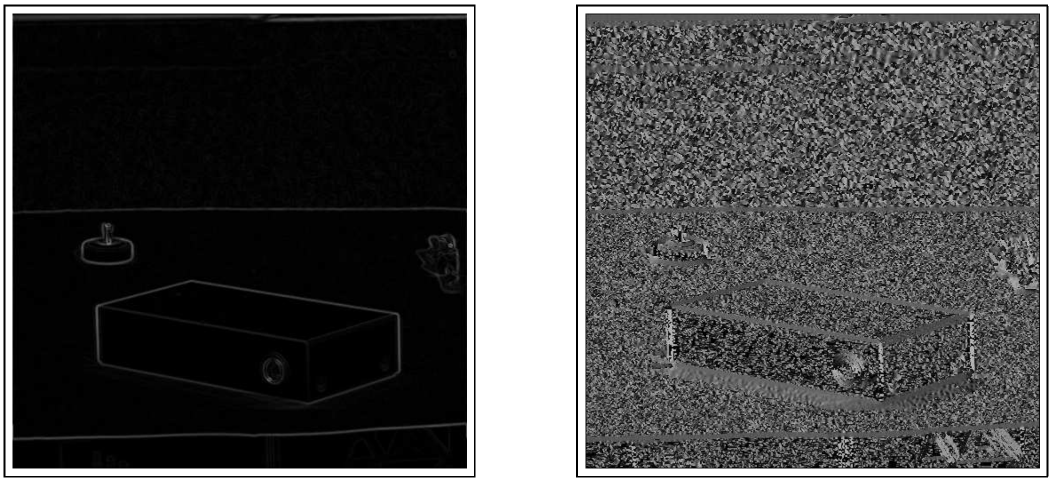
Nun ergibt sich der Winkel des Gradienten aus den Filterantworten $R_x = B * S'_x$ und $R_y = B * S'_y$ zu

$$\theta_{\mathbf{p}} := \begin{cases} \arctan\left(\frac{R_y}{R_x}\right) & \text{falls } R_x \neq 0 \\ 0° & \text{falls } R_x = 0 \end{cases} \tag{3.7}$$

und der Betrag des Gradienten zu

$$G_{\mathbf{p}} := \sqrt{R_x'^2 + R_y'^2}. \tag{3.8}$$

Abbildung 3.5 zeigt das Gradientenbetrags- und das Gradientenorientierungsbild. Die Struktur des Ausgangsbildes aus Abbildung 3.1 ist in beiden Bildern deutlich zu erkennen.



Gradientenbetragsbild　　Gradientenorientierungsbild

Abbildung 3.5: Gradientenbetrags- und Gradientenorientierungsbild

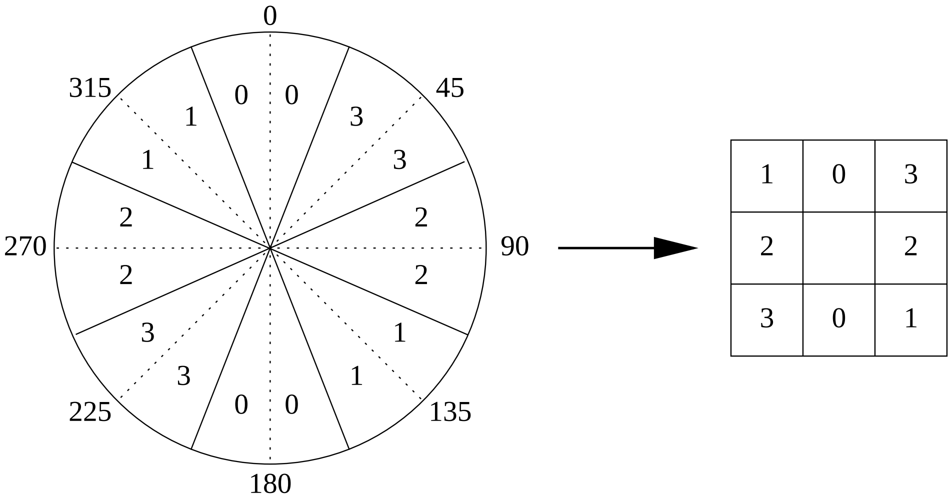


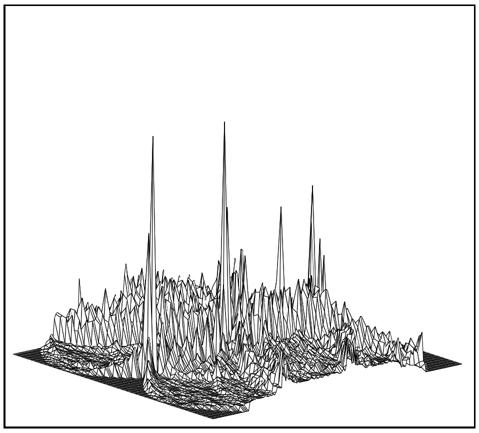
Abbildung 3.6: Suchrichtung bei der Unterdrückung von Nichtmaxima. Es gibt vier Sektoren, die mit der 8er-Nachbarschaft korrespondieren; eine Gerade durch das Zentrum verläuft durch genau einen Sektor. Die Winkelangaben erfolgen in Grad. (aus [Ja95])

Nichtmaxima, also Bildpunkte, deren Gradientenbetrag zwar über einem Schwellwert η liegt, aber die entlang des Gradienten nicht die jeweiligen Punkte mit dem größten Gradientenbetrag sind, werden unterdrückt, damit nur noch diejenigen Kantenpunkte übrig bleiben, die lokale Maxima entlang des Gradienten sind (Abbildung 3.4). Zur Überprüfung, ob es sich bei einem gegebenen Kantenpunkt um ein Nichtmaximum handelt, wird der Gradientenbetrag mit den beiden Nachbarn in Richtung des Gradienten verglichen. Welche Nachbarn in der 8er-Nachbarschaft dies sind, bestimmt sich nach dem Gradientenwinkel. Abbildung 3.6 zeigt, nach welcher Vorschrift diese Nachbarn bestimmt werden. Hat nun einer der beiden Nachbarn einen höheren Gradientenbetrag, so handelt es sich um ein Nichtmaximum, und der Kantenpunkt wird gelöscht.

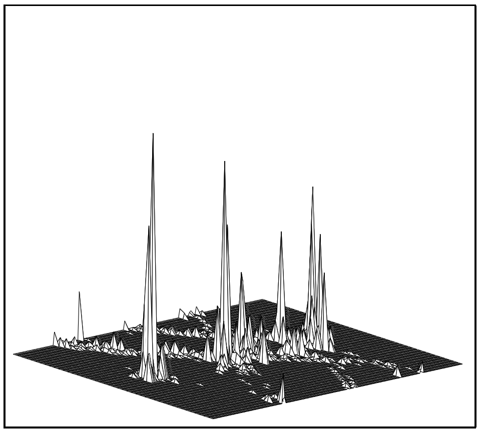
Die Wahl des Schwellwerts η hat Einfluß auf die Qualität des Ergebnisses. Im allgemeinen gute Ergebnisse liefert $\eta = 8$.

3.4 Die orientierungsselektive Hough-Transformation

Nach der Durchführung der Hough-Transformation kann nun im Akkumulatorarray nach den Maxima gesucht werden. Bei genauerer Betrachtung des Parameterraumes, also des Akkumulatorarrays, erkennt man jedoch, daß sich keine scharfen Peaks ergeben, sondern sehr breite Maxima und ein starkes Rauschen (siehe Abbildung 3.7). Dies läßt sich jedoch durch eine einfache Modifikation verbessern, nämlich die orientierungsselektive Hough-Transformation [Pr90].



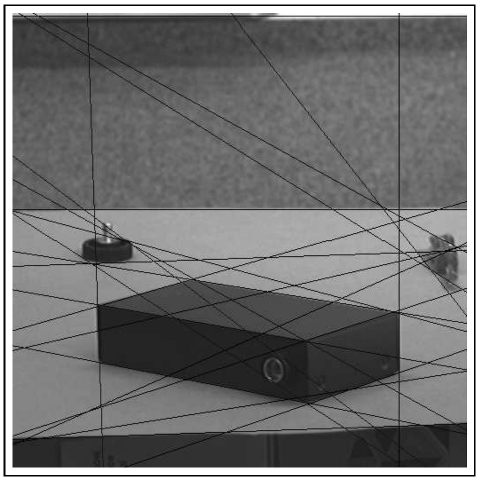
Standard-Hough-Transformation



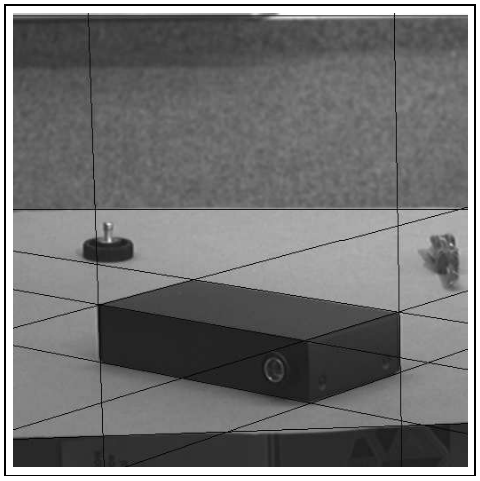
orientierungsselektive Hough-Transformation mit $\Delta\theta = 5°$

Abbildung 3.7: 3D-Darstellung des Akkumulatorarrays

Bei der Standard-Hough-Transformation werden alle Parametersätze in das Akkumulatorarray eingetragen, die einen Punkt erzeugt haben könnten. Soll allerdings dieser Punkt $\mathbf{p}$ auf einer Geraden g liegen, so muß die Orientierung $\theta_{\mathbf{p}}$ des Gradienten mit der Orientierung θ_g der Geraden übereinstimmen. Es kann also von der



Standard-Hough-Transformation



orientierungsselektive Hough-Transformation mit $\Delta\theta = 5°$

Abbildung 3.8: Ergebnis der Standard- und der orientierungsselektiven Hough-Transformation

Bedeutung der Orientierung des Gradienten und deren geometrischer Interpretation bezüglich der Geraden Gebrauch gemacht werden. Demnach braucht also nur noch zu dem gegebenen Gradientenwinkel $\theta_{\mathbf{p}}$ das zugehörige r berechnet und in das Akkumulatorarray eingetragen zu werden. Durch Rauschen im Grauwertbild ist der berechnete Gradientenwinkel natürlich nicht exakt, so daß ein Winkeltoleranzbereich $\Delta\theta$ in beiden Richtungen eingeführt werden muß, damit für den berechneten Gradientenwinkel $\theta_{\mathbf{p}}$ alle Winkel im Intervall

$$I(\theta_{\mathbf{p}}, \Delta\theta) := [\theta_{\mathbf{p}} - \Delta\theta, \theta_{\mathbf{p}} + \Delta\theta] \tag{3.9}$$

bearbeitet werden. Bei der Wahl von $\Delta\theta$ ist zu beachten, daß die Robustheit des Verfahrens abnimmt, je kleiner $\Delta\theta$ ist, da die Gradientenorientierung nicht exakt berechnet ist. Andererseits wird das Akkumulatorarray umso verrauschter, je größer $\Delta\theta$ ist. Je nach Anwendung hat sich in der Praxis $2° \leq \Delta\theta \leq 10°$ bewährt. Als Ergebnis weist das Akkumulatorarray deutlich schärfere Peaks und ein geringeres Rauschen auf, was Abbildung 3.7 verdeutlicht. Abbildung 3.8 enthält die aus den Akkumulatorarrays aus Abbildung 3.7 extrahierten Geraden und zeigt deutlich, daß die orientierungsselektive Hough-Transformation (OHT) ein wesentlich besseres Ergebnis liefert als die Standard-Hough-Transformation (SHT).

Die Hough-Transformation aus Gleichung 3.3 lautet nun:

$$A[s_r r, s_\theta \theta] := \sum_{\mathbf{p} \in B} f(\mathbf{p}, r, \theta) \tag{3.10}$$

mit

$$f(\mathbf{p}, r, \theta) := \begin{cases} 1 & \text{falls } \mathbf{f_p}(r, \theta) = 0 \wedge \theta_{\mathbf{p}} \in I \\ 0 & \text{sonst} \end{cases} \tag{3.11}$$

Die orientierungsselektive Hough-Transformation bietet allerdings nicht nur einen qualitativen Vorteil, sondern ist auch noch wesentlich effizienter als das Standardverfahren, da nicht der gesamte Winkelbereich von 180° bearbeitet wird, sondern nur der viel kleinere Winkeltoleranzbereich $2 \cdot \Delta\theta$. Das Berechnen der Gradientenwinkel erzeugt keinen zusätzlichen Aufwand, sie werden bei der Kantenextraktion schon berechnet und können gespeichert werden. Die orientierungsselektive Hough-Transformation mit $\Delta\theta = 90°$ entspricht der Standard-Hough-Transformation, denn $I(\theta, 90°) = [0°, 180°[$ für alle θ, und die Bedingung $\theta_{\mathbf{p}} \in I$ ist damit immer wahr.

Bei der Implementation dieses Verfahrens tritt ein Problem auf, das sowohl hier als auch bei der Maximumsuche beachtet werden muß. Der Parameterraum ist in θ-Richtung zyklisch, d.h., möchte man für $\Delta\theta = 10°$ und $\theta_{\mathbf{p}} = 5°$ das Intervall $[-5°, 15°]$ bearbeiten, so entspricht dies $[175°, 180°[\cup [0°, 15°]$. Für Winkel $\theta \notin [0°, 180°[$ kehrt sich darüber hinaus auch noch das Vorzeichen des zugehörigen r um. Auf diese Möbiusstruktur wird in Kapitel 3.6.2 noch genauer eingegangen.

3.5 Die gewichtete, orientierungsselektive Hough-Transformation

Auch durch den Übergang von der Standard-Hough-Transformation zur orientierungsselektiven Hough-Transformation ist das Rauschen im Parameterraum nicht vollständig beseitigt. Es entsteht eine große Anzahl an nicht signifikanten lokalen Maxima. Durch eine Gewichtung der Hough-Transformation kann das Rauschen weiter reduziert werden, es wird also nicht jede betroffene Zelle um eins erhöht, sondern vielmehr um einen Wert zwischen null und eins in Abhängigkeit des Gradientenbetrags und des Gradientenwinkels [Ba81]. Dieses Verfahren vermindert das Rauschen im Parameterraum erneut deutlich, jedoch erfordert es ein geeignetes Maß für den Gradientenbetrag. Aus diesem Grund wird im folgenden Abschnitt ein Verfahren zur Normalisierung des Kontrasts vorgestellt, anschließend wird im Abschnitt 3.5.2 die dadurch notwendig gewordene effiziente Berechnung von Bildglättungen mit Binomialfiltern beschrieben. Zum Schluß im Abschnitt 3.6 wird die gewichtete, orientierungsselektive Hough-Transformation vorgestellt.

3.5.1 Kontrastnormalisierung

Um ein Maß dafür zu finden, wie gut eine Gerade eine Kante beschreibt, können der Gradientenwinkel und der Gradientenbetrag genutzt werden. Der Gradientenwinkel wird genutzt, um die Abweichung der Gradientenwinkel der Kantenpunkte zum Winkel der Geraden zu messen, der Gradientenbetrag macht eine Aussage über die Stärke der Kante. Diese Aussage ist allerdings nur schlecht als Maß zu verwenden, denn der Gradientenbetrag ist ein absoluter Wert, dessen Bedeutung für das Vorhandensein einer Kante jedoch abhängig von den Werten der umgebenden Punkte ist. Um ein Maß für das Vorhandensein einer Kante zu finden, ist es also sinnvoll, die Umgebung eines Punktes mit einzubeziehen, denn auch einem schwacher Grauwertanstieg ist mit hoher Wahrscheinlichkeit eine Kante zuzuordnen, wenn die Gradientenbeträge in der Umgebung deutlich niedriger sind. Der absolute Betrag spielt dabei nur eine untergeordnete Rolle. Interessant ist also weniger die Stärke einer Kante sondern vielmehr das Vorhandensein einer Kante [Kr98a].

Der normalisierte Gradient $E_{\mathbf{p}}$ eines Punktes $\mathbf{p}$ ergibt sich aus der lokalen und der globalen Energie des Gradientenbetragsbildes. Die globale Energie E^{glob} wird definiert als durchschnittlicher Gradientenbetrag des gesamten Bildes, also

$$E^{glob} := \frac{1}{XY} \sum_{\mathbf{p} \in B} G_{\mathbf{p}}. \tag{3.12}$$

Die lokale Energie $E_{\mathbf{p}}^{loc}$ eines Punktes $\mathbf{p}$ berechnet sich aus den Gradientenbeträgen von $\mathbf{p}$ und von Punkten in einer Umgebung. Wird das Gradientenbetragsbild mit einem Binomialfilter geglättet, so erhält man als Ergebnis für jeden Punkt die lokale Energie als gewichteten Mittelwert einer gewissen Umgebung, bei der der Gradientenbetrag von Punkten umso stärker gewichtet wird, je näher sie an $\mathbf{p}$ liegen. Die Maskengröße n_E des Binomialfilters kann z.B. mit $n_E \in \{5, 7, 9, 11\}$ gewählt werden.

Das Kriterium für einen Kantenpunkt ist nun nicht mehr ein globaler Schwellwert, sondern die Tatsache, daß die lokale Energie deutlich über der Durchschnittsenergie liegt. Die Durchschnittsenergie ist definiert als

$$E_{\mathbf{p}}^{av} := \varrho E_{\mathbf{p}}^{loc} + (1 - \varrho) E^{glob}, \tag{3.13}$$

ϱ ist der Gewichtungsfaktor, der bestimmt, wie groß der Anteil der lokalen Energie an der Durchschnittsenergie ist. Zwei markante Werte $\mu_{\mathbf{p}}, \nu_{\mathbf{p}}$ charakterisieren nun dieses Kriterium: Liegt die lokale Energie von $\mathbf{p}$ unter $\mu_{\mathbf{p}}$, so liegt kein Kantenpunkt vor, liegt die lokale Energie über $\nu_{\mathbf{p}}$, so liegt ein Kantenpunkt vor. Diese markanten Werte sind wie folgt definiert:

$$\mu_{\mathbf{p}} := \mu E_{\mathbf{p}}^{av} \tag{3.14}$$

$$\nu_{\mathbf{p}} := \nu E_{\mathbf{p}}^{av}, \tag{3.15}$$

wobei μ und ν global bestimmen, wie weit die lokale Energie eines jeden Punktes über dem Durchschnitt liegen muß, damit es sich um einen Kantenpunkt handelt bzw. wie weit unter dem Durchschnitt die lokale Energie höchstens liegen darf, damit es sich nicht um einen Kantenpunkt handelt.

Es mag zunächst ungewöhnlich erscheinen, daß die lokale Energie $E_{\mathbf{p}}^{loc}$ zur Durchschnittsenergie beiträgt, denn die Kontrastnormalisierungsfunktion (3.16) ist Funktion der lokalen Energie. Damit ist die Durchschnittsenergie allerdings für jeden Punkt eine andere, und dies erlaubt es, für einen kleinen Gradientenbetrag in der lokalen Umgebung eine geringere Forderung für das Vorhandensein einer Kante, für einen großen Gradientenbetrag in der lokalen Umgebung aber auch eine größere Forderung zu stellen. Der normalisierte Kontrast hängt somit von der lokalen sowie der globalen Energie ab.

Mit einem Standardparametersatz $\varrho = \frac{1}{4}$, die globale Energie wird also deutlich stärker gewichtet, $\mu = 1.0$ und $\nu = 2.3$ werden sehr gute Ergebnisse erzielt.

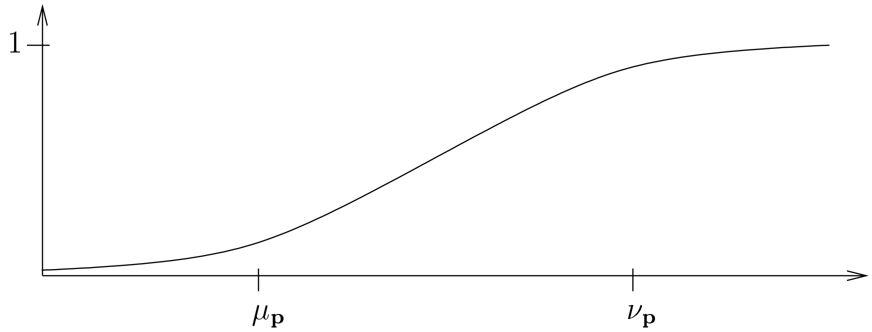


Abbildung 3.9: Die Kontrastnormalisierungsfunktion aus Gleichung (3.16) beginnt bei $\mu_{\mathbf{p}}$ deutlich zu steigen und hört bei $\nu_{\mathbf{p}}$ auf. Der Gradientenbetrag wird somit auf das Intervall $[0, 1]$ normiert.

Für den unsicheren Bereich zwischen $\mu_{\mathbf{p}}$ und $\nu_{\mathbf{p}}$ wird ein fließender und kein abrupter Übergang gewählt, so daß sich der normalisierte Gradient aus einer verschobenen und skalierten sigmoiden Funktion ergibt:

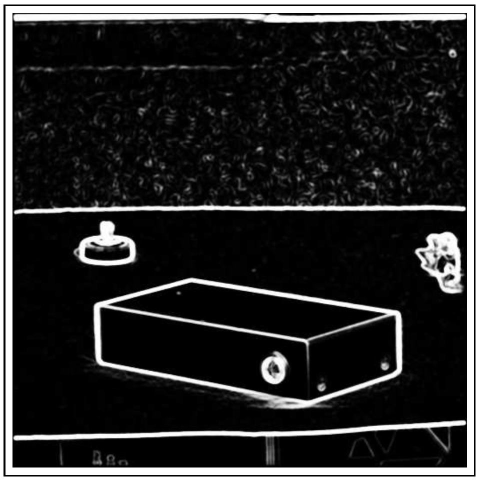
$$E_{\mathbf{p}} := \frac{1}{2} \tanh \left(\frac{3}{\nu_{\mathbf{p}} - \mu_{\mathbf{p}}} \left(E_{\mathbf{p}}^{loc} - \frac{\nu_{\mathbf{p}} + \mu_{\mathbf{p}}}{2} \right) \right) . \tag{3.16}$$

Abbildung 3.9 zeigt die Funktion, die so skaliert und verschoben ist, daß sie bei $\mu_{\mathbf{p}}$ anfängt und bei $\nu_{\mathbf{p}}$ aufhört zu steigen. Diese Funktion erfüllt ein Überdurchschnittskrierium, nach dem Gradientenbeträge über diesem Durchschnitt als Kante gewertet werden. Dabei ist es unerheblich, wie weit der Gradientenbetrag über dem Durchschnitt liegt. Eine Kante wird somit als Kante interpretiert, sobald sie einen markanten Wert überschritten hat, es wird kaum zwischen starken und schwachen Kanten unterschieden. Ebenso verhält es sich mit der Aussage, ob ein Punkt kein Kantenpunkt ist.

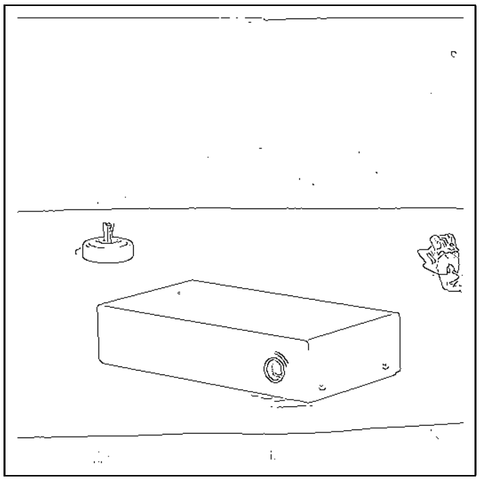
Die Normalisierung des Gradienten auf das Intervall $[0, 1]$ ermöglicht nun eine sinnvolle Gewichtung der Hough-Transformation und auch später eine sinnvolle Konfidenzbewertung von Liniensegmenten (siehe Abschnitt 4.1.1).

Die Abbildung 3.10 zeigt das normalisierte Gradientenbetragsbild aus Abbildung 3.5 und außerdem den Grund, warum dieses Verfahren keine Verbesserung des in Abschnitt 3.3 beschriebenen Verfahrens ist: Die Kontrastnormalisierung verstärkt lokal starke aber global schwache Grauwertübergänge, wie sie auch durch starkes Rauschen oder aber beispielsweise bei Textur entstehen. Man sieht deutlich störende Punkte, die schwache senkrechte Kante des Quaders wird allerdings auch nicht besser extrahiert (vgl. Abbildung 3.4).

Dies hat zur Folge, daß eine Menge isolierter Punkte als Kantenpunkte erkannt werden, die das Rauschen im Parameterraum der Hough-Transformation verstärken.



normalisierter Gradientenbetrag



Unterdrückung von Nichtmaxima des normalisierten Gradienten

Abbildung 3.10: Visualisierung der Kontrastnormalisierung

Das Kantenextraktionsverfahren aus Abschnitt 3.3 ignoriert diese schwachen Grauwertübergänge aufgrund der Wahl des globalen Schwellwerts.

3.5.2 Schnelle Faltung mit Binomialfiltern

Die Kontrastnormalisierung erfordert zur Berechnung der lokalen Energie eine Glättung des Gradientenbetragsbildes, und auch das Ergebnis der Kantenextraktion wird durch vorherige Rauschminderung durch eine Bildglättung verbessert. Es stellt sich daher die Frage, welcher Glättungsfilter zu wählen ist. Wegen der Isotropie ist die Gaußsche Filtermaske für den kontinuierlichen Fall die beste Wahl, allerdings ist die diskrete Approximation – die Binomialmaske – nicht mehr exakt isotrop. Trotzdem ist auch im diskreten Fall der Binomialfilter die beste Wahl, was [Ja93] ausführlich beweist.

Die Faltung mit einem $n \times n$-Binomialfilter erfordert für jeden Punkt $2n^2 - 1$ Operationen. Selbst bei einer kleinen Maske ist der Rechenaufwand bereits sehr groß. Es gibt allerdings mehrere Möglichkeiten, diesen quadratischen Aufwand auf linearen Aufwand zu reduzieren [Ja93].

Binomialmasken sind separabel, deshalb läßt sich die Faltung mit einer 2D-Binomialmaske in zwei Faltungen mit je einer horizontalen und einer vertikalen 1D-Binomialmaske zerlegen. Für das Beispiel einer 5×5-Binomialmaske ergibt sich somit

$$\frac{1}{256}\begin{pmatrix} 1 & 4 & 6 & 4 & 1 \\ 4 & 16 & 24 & 16 & 4 \\ 6 & 24 & 36 & 24 & 6 \\ 4 & 16 & 24 & 16 & 4 \\ 1 & 4 & 6 & 4 & 1 \end{pmatrix} = \frac{1}{16}\begin{pmatrix} 1 & 4 & 6 & 4 & 1 \end{pmatrix} * \frac{1}{16}\begin{pmatrix} 1 \\ 4 \\ 6 \\ 4 \\ 1 \end{pmatrix}. \tag{3.17}$$

Für eine Faltung mit einem $n \times n$-Binomialfilter sind auf diese Weise nur noch $4n - 2$ Operationen notwendig, der Aufwand wächst also nur noch linear mit der Maskengröße.

Ein theoretisch geringerer wenn auch immer noch linearer Aufwand läßt sich mit einer weiteren Zerlegung der 1D-Binomialmaske der Länge n erreichen. Eine $(n-1)$-fache Faltung mit der Elementarmaske $(1 \;\; 1)$ und anschließende Multiplikation mit $2^{-(n-1)}$ erzeugt die 1D-Binomialmaske der Länge n, der Rechenaufwand für die obige Faltung beträgt $2n + 1$ Operationen. Eingespart werden hier vor allem Multiplikationen, die allerdings bei der heutigen Rechnerarchitektur oft sogar schneller sind als Additionen. Auch die ausgefeilten Optimierungstechniken heutiger Compiler machen den theoretischen Vorteil des geringeren Rechenaufwands zunichte, so daß das oben beschriebene Verfahren mit separierten Filtern deutlich schneller ist als die Zerlegung in Elementarmasken.

Eine letzte Möglichkeit zur Effizienzsteigerung bei der Bildglättung ist das Zusammenfassen der Koeffizienten. Fast alle der n^2 Koeffizienten einer $n \times n$-Binomialmaske tauchen mehrfach auf. Durch Ausklammern läßt sich der Aufwand gegenüber der Separierung zwar theoretisch nicht verringern, er bleibt quadratisch, jedoch ist dieses Verfahren bei Masken bis zur Größe 11 das schnellste der getesteten. Dies läßt sich wiederum auf die Optimierungstechniken des Compilers zurückführen. Nachteil dieses Verfahrens ist, daß für jede Maske eine eigene Routine geschrieben werden muß, während für alle separablen Masken die gleiche Routine genutzt werden kann.

Bis zur Maskengröße 11 wurden eigene Routinen implementiert, da Masken dieser Größe in den allermeisten Fällen genutzt werden, der allgemeine Fall wurde mit Hilfe der Separabilität der Masken implementiert.

3.5.3 Gewichtung der Hough-Transformation

Sowohl die Standard- als auch die orientierungsselektive Hough-Transformation erhöhen eine betroffene Zelle im Akkumulatorarray genau dann um eins, wenn eine Bedingung erfüllt ist. Man kann vermuten, daß eine weitere Glättung des Parameterraums erreicht wird, wenn es keinen scharfen Übergang zwischen *Bedingung erfüllt* und *Bedingung nicht erfüllt* gibt, sondern vielmehr einen fließenden Übergang, indem jedem Parametersatz eine Wahrscheinlichkeit zugewiesen wird, mit der ein Parametersatz den entsprechenden Punkt auch wirklich erzeugt hat.

Solch eine Wahrscheinlichkeit $p(\mathbf{p}, \theta)$ eines Punktes $\mathbf{p}$ bezüglich der Hough-Geraden durch $\mathbf{p}$ mit dem Winkel θ kann mit Hilfe des normalisierten Gradienten aus Abschnitt 3.5.1 und des Gradientenwinkels wie folgt berechnet werden: Sei $E_\mathbf{p}$ der auf $[0, 1]$ normalisierte Gradient von $\mathbf{p}$, $\theta_\mathbf{p}$ dessen Gradientenwinkel und θ der Winkel der Hough-Geraden, dann ist $(1 - \frac{||\theta - \theta_\mathbf{p}||}{\Delta\theta})$ die auf $[0, 1]$ normierte Abweichung des Winkels der Hough-Geraden vom Gradientenwinkel des Punktes mit der maximalen

Abweichung $\Delta\theta$. Definiere nun

$$p(\mathbf{p},\theta) := E_{\mathbf{p}} \cdot \left(1 - \frac{||\theta - \theta_{\mathbf{p}}||}{\Delta\theta}\right) \tag{3.18}$$

als Produkt des normalisierten Gradienten und der Winkelabweichung. Dieser Wert macht eine Aussage darüber, mit welcher Wahrscheinlichkeit $\mathbf{p}$ Kantenpunkt auf der entsprechenden Hough-Geraden mit dem Winkel θ ist. Ein solches Maß findet sich auch in den folgenden Abschnitten wieder, denn zur Bestimmung der Signifikanz einer Hough-Geraden ist eine derartige Wahrscheinlichkeit von großem Nutzen.

Da hier und im folgenden oft von der Differenz zweier Winkel θ_1, θ_2 Gebrauch gemacht wird, wird $||\theta_1 - \theta_2||$ als der Differenzwinkel definiert, also

$$||\theta_1 - \theta_2|| := \min\{|\theta_1 - \theta_2|, 180^\circ - |\theta_1 - \theta_2|\} \tag{3.19}$$

im Gegensatz zum Betrag der Differenz zweier reller Werte $|\theta_1 - \theta_2|$.

Die orientierungsselektive Hough-Transformation aus Gleichung 3.10 wird nun wie folgt modifiziert:

$$A[s_r r, s_\theta \theta] := \sum_{\mathbf{p}\in B} f(\mathbf{p}, r, \theta) \tag{3.20}$$

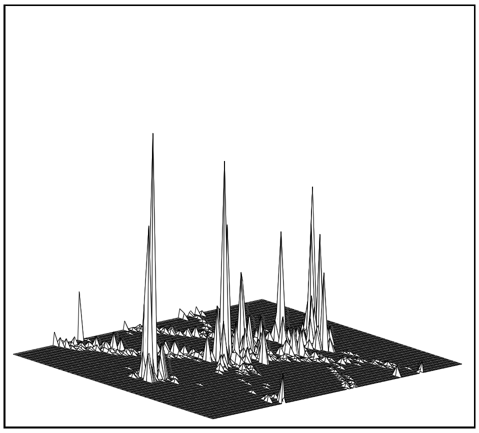
mit

$$f(\mathbf{p}, r, \theta) := \begin{cases} p(\mathbf{p},\theta) & \text{falls } \mathbf{f_p}(r,\theta) = 0 \wedge \theta_{\mathbf{p}} \in I \\ 0 & \text{sonst} \end{cases} \tag{3.21}$$

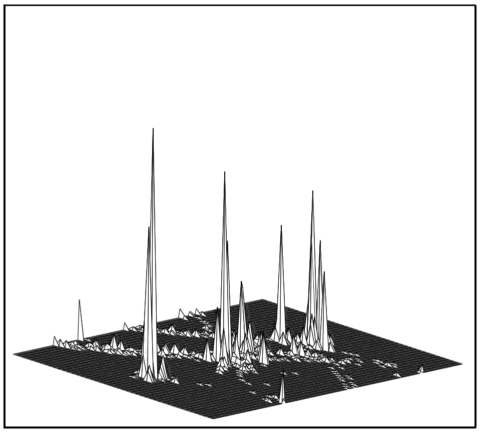
Es werden also weiterhin für jeden Kantenpunkt, den das Verfahren aus Abschnitt 3.3 ermittelt hat, lediglich die Gradientenwinkel aus I in Betracht gezogen, jedoch wird die Zelle (r, θ) des Akkumulatorarrays nicht mehr um eins erhöht, sondern um einen Wert zwischen null und eins, der der oben beschriebenen Wahrscheinlichkeit entspricht, daß der Punkt $\mathbf{p}$ Kantenpunkt auf der durch den Parametersatz (r, θ) definierten Geraden ist.

Tatsächlich glättet diese Erweiterung der orientierungsselektiven Hough-Transformation den Parameterraum noch einmal merklich (siehe Abbildung 3.11), sichtbar ist dies lediglich im Umkreis lokaler Maxima. Zwar haben signifikante lokale Maxima weiterhin lokale Nebenmaxima, jedoch sinkt deren Anzahl beträchtlich. Dazu trägt vor allem auch der in Abschnitt 3.3 beschriebene modifizierte Sobel-Operator bei, der die Gewichtung der Hough-Transformation mit der Winkelabweichung durch seine Genauigkeit erst sinnvoll macht. Bei Verwendung des Sobel-Operators zur Berechnung des Gradientenwinkels ist nur eine leichte Glättung des Parameterraums beobachtbar. Dies ist einleuchtend, da die Gewichtung des Parameterraums stark vom Gradientenwinkel abhängt und bei stark ungenauem Gradientenwinkel auch falsche Punkte stärker gewichtet werden.

Darüber hinaus ist das Problem der ungewichteten Hough-Transformation beseitigt, daß die Akkumulatorzellen zweier lokaler Maxima mit einer zu hohen Wahrscheinlichkeit denselben Wert haben. Einige Verfahren zur Maximumsuche stellen fest, ob



orientierungsselektive
Hough-Transformation mit $\Delta\theta = 5°$



gewichtete, orientierungsselektive
Hough-Transformation mit $\Delta\theta = 5°$

Abbildung 3.11: 3D-Darstellung des Akkumulatorarrays der gewichteten OHT im Vergleich mit der ungewichteten OHT

eines von zwei Maxima eher einer Geraden im Bild entspricht als das andere und gehen davon aus, daß es sich dabei um das stärkere handelt. Problematisch ist dies natürlich, wenn beide Zellen denselben Wert haben. Durch die Gewichtung werden nun Fließkommazahlen addiert, so daß die Wahrscheinlichkeit dieses Problemfalls vernachlässigbar gering ist (vgl. Abschnitte 3.6.3 und 3.6.4).

3.6 Finden der Maxima im Akkumulatorarray

Die Suche nach den lokalen Maxima im Akkumulatorarray stellt ein ernsthaftes Problem dar. In diesem Abschnit soll die Maximumsuche eingehender diskutiert werden. Abschnitt 3.6.1 erläutert, warum die Maximumsuche so problematisch ist. In Abschnitt 3.6.2 wird genauer auf die Zyklizität des Akkumulatorarrays eingegangen, die bei der Maximumsuche beachtet werden muß. Abschnitt 3.6.3 stellt Ansätze aus der Literatur vor, und in Abschnitt 3.6.4 wird ein selbst entwickeltes Distanzverfahren vorgestellt, das in 3.6.5 mit bisherigen Verfahren verglichen wird. [IK88, Le93] bieten eine gute Übersicht über die vorhandene Literatur zum Thema Hough-Transformation.

3.6.1 Problem der Maximumsuche

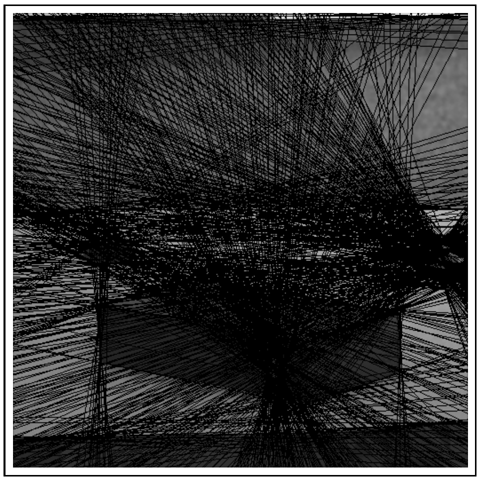
Das Auffinden von lokalen Maxima in einem zweidimensionalen Array ist einfach durch einen Vergleich jedes Punktes mit seinen acht Nachbarn zu lösen. Da jeder Kantenpunkt nicht nur für den Paramersatz derjenigen Geraden votiert, die mit der Kante korrespondiert, sondern auch noch für einige andere Parametersätze, und da darüber hinaus auch *falsche* Kantenpunkte und Kantenpunkte *uninteressanter* Kanten für diese Parametersätze votieren, enthält das Akkumulatorarray keine scharfen, eindeutig erkennbaren lokalen Maxima, sondern vielmehr eine Hügellandschaft (Abbildung 3.7). Jedes lokale Maximum besitzt viele Nebenmaxima in der Umgebung. Ferner hängt der Betrag des Maximums von der Länge des Kantenstücks ab, denn mindestens sämtliche Kantenpunkte haben für diesen Parametersatz votiert, dazu noch einige zufällige Votierungen. Lange Kanten erzeugen also lokale Maxima mit einem höheren Betrag als kurze Kanten, so daß die gewünschte Mindestlänge einer Kante die Festlgung eines Schwellwerts erleichtert, jedoch liegen die Nebenmaxima der langen Kanten meist auch über diesem Schwellwert.

Aus diesen Gründen reicht dieser einfache Ansatz nicht aus, um die signifikanten Maxima zu finden, denn ein signifikantes Maximum, welches mit einer tatsächlich im Bild vorhandenen Kante korrespondiert, kann auf diese Weise nicht von einem nicht signifikanten Maximum, das lediglich durch Rauschen erzeugt wurde, unterschieden werden. Abbildung 3.7 zeigt deutlich, daß das Akkumulatorarray tatsächlich sehr stark verrauscht ist. Trotz eines geeigneten Schwellwerts erhält man auf diese Weise selbst in sehr einfachen Bildern eine dreistellige Anzahl an lokalen Maxima, die im Akkumulatorarray grüppchenweise sehr eng zusammen liegen. Zeichnet man diese Geraden in das Ausgangsbild ein, so erkennt man Geradenbüschel an jeder Kante, die durch die hohe Anzahl an Nebenmaxima um das lokale Maximum entstehen. Beobachtungen ergeben hierbei einen maximalen Schnittwinkel der beiden äußersten zu einem Geradenbüschel gehörenden Geraden von 15°. Außerdem läßt sich beobachten, daß sich alle Geraden eines Büschels in einer kleinen Umgebung um die Kante schneiden (Abbildung 3.12).

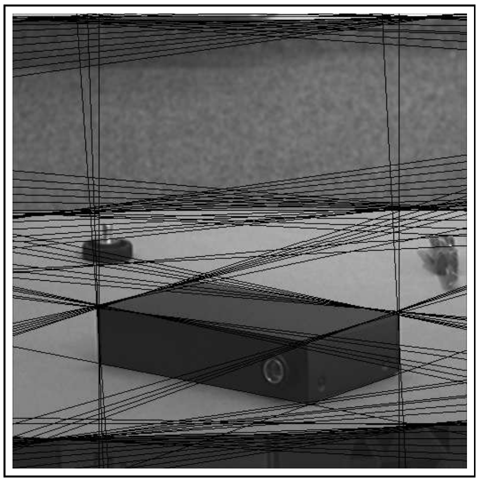
Eine Verminderung des Rauschens im Akkumulatorarray ist also in jedem Fall nützlich. Beide in Abschnitt 3.4 und 3.5 beschriebenen Modifikationen der Standard-Hough-Transformation vermindern das Rauschen, jedoch erhält man immer noch – wenn auch wesentlich kleinere – Grüppchen von lokalen Maxima (Abbildung 3.12).

Der geradlinige Ansatz, von jedem solchen Grüppchen im Akkumulatorarray den jeweils stärksten Kandidaten zu wählen, löst das Problem nur bedingt, denn der Betrag eines lokalen Maximums sagt allein nichts über dessen Signifikanz aus, auch wenn die Wahrscheinlichkeit dafür mit dem Betrag steigt. Zwar haben signifikante Maxima einen höheren Betrag als ihre durch Ungenauigkeit im Gradientenwinkel und durch Rauschen (abhängig von der Größe des gewählten $\Delta\theta$) erzeugten lokalen Nebenmaxima, jedoch erzeugen zwei eng beieinander liegende parallele Geradensegmente im Bild zwei eng beieinander liegende lokale Maxima, die durch diesen Ansatz unweigerlich verloren gehen.

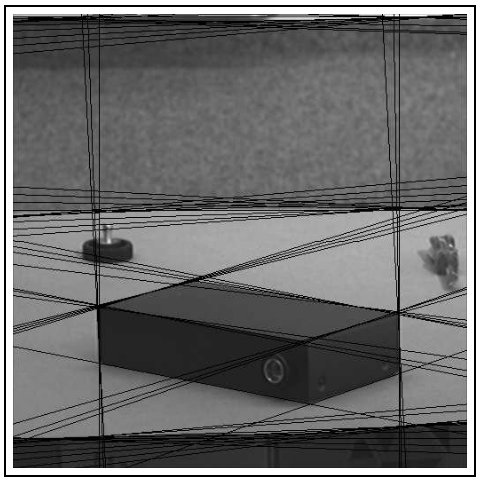
Die Abbildungen 3.13 (a)–(d) zeigen diverse solcher Problemfälle, auf die in Abschnitt 3.6.5 zur Verdeutlichung zurückgegriffen wird. In (a) werden zwei durch ein Schranktürenpaar erzeugte, eng beieinander liegende Parallelen gezeigt, auch an der Unterseite eines Schrankes entstehen nochmals zwei markante Parallelen. An Zimmertüren (b) ergeben sich durch den Türrahmen ebenfalls markante Strukturen durch enge Parallelen, daher sind diese Fälle für die Navigation von Robotern beson-



Standard-Hough-Transformation,
Schwellwert 40



orientierungsslektive
Hough-Transformation



gewichtete, orientierungsselektive
Hough-Transformation

Abbildung 3.12: Problem der Geradenbüschel bei der Maximumsuche

ders interessant. Bei der Objekterkennung spielen (c) und (d) eine große Rolle. Der Quader in (c) wurde in ungünstigem Winkel aufgenommen, so daß sich die Vorder- sowie die Seitenkante in einem so engen Winkel (kleiner als der bereits erwähnte maximale Winkel eines Geradenbüschels von 15°) schneiden, so daß die Seitenkante auch zu dem von der Vorderkante erzeugten Geradenbüschel gehören könnte. Auch in (d) schneiden sich zwei Geraden durch ungünstige Projektion mit einem kleinen Winkel, und das Ziel ist auch hier, die Signifikanz beider Geraden festzustellen und sie nicht als Geradenbüschel zu eliminieren.

Da die Maximumsuche erst das Ergebnis der Hough-Transformation liefert und die Qualität der Maximumsuche einen starken Einfluß auf die Qualität des Ergebnisses hat, wurde in dieser Arbeit besonders viel Wert auf diesen Teil des Verfahrens ge-

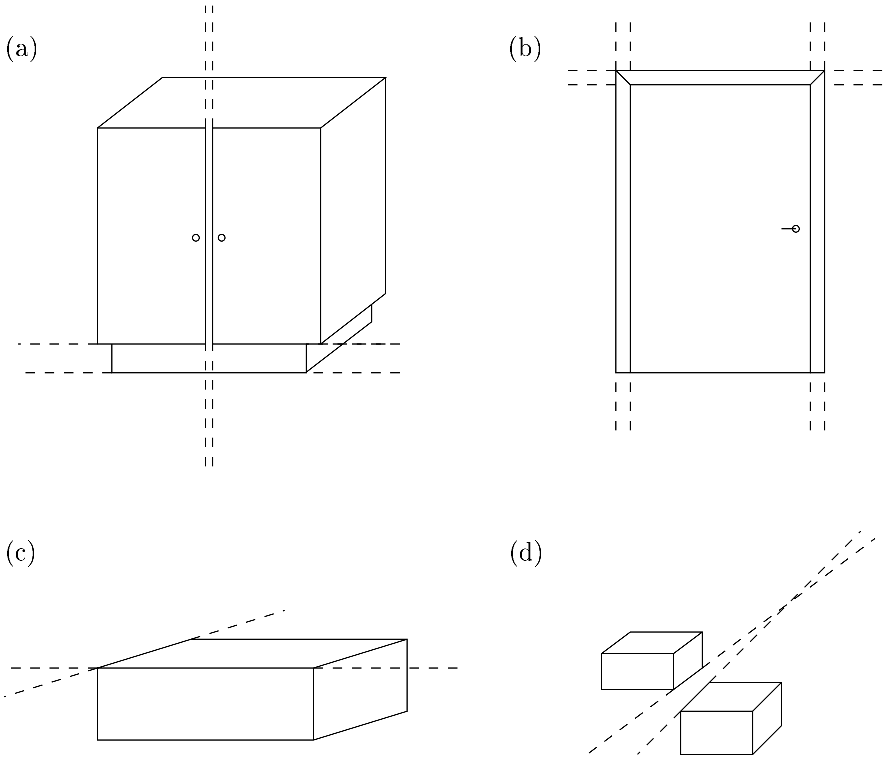


Abbildung 3.13: (a) und (b) zeigen zwei in der Robotik auftretende Problemfälle, die bei der Maximumsuche beachtet werden müssen. (c) und (d) zeigen zwei bei der Objekterkennung auftretende Problemfälle.

legt. Die Strategie ist zweigeteilt: Zuerst Verminderung des Rauschens so weit wie möglich, ohne die Lage der Maxima zu verändern. Dann finde aus der Menge der lokalen Maxima diejenigen heraus, die korrespondierende Kanten im Bild haben.

3.6.2 Zyklizität des Parameterraums

Auf ein Detail wurde bereits hingewiesen, welches auch alle Verfahren zur Maximumsuche beachten müssen: die Zyklizität des Parameterraumes. Nicht jede Zelle $z = [r, \theta]$ des Akkumulatorarrays besitzt zwei Nachbarn in r-Richtung, nämlich

$$z' = \left[r - \frac{1}{s_r}, \theta\right] \quad \text{und} \quad z'' = \left[r + \frac{1}{s_r}, \theta\right]. \tag{3.22}$$

Faktisch ist der Wertebereich von r zwar $]-\infty, +\infty[$, allerdings kann eine Gerade, die im Bild sichtbar ist, keinen größeren Abstand als $\frac{e}{2}$ zu dem in der Bildmitte liegenden Ursprung haben. Werte, die über $\pm\frac{e}{2}$ hinausgehen, brauchen daher nicht betrachtet zu werden, e sei hier die Länge der Bilddiagonalen.

In θ-Richtung hat jede Zelle $z = [r, \theta]$ des Akkumulatorarrays zwei Nachbarn, nämlich

$$z' = \left[r, \theta - \frac{1}{s_\theta}\right] \quad \text{und} \quad z'' = \left[r, \theta + \frac{1}{s_\theta}\right]. \tag{3.23}$$

Wegen der π-Periodizität gilt für die Nachbarn

$$z' = \left[-r, 180 - \frac{1}{s_\theta}\right] \quad \text{falls} \quad \theta - \frac{1}{s_\theta} < 0 \tag{3.24}$$

$$z' = \left[-r, \frac{1}{s_\theta}\right] \quad \text{falls} \quad \theta + \frac{1}{s_\theta} \geq 180. \tag{3.25}$$

Die Zyklizität ist einem Möbiusband ähnlich, es muß also auch noch das Vorzeichen von r umgekehrt werden.

Da $0° \leq \theta < 180°$ und nicht $0° \leq \theta < 360°$, legt das Vorzeichen von r fest, auf welcher *Seite* des Ursprungs die Gerade liegt. Formal begründet sich dies damit, daß bei der Hough-Transformation nur eine halbe Periode jeder Sinusodialkurve in das Akkumulatorarray eingetragen wurde und sich damit ein Vorzeichenwechsel ergibt. Diese Zyklizität muß also bei der Maximumsuche von den im folgenden Abschnitt genannten Verfahren beachtet werden.

3.6.3 Die Maximumsuche in der Literatur

Viele Veröffentlichungen zum Thema Hough-Transformation gehen ausführlich auf die Genauigkeit und die Effizienz der Hough-Transformation ein, es finden sich jedoch nur wenige Angaben zur Suche nach den lokalen Maxima im Akkumulatorarray. Dies mag sich dadurch begründen, daß das Problem der Maximumsuche intuitiv einfacher erscheint als es tatsächlich ist, und daher oftmals vernachlässigt wird.

An dieser Stelle werden nun kurz einige Ansätze aus der Literatur sowie ein eigener Ansatz informell vorgestellt und diskutiert. Die hier vorgestellten Verfahren haben alle die Eigenschaft, die Maximumsuche ausschließlich auf das Akkumulatorarray zu beschränken. Das Ausgangsbild wird zur Lösung dieses Problems nicht herangezogen. In Abschnitt 3.6.5 werden die angesprochenen Nachteile der Verfahren überprüft und mit dem in Abschnitt 3.6.4 vorgestellten Distanzverfahren, das die Signifikanz der im Akkumulatorarray gefundenen Maxima im Bild überprüft, verglichen.

Die hier angesprochenen Nachteile der verschiedenen Verfahren beziehen sich vor allem auf das Ziel dieser Arbeit, nämlich die Akkumulation einer stabilen Objektrepräsentation von polyedrischen Objekten über eine Bewegung auf der Basis von Geraden, die die Kanten beschreiben. Dies setzt voraus, daß möglichst für jede Kante eine Hough-Gerade gefunden wird. Wird für eine Kante während einer Bewegung in keiner Ansicht eine Hough-Gerade gefunden (Parallele oder eng beieinanderliegende Objektkanten, siehe Abbildung 3.13), so fehlt diese Kante in der Repräsentation völlig, während eine in einer Ansicht fehlerhafte Hough-Gerade über die Bewegung an Gewicht verliert.

Pauli [Pa98] schlägt vor, im Akkumulatorarray das globale Maximum zu suchen, einen Bereich um dieses Maximum zu löschen und anschließend neu zu beginnen. Diese Heuristik nutzt die Tatsache, daß die Parametersätze der Geraden eines Büschels ähnlich sind und deshalb im Akkumulatorarray nahe beieinander liegen. Die Anzahl der zu suchenden Maxima muß a priori festgelegt werden, was ein Problem darstellt. Darüber hinaus muß noch die Größe und Form des zu löschenden Bereichs festgelegt werden. Pauli schlägt hierbei ein rechteckiges Fenster mit dem gefundenen Maximum als Mittelpunkt vor. Es werden also alle Zellen gelöscht, deren zugehörige Parameter sowohl in r als auch in θ nur eine gewisse a priori festgelegte Differenz haben, die sich wie in Gleichung (3.27) berechnet.

Die Zahl der zu suchenden Maxima kann bei diesem Verfahren auch durch einen Schwellwert ersetzt werden; es wird so lange das globale Maximum gesucht, bis es den Schwellwert unterschreitet. Problemfälle wie in Abbildung 3.13 werden auf diese Weise allerdings nicht berücksichtigt, denn wird der zu löschende Bereich zu klein gewählt, so werden zwar einige Problemfälle erkannt, man erhält aber auch viele nicht signifikante Geraden durch Nebenmaxima. Wird der zu löschende Bereich größer gewählt, um die Nebenmaxima zu eliminieren, so werden auch die Problemfälle eliminiert. Dieses Phänomen läßt sich sehr einfach begründen: Jedes Geradensegment im Bild erzeugt neben einem lokalen Maximum im Akkumulatorarray auch weitere lokale Nebenmaxima. Existiert nun ein Geradensegment im Bild, dessen lokales Maximum mit einem Nebenmaximum eines anderen Segments zusammefällt, so läßt sich das lokale Maximum nicht mehr von dem Nebenmaximum eines anderen lokalen Maximums unterscheiden. Der Verlust der Lokalität bei der Hough-Transformation ist der Grund für diesen Nachteil.

Da beobachtbar ist, daß die von einem lokalen Maximum erzeugten Nebenmaxima im Akkumulatorarray nahe beieinander liegen, liegt es nahe, den Ansatz von Pauli um eine Metrik zu erweitern. Diese Metrik weist zwei Geraden anhand der zugehörigen Parametersätze einen Abstand zu, der klein ist, wenn die Geraden im Bild zu einem Büschel gehören könnten.

Seien also $g_1 = (r_1, \theta_1)$ und $g_2 = (r_2, \theta_2)$ zwei durch lokale Maxima im Akkumulatorarray repräsentierte Geraden. Definiere nun

$$d(g_1, g_2) := \alpha d_r(r_1, r_2, \theta_1, \theta_2) + (1 - \alpha) d_\theta(\theta_1, \theta_2) \tag{3.26}$$

mit

$$\begin{aligned} d_\theta(\theta_1, \theta_2) &:= ||\theta_1 - \theta_2|| = \min\{|\theta_1 - \theta_2|, 180 - |\theta_1 - \theta_2|\} \\ d_r(r_1, r_2, \theta_1, \theta_2) &:= \begin{cases} |r_1 - r_2| & \text{falls } d_\theta(\theta_1, \theta_2) = |\theta_1 - \theta_2| \\ |r_1 + r_2| & \text{falls } d_\theta(\theta_1, \theta_2) \neq |\theta_1 - \theta_2| \end{cases} \end{aligned} \tag{3.27}$$

Der Abstand zweier Geraden ist also die mit $0 \leq \alpha \leq 1$ gewichtete Summe aus den Differenzen der einzelnen Parameter. An dieser Stelle muß die Zyklizität des Akkumulatorarray in θ-Richtung beachtet werden (Abschnitt 3.6.2), daher die etwas kompliziert anmutende Definition für d_θ und d_r.

Bei genauerer Betrachtung fällt auf, daß diese Metrik eine formale Beschreibung der von Pauli verlangten Form und Größe des zu löschenden Bereichs ist. In diesem Fall handelt es sich dabei um eine Ellipse, wobei der Gewichtungsfaktor α das Verhältnis der beiden Radien zueinander festlegt. Die Größe wird duch den maximal erlaubten Abstand $d(g_1, g_2)$ bemessen. Darüber hinaus gilt dasselbe wie für das bereits zu dem Verfahren von Pauli erwähnte.

Mehrfach erwähnt wurde, daß die Nebenmaxima eines lokalen Maximums im Akkumulatorarray nahe beineinander liegen. Diese Tatsache wird sowohl von Pauli als auch von der oben beschriebenen Metrik ausgenutzt. Es wird vorausgesetzt, daß der Abstand zweier Geradenpaare, die sich mit demselben Winkel schneiden, an jedem Ort im Bild gleich ist. Dies ist jedoch nicht der Fall. Der Abstand d ist vielmehr abhängig vom Abstand r der Geraden zum Ursprung. Schneiden sich zwei Geraden im Ursprung, so ist $d_r = 0$. Schneiden sie sich in der äußersten Ecke des Bildes, so ist $d_r > 0$ abhängig von der Bildgröße. Tatsächlich ist nur d_θ in beiden Fällen gleich. Dies ist ein Nachteil einer solchen Metrik.

Auch Kunze [Ku99] und Schmidt [Sch98] versuchen, die Gruppierung der Maxima direkt im Akkumulatorarray zu beseitigen, jedoch auf eine andere Weise. Zuerst wird ein globaler Schwellwert auf das Akkumulatorarray angewendet, alle Werte darunter werden auf Null gesetzt. Dadurch wird das Rauschen unterdrückt, und es entstehen zusammenhängende Bereiche im Akkumulatorarray, deren jeweiliges Maximum gewählt wird. Der Zusammenhang wird durch die 8er-Nachbarschaft definiert. Die Problemfälle aus Abbildung 3.13 werden damit allerdings auch nicht konsequent berücksichtigt. Sehr schwierig ist es auch, den *richtigen* Schwellwert zu wählen, denn ist er zu hoch, so entstehen viele kleine Zusammenhangskomponenten und demnach viele, meist falsche lokale Maxima, ist er zu klein, so werden signifikante Geraden zusammengefaßt. Abbildung 3.14 zeigt vereinfacht die Wirkung des Schwellwerts.

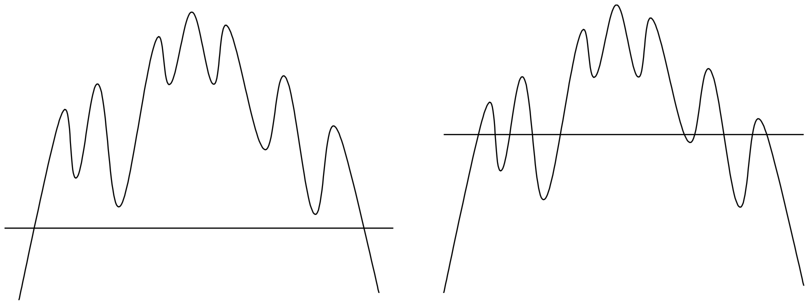


Abbildung 3.14: Diese vereinfachte 1D-Darstellung des Akkumulatorarrays zeigt die Problematik bei der Wahl des Schwellwerts. Links ist er vielleicht zu niedrig gewählt, denn es ist nicht klar, ob hier ein oder mehrere signifikante lokale Maxima enthalten sind. Rechts ist er zu hoch gewählt, es entstehen viele Zusammenhangskomponenten.

O'Gorman und Sanderson [OS84] haben ein Verfahren – den *converging squares* *algorithm* – entwickelt, das allgemein Maxima in mehrdimensionalen Arrays findet. Natürlich kann dieses Verfahren auch zur Maximumsuche im Akkumulatorarray eingesetzt werden, jedoch läßt sich nur ein einziges Maximum suchen, denn es wird nicht das globale Maximum gesucht, sondern vielmehr das lokale Maximum, das in dem Bereich größter Dichte liegt. Mit Hilfe einer Auflösungspyramide wird ein immer kleinerer Bereich des Arrays betrachtet, so daß zum Schluß der Bereich größter Dichte übrig bleibt. Prinzipiell läßt sich dieses Verfahren zur Maximumsuche nutzen, da es das Problem der Gruppierung der Nebenmaxima um ein lokales Maximum betrachtet, jedoch muß ein gefundenes Maximum wie bei dem von Pauli beschriebenen Verfahren gelöscht werden, um einen weiteren Suchvorgang starten zu können. Auf diese Problematik wurde bereits eingegangen.

Abbildung 3.15: Schmetterlingsstruktur von Maxima mit Akkumulatorarray

Da Rauschen das Hauptproblem bei der Maximumsuche ist, versuchen Leavers und Boyce [LB87], das Akkumulatorarray mit geeigneten Filtern zu glätten. Die Struktur eines Maximums, das von einem Liniensegment im Bild erzeugt wird, wird näher untersucht und als *Schmetterlingsstruktur* beschrieben. Jedes Maximum weist diese typische Struktur auf, bei der das Rauschen eine Verteilung aufweist, die Schmetterlingsflügeln ähnelt (siehe Abbildung 3.15). Aus dieser Erkenntnis wird eine Faltungsmaske abgeleitet, die genau diese Maxima verstärken und zufällige Maxima unterdrücken soll. Die Faltungsmaske ist einfacher Struktur, der linke und rechte Nachbar wirkt verstärkend, der obere und untere stark abschwächend. Die Maske sieht wie folgt aus:

$$\begin{pmatrix} 0 & -2 & 0 \\ 1 & 2 & 1 \\ 0 & -2 & 0 \end{pmatrix} \tag{3.28}$$

Ein Vorteil dieses Verfahrens ist zweifelsohne die Parameterlosigkeit, auch vermindet es das Rauschen im Akkumulatorarray. Jedoch wird auch hier das Problem der Signifikanz zweier eng beieinander liegender lokaler Maxima nicht gelöst, da es ausschließlich auf dem Akkumulatorarray arbeitet. Darüber hinaus sind die Geraden etwas ungenauer, wie in Abschnitt 3.6.5 gezeigt wird, da ein lokales Maximum im Akkumulatorarray durch die Filterung verschoben werden kann.

Einen alternativen Ansatz bieten Vinod et al. [Vi92]. Mittels eines neuronalen Netzes wird die markante Verteilung der Werte im Akkumulatorarray in einer lokalen Umgebung analysiert. Das Netz ist in der Lage, die signifikanten Peaks anhand dieser markanten Struktur zu erkennen und von nicht-signifikanten zu unterscheiden. Die in dieser Arbeit angeführten Ergebnisse sind vielversprechend, jedoch räumen

Vinod et al. ein, daß das Netz für andere Anwendungen ggfs. in geringem Rahmen neu trainiert werden muß.

Viele der vorgestellten Verfahren reagieren sehr sensitiv auf Veränderungen der Parameter, wünschenswert wäre jedoch ein Verfahren, das die Mehrzahl der Aufgaben mit einem Standardparametersatz bewältigt. Darüber hinaus versuchen alle Verfahren, das Problem ausschließlich im Akkumulatorarray zu lösen. Da jedoch die Lokalität, also die Lage der zu einer Geraden gehörenden Kante, bei der Hough-Transformation verloren geht, ist ein Vergleich der gefundenen Geraden mit dem Bild nützlich, um diejenigen Geraden zu finden, die tatsächlich mit der Objektberandung übereinstimmen. Das folgende Distanzverfahren erfüllt diese Bedingungen.

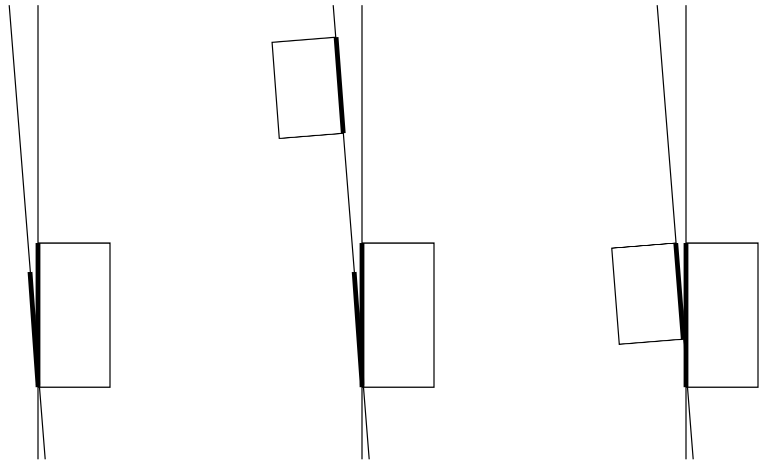
3.6.4 Distanzverfahren zur Maximumsuche

Um zu entscheiden, ob zwei eng beieinander liegende lokale Maxima tatsächlich zwei verschiedene im Bild vorhandene Geradensegmente beschreiben oder nicht, wird im folgenden eine Metrik eingeführt, die die durch die Parameter zweier Maxima beschriebenen Geraden mit dem Bild vergleicht und ihre signifikanten Bereiche bestimmt. Die Idee ist, daß der durch die Metrik bestimmte Abstand zweier Geraden genau dann klein ist, wenn die signifikanten Segmente beider Geraden lokal im Bild übereinstimmen, dieselbe Kante also beide Maxima erzeugt hat. Der Abstand ist groß, wenn eine Gerade ein signifikantes Segment besitzt, das für die andere Gerade nicht signifikant ist. Auf diese Weise wird bestimmt, ob die Geraden zu einem Geradenbüschel oder aber beide zu verschiedenen Kanten gehören. Die globale Information wird also durch lokale Information verifiziert.

Mit anderen Worten, laufen zwei Geraden im Bild entlang nur einer Kante, so handelt es sich um ein Geradenbüschel, läuft jedoch eine Gerade entlang einer Kante, entlang derer die zweite Gerade nicht verläuft, so sind die Geraden unabhängig. Abbildung 3.16 visualisiert diese Unterscheidung, die von der Metrik gemacht werden soll.

Empirisch läßt sich feststellen, daß ein signifikantes Maximum lokale Nebenmaxima in der Form hat, daß sich die zugehörigen Geraden im Bild an ihren starken Bereichen mit einem maximalen Winkel von $15°$ schneiden (vgl. Abschnitt 3.6.1). Als *starke Bereiche* von Geraden werden im folgenden diejenigen Segmente der Geraden bezeichnet, an denen die Orientierung des Gradienten mit der Orientierung der Geraden weitestgehend übereinstimmt und der Gradientenbetrag hoch ist. Es ist also der Bereich der Geraden, der mit einer Objektkante korrespondiert.

Die Frage ist nun, wie man feststellt, ob die signifikanten Bereiche beider Geraden einer oder zwei Kanten im Bild zugeordnet werden können. Da sich die Geraden jedoch an einem starken Bereich beider Geraden schneiden, läßt sich dies unter Zuhilfenahme des Abstandes zum Schnittpunkt feststellen. Betrachtet man auf jeder Geraden jeweils einen Punkt mit demselben Abstand zum Schnittpunkt, so unterscheidet man zwei Fälle: Liegen beide Punkte im starken Bereich der Geraden, so



(a) (b) (c)

Abbildung 3.16: Von der Metrik zu unterscheidende Fälle

spricht dies dafür, daß beide Geraden von derselben Kante erzeugt wurden. Liegt nur ein Punkt an einer Kante, so wurden die Geraden von verschiedenen Kanten erzeugt. Abbildung 3.16 zeigt in (a) und (b) diese zu unterscheidenden Fälle. Die starken Bereich der Geraden sind markiert. Ein Sonderfall wird in (c) gezeigt. Hier werden beide Geraden von verschiedenen Kanten erzeugt, sie erwecken jedoch den Eindruck, daß es sich um dieselbe Kante handelt.

Dieses Problem läßt sich wie folgt beseitigen: Der Einflußbereich einer Kante beträgt im Mittel 2.5 Pixel zu beiden Seiten der Kante. Dies bedeutet, daß auch Geraden, die in dieser Entfernung an der Kante vorbei verlaufen, dort einen starken Bereich aufweisen. Dies bedeutet aber auch, daß eine Gerade, die einen starken Bereich in größerer Entfernung als 2.5 Pixel von einer anderen Geraden beseitzt, Fall (b) zuzuordnen ist. Formal läßt sich dies über die Entfernung zum Schnittpunkt bestimmen, denn beträgt diese Entfernung mehr als $2.5 \cdot \cot \chi$, wobei χ der Schnittwinkel der Geraden ist, so beträgt der Abstand beider Punkte mehr als 2.5. Die Akkumulation dieser punktweisen Informationen läßt eine endgültige Aussage über die Signifikanz beider Geraden zu.

Es ist also ein Konfidenzmaß $c_g(\mathbf{p})$ zu definieren, das dem Punkt $\mathbf{p} = (x, y)$ einen hohen Wert zuweist, wenn $\mathbf{p}$ tatsächlich Kantenpunkt auf der Geraden $g = (r, \theta)$ ist, also zu dem starken Bereich der Geraden gehört. Ansonsten wird $\mathbf{p}$ ein niedriger Wert zugewiesen. Diese Konfidenz wird definiert als Produkt des normalisierten Gradienten und der Abweichung des Geradenwinkels vom Gradientenwinkel des Punktes:

$$c_g(\mathbf{p}) := E_{\mathbf{p}} \cdot f_\theta(||\theta - \theta_{\mathbf{p}}||) \tag{3.29}$$

mit

$$f_\theta(\theta) := \begin{cases} \frac{90^\circ - \theta}{160^\circ} & \text{falls } \theta > 10^\circ \\ \frac{10^\circ - \theta}{20^\circ} + 0.5 & \text{falls } \theta \leq 10^\circ \end{cases} \tag{3.30}$$

Abbildung 3.17 zeigt die Funktion f_θ, die geringen Winkelabweichungen eine hohe Konfidenz, größeren Abweichungen aber eine stark fallende Konfidenz zuweist.

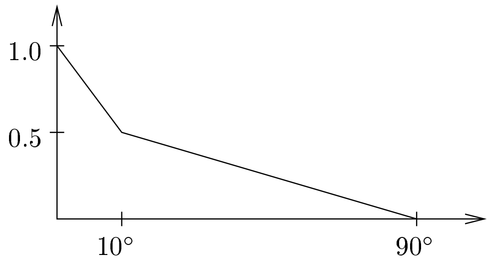


Abbildung 3.17: Winkelabweichungsfunktion

Seien nun $g = (r, \theta)$ und $g' = (r', \theta')$ zwei nicht parallele, durch lokale Maxima im Akkumulatorarray bestimmte Geraden. Sei $\mathbf{s} := (x_\mathbf{s}, y_\mathbf{s})$ der Schnittpunkt von g, g'. Bezeichne $g(t)$ den Punkt (x_t, y_t) mit Abstand t von $\mathbf{s}$. Es ist also

$$x_t = x_\mathbf{s} + t \sin\theta \tag{3.31}$$

$$y_t = y_\mathbf{s} - t \cos\theta. \tag{3.32}$$

Mit Hilfe der Konfidenz $c_g(\mathbf{p})$ läßt sich nun eine Aussage darüber machen, ob beide Geraden ihre starken Bereiche von derselben Kante im Bild haben oder von verschiedenen. In Fall (a) aus Abbildung 3.16 handelt es sich dabei um ein Geradenbüschel, die schwächere der beiden Geraden (gemessen am Wert im Akkumulatorarray) kann gelöscht werden. In Fall (b) und (c) repräsentieren die Geraden verschiedene Liniensegmente im Bild.

Werden nun beide Geraden simultan in äquidistanten Schritten vom Schnittpunkt ausgehend in beiden Richtungen abgetastet, so lassen sich die Konfidenzen von jeweils zwei Punkten gleicher Entfernung t zum Schnittpunkt betrachten, also $c_g(g(t))$ und $c_{g'}(g'(t))$. Sind beide Konfidenzen hoch, so handelt es sich dabei um dieselbe Kante im Bild. Ist jedoch $c_g(g(t))$ groß und $c_{g'}(g'(t))$ klein, so hat die Kante im Bild nur Einfluß auf eine der beiden Geraden im Bild. Es muß also eine Funktion $\mathbf{f}_t$ definiert werden, die ersterem einen niedrigen und letzterem einen hohen Wert zuweist. Durch Aufsummierung dieser Werte der Punkte ergibt sich der Abstand der Geraden.

Wir definieren zunächst eine Wahrheitsfunktion, die *true* liefert, falls die Bedingungen für die Metrik gegeben sind, also der Schnittwinkel $\chi := ||\theta - \theta'||$ der Geraden kleiner als 15° ist, der Schnittpunkt im Bild liegt und die Konfidenz des Schnittpunkts hoch ist.

$$\Theta(g, g') := \begin{cases} \textit{true} & \text{falls } \chi \leq 15^\circ \wedge \mathbf{s} \in B \wedge \\ & \qquad (c_g(\mathbf{s}) > \frac{1}{2} \wedge c_{g'}(\mathbf{s}) > \frac{1}{2}) \\ \textit{false} & \text{sonst} \end{cases} \tag{3.33}$$

Nun definieren wir die eben beschriebene Funktion $\mathbf{f}$, die beide Geraden abtastet und entsprechend der Funktion $\mathbf{f}_t$ hohe oder niedrige Werte aufsummiert.

$$\mathbf{f}(g,g') := \sum_t \mathbf{f}_t(g,g') \tag{3.34}$$

mit

$$\mathbf{f}_t(g,g') := \begin{cases} c_g(g(t)) & \text{falls } (c_g(g(t)) > \frac{1}{2} \wedge c_{g'}(g'(t)) \leq \frac{1}{2}) \text{ oder} \\ & \quad (c_g(g(t)) > \frac{1}{2} \wedge c_{g'}(g'(t)) > \frac{1}{2} \wedge t \geq 2.5 \cdot \cot\chi) \\ 0 & \text{sonst} \end{cases} \tag{3.35}$$

Die Funktion $\mathbf{f}_t$ erfüllt die oben geforderten Bedingungen. Betrachtet werden hier nur die Fälle, in denen die Funktion einen hohen Wert haben soll. Dies ist im einzelnen der bereits erwähnte Fall, daß nur ein Punkt Support von einer Kante bekommt (erste Zeile in Formel 3.35). Außerdem ist dies der Fall, daß die beiden Punkte zu verschiedenen Kanten gehören (zweite Zeile in Formel 3.35), beide Punkte als eine hohe Konfidenz haben, sie jedoch so weit auseinander liegen, daß sie nicht beide im Einflußbereich derselben Kante liegen. Ansonsten liefert die Funktion Null.

Da eine Kante auf die optimale Gerade mehr Einfluß hat als auf eine andere Gerade, ist im allgemeinen $\mathbf{f}(g,g') \neq \mathbf{f}(g',g)$. Definiere das Minimum als die Metrik, die $+\infty$ ist, falls die Bedingungen nicht zutreffen, $\Theta(g,g')$ also *false* liefert.

$$\mathbf{d}(g,g') := \begin{cases} \min\{\mathbf{f}(g,g'),\mathbf{f}(g',g)\} & \text{falls } \Theta(g,g') = \mathit{true} \\ +\infty & \text{sonst} \end{cases} \tag{3.36}$$

$\mathbf{d}(g,g') < 40$ ist eine gute Wahl, um die Zugehörigkeit beider Geraden zu einem Geradenbüschel zu ermitteln. Wähle in diesem Fall diejenige Gerade aus, die den höheren Wert im Akkumulatorarray besitzt. Der Schwellwert für das Akkumulatorarray spielt bei diesem Verfahren eine untergeordnete Rolle, er ist nur noch Maß für die gewünschte minimale Länge eines Liniensegments. Die Wahl hat jedoch keinen Einfluß auf das Auftreten von Geradenbüscheln (vgl. Abschnitt 3.6.5).

Bei der Implementation des Verfahrens ist zu beachten, daß die Geraden ausgehend vom Schnittpunkt in beiden Richtungen abgetastet werden, und zwar beide in der gleichen Richtung, nämlich entlang des kleineren der beiden Schnittwinkel. Dieses Richtungsproblem tritt allerdings nur dann auf, wenn $|\theta - \theta'| > 90°$ ist. In diesem Fall bestimmen sich (x_t, y_t) zu

$$x_t = x_{\mathbf{s}} + t\sin\theta \tag{3.37}$$
$$y_t = y_{\mathbf{s}} + t\cos\theta. \tag{3.38}$$

3.6.5 Vergleich der Verfahren

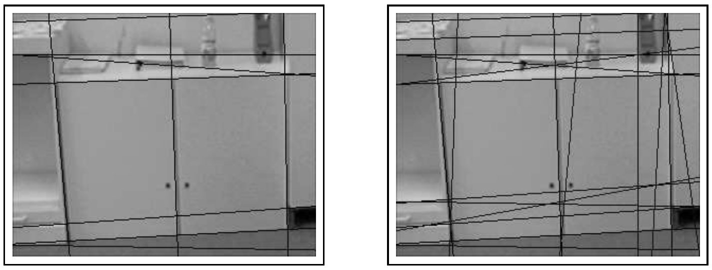
An einigen exemplarischen Beispielen werden nun die qualitativen Eigenschaften der verschiedenen vorgestellten Verfahren zur Suche von Maxima im Akkumulatorarray

Kantenbild Distanzverfahren Leavers und Boyce

Abbildung 3.18: Vergleich des Distanzverfahrens mit dem von Leavers und Boyce

verdeutlicht. Für weitere Vergleiche sowie die Bestimmung eines optimalen Parametersatzes für die Hough-Transformation unter Nutzung des Distanzverfahrens sei auf Abschnitt 3.7 verwiesen.

Das vom Distanzverfahren mit Schwellwert 20 und Distanz 40 gelieferte Ergebnis zum Vergleich mit den anderen Verfahren findet sich in Abbildung 3.18. Einige gefundene Geraden erscheinen auf den ersten Blick überflüssig, bei genauerem Hinsehen korrespondieren sie jedoch mit – wenn auch recht kurzen – Kantenstücken im Bild, vergleiche hierzu das Kantenbild. Diese Abbildung zeigt auch die leichte Ungenauigkeit, mit der das Verfahren von Leavers und Boyce arbeitet. Die deutliche Oberkante des Schranks fällt völlig weg, die verschiedenen kurzen Kanten unten links und rechts werden zu einer Geraden zusammengefaßt.



10 Maxima 20 Maxima

Abbildung 3.19: Verfahren von Pauli

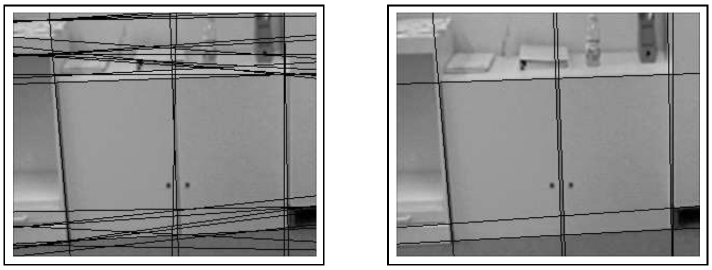
Abbildung 3.19 zeigt das Verfahren von Pauli mit verschiedener Zahl vorgegebener Maxima. Die Höhe und die Breite des zu löschenden Fensters ist mit 20 vorgegeben. Bei passender Anzahl der gefundenen Maxima liefert dieses Verfahren ein gutes Ergebnis, jedoch werden die Parallelen am Schrank eliminiert.

Abbildung 3.20 verdeutlicht die Sensitivität des Abstandsparameters gegenüber kleinen Änderungen. Bei vorgegebener Distanz 10 erhält man ein gutes Ergebnis bis auf die Elimination der Parallelen am Schrank. Kleine Änderungen dieser Distanz liefern jedoch stark veränderte Ergebnisse.

Das Verfahren von Kunze (Abbildung 3.21) hängt sehr stark vom Schwellwert ab. Bei

Distanz 5 Distanz 10 Distanz 15

Abbildung 3.20: Verfahren von Pauli, Löschen über Distanzfunktion



Schwellwert 20 Schwellwert 40

Abbildung 3.21: Verfahren von Kunze

guter Wahl des Schwellwerts liefert dieses Verfahren ein gutes Ergebnis, es werden sogar die Parallelen zwischen den Schranktüren erkannt, nicht jedoch am linken Rand des Schranks. Ein großer Nachteil ist die Sensitivität des Schwellwerts. Der Schwellwert ist nicht nur Maß für die Mindestlänge eines Kantenstücks, sondern ruft bei zu niedriger Wahl auch Geradenbüschel hervor.

Man sieht deutlich, daß die qualitativen Nachteile, wie sie in Abschnitt 3.6.3 angeführt wurden, tatsächlich nur an wenigen bestimmten – wenn auch gelegentlich wichtigen – Stellen auftreten. Darüber hinaus reagieren viele Verfahren sehr empfindlich auf kleine Veränderungen bei den zugehörigen Parametern. Das Distanzverfahren liefert hierbei das beste Ergebnis, auch läßt es sich sehr gut mit einem Standardparametersatz betreiben, der nur in seltenen Fällen an die Aufgabe angepaßt werden muß (vgl. Abschnitt 3.7).

3.7 Experimente

In diesem Abschnitt wird der vollständige Ablauf der Extraktion von Geraden aus einem Grauwertbild zusammengefaßt, der optimale Parametersatz für dieses Verfahren gefunden, und schließlich werden die Verfahren zur Maximumsuche mit deren jeweiligen Standardparametersätzen mittels zehn Bildern verglichen.

Um die Geraden in einem Grauwertbild zu bestimmen, müssen nacheinander folgende Schritte abgearbeitet werden.

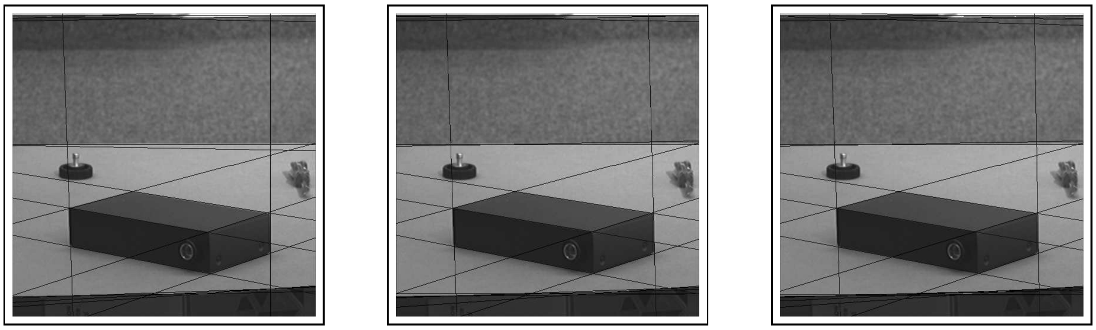
1. Bildglättung mit einem 3×3-Binomialfilter
2. Kantenextraktion mittels des modifizierten Sobel-Operators und Unterdrückung von Nichtmaxima
3. Kontrastnormalisierung
4. gewichtete, orientierungsselektive Hough-Transformation
5. Suche der Maxima im Akkumulatorarray mit dem Distanzverfahren

Diese Verfahrensweise liefert durchweg gute Ergebnisse, die Parameter für die zehn in den Abbildungen 3.25 und 3.26 gezeigten Bilder mußten nicht verändert werden, sie wurden auch bereits in den zugehörigen Abschnitten beschrieben und deren Wahl eingehend begründet. An dieser Stelle werden sie noch einmal kurz zusammengefaßt und einige wenige auf Veränderungen getestet.

Der Schwellwert η für die Binarisierung des Bildes ist mit 8 voreingestellt. Lediglich bei extrem kontrastarmen oder -reichen Bildern ist eine Veränderung von Nutzen.

Der Standardparametersatz der Kontrastnormalisierung braucht nicht verändert zu werden, er ist auf $(\mu, \nu, n_E, \varrho) = (1.0, 2.3, 5, 0.25)$ voreingestellt.

Der Winkelfehler $\Delta\theta$ der orientierungsselektiven Hough-Transformation liefert mit dem Wert 10 die robustesten Ergebnisse. Wegen der Ungenauigkeit bei der Berechnung des Gradientenwinkels sinkt die Robustheit bei zu niedriger Wahl des Fehlers. Ein zu hoher Wert erzeugt mehr Rauschen im Akkumulatorarray und erschwert somit die Maximumsuche.



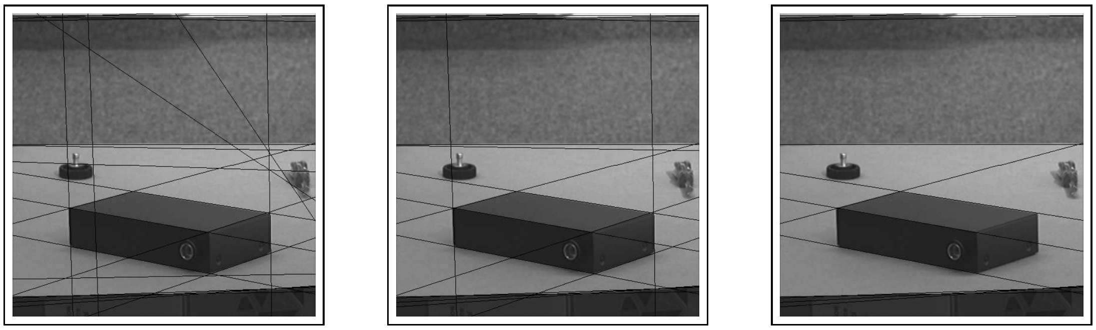
Skalierfaktor 1 Skalierfaktor 2 Skalierfaktor 3

Abbildung 3.22: Wahl des Skalierfaktors

Der Skalierfaktor des Akkumulatorarrays verspricht exaktere Ergebnisse bei einer höheren Abtastrate des Parameterraums und die damit genauere Möglichkeit der Bestimmung der Parameter der Geraden auch auf Nachkommastellen. Abbildung

3.22 zeigt nur minimale Veränderungen bei einer Erhöhung des Skalierfaktors. Da jedoch auch der Aufwand quadratisch mit dem Skalierfaktor steigt, ist 1 hier der beste Kompromiß.

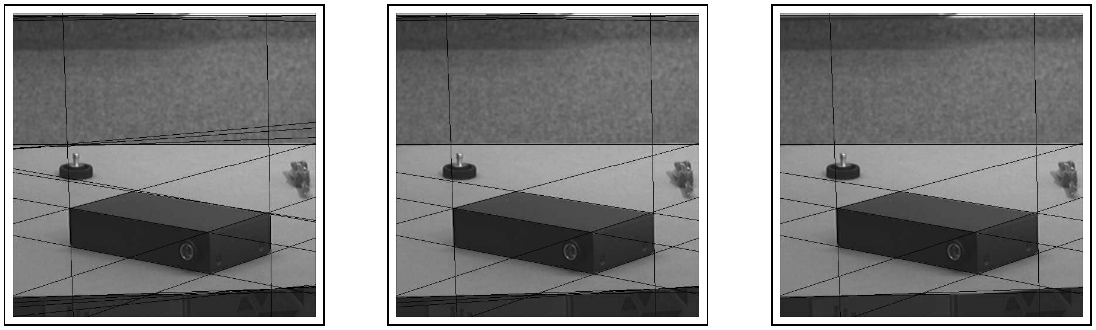
Eine wesentliche Rolle spielen hingegen die beiden Parameter des Distanzverfahrens zur Maximumsuche. Sowohl der Schwellwert σ als auch die vorgegebene minimale Distanz $\mathbf{d}$ sind ausschlaggebend dafür, daß alle signifikanten Geraden gefunden werden, aber auch nur wenige nicht signifikante. Dabei sollen diese Parameter strikt getrennt sein, der Schwellwert als Maß für die Mindestlänge eines Liniensegments im Bild, die Distanz als Entscheidungskriterium für die Signifikanz zweier Geraden.



Schwellwert 10 | Schwellwert 20 | Schwellwert 80

Abbildung 3.23: Wahl des Schwellwerts

Abbildung 3.23 zeigt, daß der Schwellwert tatsächlich diese geforderte Bedingung erfüllt. Wird er sehr hoch gewählt, so werden nur die längsten Kanten gefunden, wird er niedrig gewählt, so erscheinen auch kurze Kanten im Bild, jedoch keine Geradenbüschel. Es ist jedoch zu beachten, daß die Mindestlänge einer Kante von der Bildgröße abhängt. Verkleinert man beispielsweise ein Bild, so verringern sich auch die Längen der Kanten, Geradenbüschel sollen natürlich trotzdem nicht auftreten.



Distanz 10 | Distanz 40 | Distanz 100

Abbildung 3.24: Wahl der Distanz

Für die Elimination von Büscheln ist eine geeignete Wahl der Distanz notwendig, jedoch fällt diese Wahl nicht schwer. Der Standardwert 40 liefert durchweg gute Ergebnisse, bei zu niedriger Wahl erhält man Geradenbüschel (vgl. 3.24), eine Erhöhung des Werts hat jedoch kaum Konsequenzen.

Zum besseren Vergleich der Verfahren zur Maximumsuche finden sich in den Abbildungen 3.25 und 3.26 zehn Bilder mit den jeweiligen gefundenen Geraden. In der linken Spalte sind die Ergebnisse des Distanzverfahrens mit den Parametern $\mathbf{d} = 40, \sigma = 20$, in der Mitte des Verfahrens von Pauli mit $\sigma = 20$ und der Metrik $d(g_1, g_2) = 10$ und rechts des Verfahrens von Kunze mit $\sigma = 20$ abgebildet.

Das hier vorgestellte Verfahren zur Extraktion von Geraden aus Grauwertbildern arbeitet effizient mit sehr guten Ergebnissen. Eine Anpassung von Parametern an die Aufgabenstellung ist in der Regel nicht notwendig, so daß es bestens zur Gewinnung einer Objektrepräsentation gut geeignet ist.

3.8 Zusammenfassung

Am Anfang dieses Kapitels wurde die Hough-Transformation zur Extraktion von Bildprimitiva gewählt. Ausschlaggebend für diese Wahl war einerseits die Art der zu extrahierenden Bildprimitiva – nämlich Kanten polyedrischer Objekte – und andererseits die vielversprechenden Ergebnisse, schließlich hat sich die Hough-Transformation auf diesem Gebiet lange Zeit bewährt. Der Vorteil dieses globalen Verfahrens gegenüber lokalen Verfahren wurde diskutiert, die Nachteile der Globalität traten jedoch bei der Maximumsuche auf.

Das starke Rauschen im Akkumulatorarray sowie die Notwendigkeit der Wahl eines Schwellwerts stellten die Signifikanz von gefundenen Maxima in Frage. Aufgrund der fehlenden Lokalität im Akkumulatorarray wurde ein Vergleich mit dem Bild notwendig, um zu entscheiden, welche gefundenen Maxima denn tatsächlich Kanten im Bild repräsentieren. Da einige in der Literatur erwähnte Verfahren nachteilig sind, wurde ein Verfahren entwickelt, das die meisten problematischen Fälle bewältigt und nur wenig redundante Information erzeugt. Es wurde vor allem Wert auf die Bewältigung problematischer Fälle gelegt, da aus dem Ergebnis der Maximumsuche eine Objektrepräsentation gewonnen werden soll.

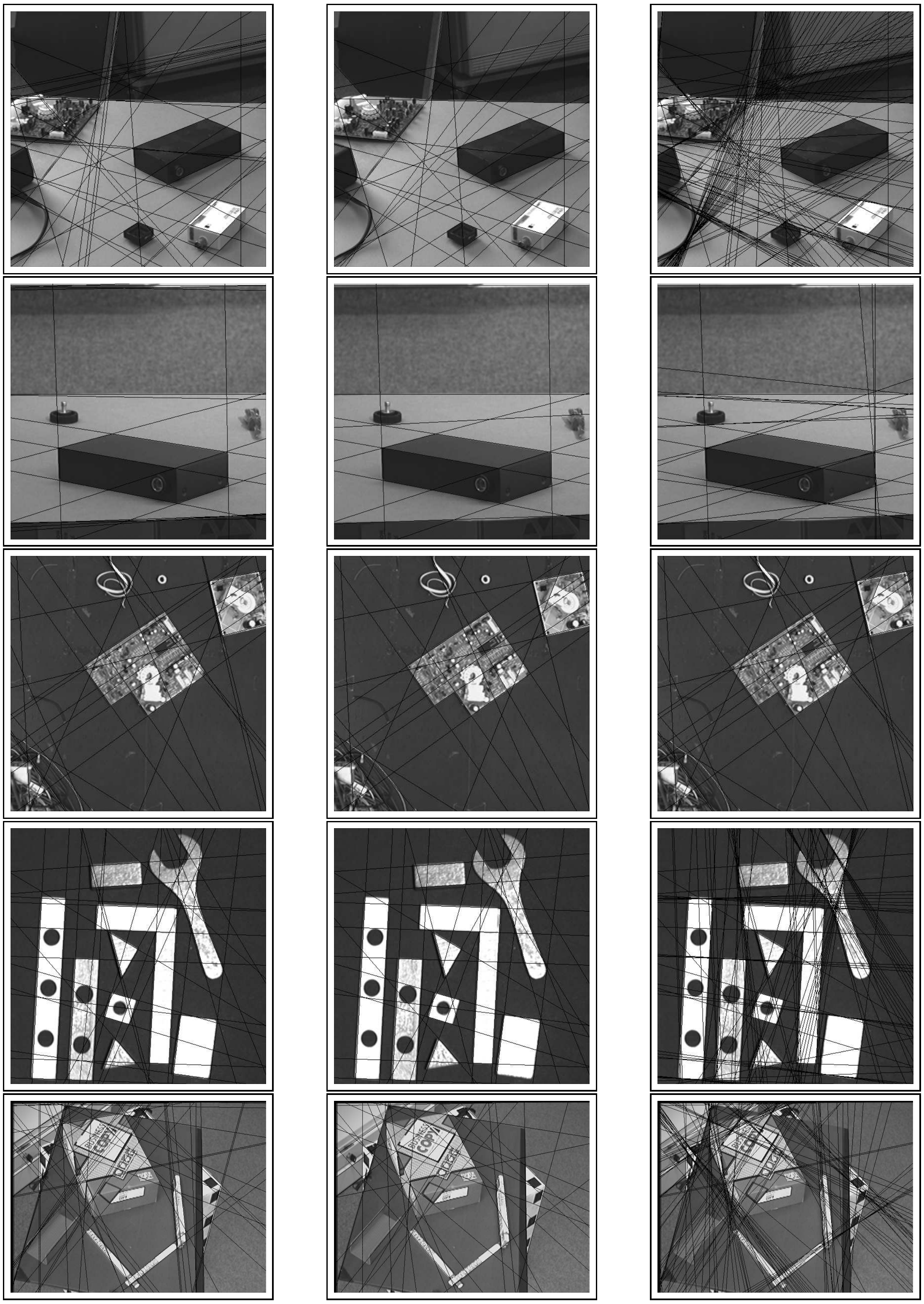


Abbildung 3.25: Vergleich der Verfahren zur Maximumsuche, erster Teil, links Distanzverfahren, Mitte Verfahren von Pauli, rechts Verfahren von Kunze

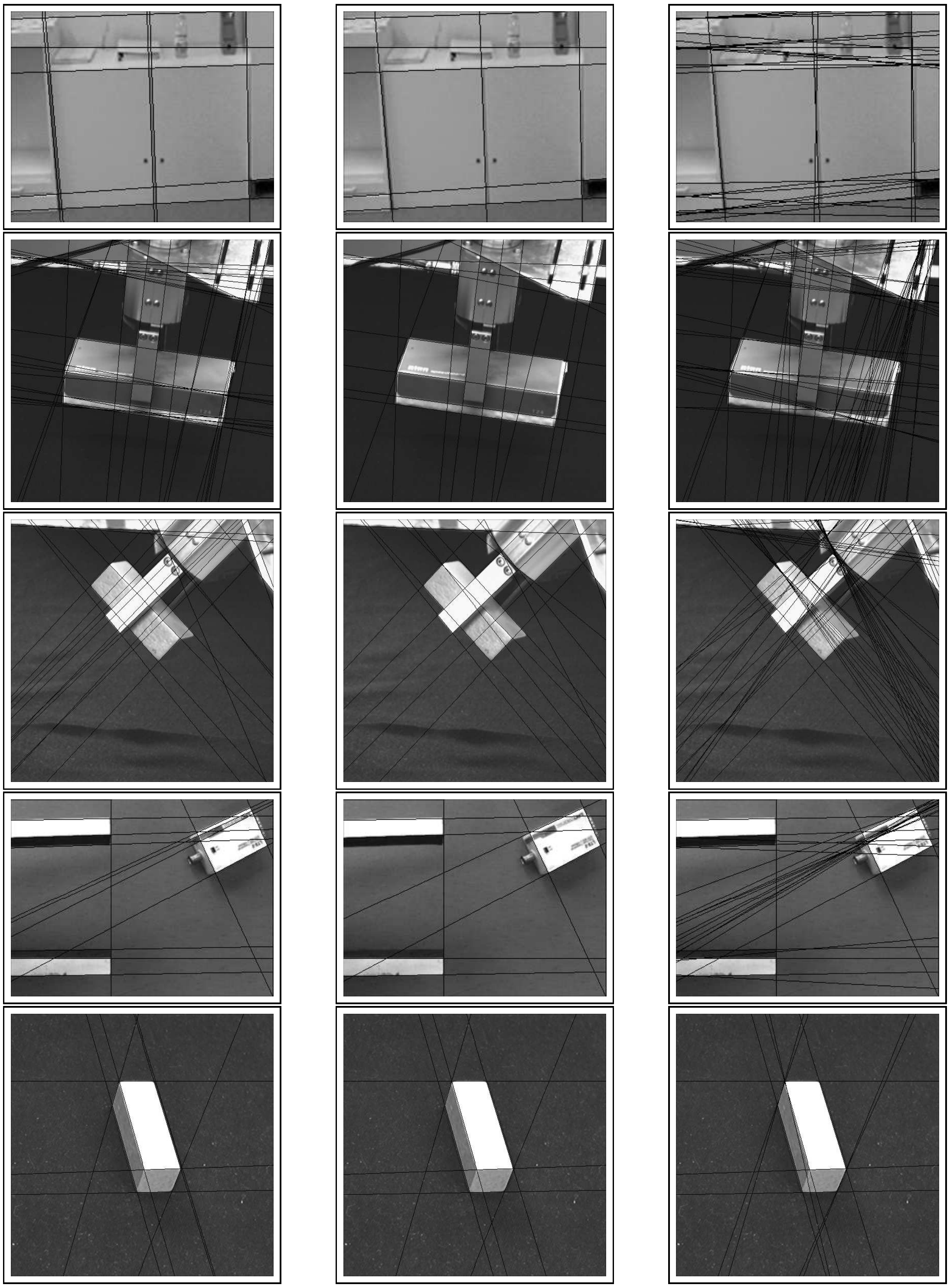


Abbildung 3.26: Vergleich der Verfahren zur Maximumsuche, zweiter Teil, links Distanzverfahren, Mitte Verfahren von Pauli, rechts Verfahren von Kunze

4 Gewinnung einer 3D-Repräsentation

Zur Akkumulation von Repräsentationen eines Objekts aus verschiedenen Ansichten in einem Wahrnehmungs-Handlungs-Zyklus wird zuerst aus jeder Ansicht eine eigenständige Repräsentation gewonnen. Jedes Element der Repräsentation ist dabei konfidenzbewertet, ihm ist also eine Konfidenz (Wahrscheinlichkeit) für das tatsächliche Auftreten einer korrespondierenden Struktur im Bild zugewiesen. Das System drückt damit aus, wie *sicher* es sich bei der Extraktion dieses Elements ist. Diese Konfidenz ermöglicht die Akkumulation der Elemente über mehrere Repräsentationen und läßt schließlich eine Gesamtaussage – ebenfalls in Form einer Konfidenz – zu jedem Element über alle aufgenommenen Ansichten zu. Es werden also frühe Entscheidungen vermieden, vielmehr wird eine spätere Entscheidung durch die Akkumulation erleichtert. Die spätere Entscheidung wird dann in einem größeren, globaleren Zusammenhang getroffen.

In Abschnitt 4.1 wird eine Entscheidung getroffen, welche Elemente eine Repräsentation enthält. Abschnitt 4.2 beschreibt das Finden von Korrespondenzen in zwei Bildern, aus denen in Abschnitt 4.3 schließlich eine 3D-Repräsentation einer Ansicht gewonnen wird. Abbildung 4.1 verdeutlicht die prinzipielle Vorgehensweise bei der Gewinnung einer 3D-Repräsentation.

Zur besseren Unterscheidung zwischen einer aus einem Bildpaar gewonnenen Repräsentation und der akkumulierten Repräsentation werden im folgenden die Begriffe Bildrepräsentation für ein Bildpaar sowie Objektrepräsentation für den akkumulierten Fall verwendet.

In diesem Kapitel wird also die Bestimmung einer Bildrepräsentation zu jedem Bildpaar beschrieben. Diese einzelnen Repräsentationen verschiedener Ansichten werden anschließend, wie in Kapitel 5 beschrieben, über eine Menge von Bildpaaren akkumuliert, um eine stabile Objektrepräsentation zu erhalten.

4.1 Wahl der Elemente einer Repräsentation

Unter einer Repräsentation eines Objekts verstehen wir eine formalisierbare Darstellung eines Objekts, die sich daher auch auf einem Computer realisieren läßt.

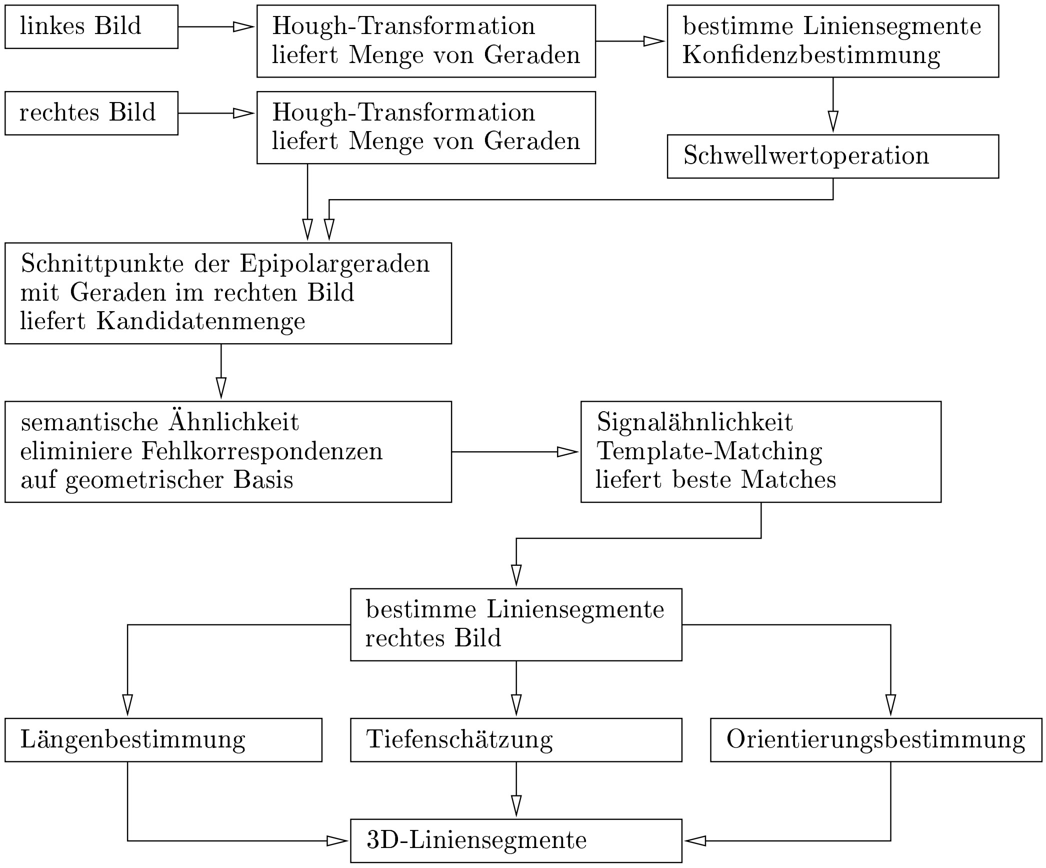


Abbildung 4.1: Vorgehen bei der Gewinnung einer Bildrepräsentation

Diese Objektrepräsentation ist in dieser Arbeit nicht explizit definiert (wie z.B. in modellbasierten Verfahren [Yu91]), sondern stabilisiert sich in enger Interaktion von Seherfahrung, Roboterhandlung und Vorwissen.

Die Hough-Transformation liefert zu jedem Bild eine Menge von Hough-Geraden, die im Bild stückweise mit Objektkanten zusammenfallen. Da Objekte in der Regel räumlich begrenzte Strukturen darstellen, definieren wir eine Objektrepräsentation, die auf lokalen Entitäten beruht. Eine Beschreibung von Objektkanten auf Basis von Liniensegmenten mit definiertem Anfang und Ende ist sicher eine vorteilhafte Darstellung, aber allein die Bestimmung von Anfang und Ende einer Kante im Bild ist sehr schwierig. Durch Rauschen oder Okklusion unterbrochene Kantenstücke zu einer Kante zusammenzufassen erfordert semantische Information und ist ebenfalls problematisch. Eckpunktinformationen werden an dieser Stelle nicht zu Hilfe genommen, siehe dazu [Ha99].

Einfacher zu extrahieren und auch einfacher zu akkumulieren sind Liniensegmente, die in 2D – also im Bild – konstante Länge haben. Grundlage für die Beschreibung von Kanten durch Mengen von Liniensegmenten sind die in Kapitel 3 bestimmten Hough-Geraden. Jede Hough-Gerade wird in Liniensegmente konstanter Länge eingeteilt, also diskretisiert. Für jedes Liniensegment läßt sich auf Basis von Gradientenbetrag und Gradientenorientierung eine Konfidenz bestimmen, die umso höher ist, je höher der Gradientenbetrag und je geringer die Orientierungsabweichung von Hough-Gerade und Gradient sind. An den Stellen, an denen die Hough-Gerade nicht entlang einer Objektkante verläuft, ist die Konfidenz nahe Null. Je besser die Objektkante anhand des Gradienten zu bestimmen ist, desto höher wird die Konfidenz, im Idealfall erreicht sie den Wert Eins. Die Konfidenz könnte also auch als Wahrscheinlichkeit des tatsächlichen Zusammenfallens eines Liniensegments mit einem Teil einer Kante interpretiert werden.

An dieser Stelle wird der Zusammenhang der Kollinearität explizit formalisiert. Eine Objektkante wird repräsentiert durch eine Menge von gleich orientierten Liniensegmenten – im allgemeinen entstanden aus einer einzigen Hough-Geraden. Rauschen und Okklusion betreffen an dieser Stelle nur noch einzelne Liniensegmente. Allerdings wird der Zusammengehörigkeit von aus einer Hough-Geraden entstandenen Liniensegmenten keine Beachtung geschenkt.Vielmehr muß ein Liniensegment globalen Support durch das Vorhandensein einer Hough-Geraden und lokalen Support durch Vergleich eines Segments der Hough-Geraden mit Gradientenbetrag und -orientierung im Bild haben. Hier sei darauf hingewiesen, daß diese explizite Formalisierung von Gestaltprinzipien problematisch ist, da, anders als bei der analytisch gut beschreibbaren starren Körperbewegung, deren Zusammenhang weitaus komplexer ist. Interessanterweise gibt es in der Entwicklungspsychologie starke Hinweise, daß der Bewegungszusammenhang von Säuglingen bereits sehr früh (nach etwa 16 Wochen), Gestaltprinzipien wie Parallelität oder Kollinearität erst verhältnismäßig spät (ab 7 Monaten) benutzt werden [KA98].

4.1.1 Liniensegmente in 2D

Ein 2D-Liniensegment läßt sich durch Lage, Orientierung und Länge eindeutig beschreiben, wobei die Länge für alle Liniensegmente in dieser Arbeit gleich ist. Die Liniensegmente sind also eine Diskretisierung aller Hough-Geraden. Darüber hinaus wird jedem Liniensegment eine Konfidenz zugeordnet. Sei $\hat{\mathcal{S}}$ die Menge alle 2D-Liniensegmente $\hat{\mathbf{s}} = (\hat{x}, \hat{y}, \theta, \hat{l}; \hat{c})$ mit $(\hat{x}, \hat{y})$ als Mittelpunkt und θ als Winkel der zugehörigen Hough-Geraden. Das Dach unterscheidet Parameter eines 2D-Liniensegment von gleichlautenden Parametern eines 3D-Liniensegments.

Gegeben sei eine Hough-Gerade mit den Parametern (r, θ) (siehe Abbildung 3.2). Zu dieser Hough-Geraden werden zuerst die beiden Schnittpunkte mit dem Bildrand berechnet, zwischen denen die Gerade im Bild verläuft. Dieses Geradensegment wird nun aufgeteilt in Liniensegmente der Länge $\hat{l}$ (in dieser Arbeit wurde $\hat{l} = 11$ Pixel verwendet). Die Konfidenz jedes Liniensegments ergibt sich nun als Durchschnitt der Konfidenzen von jedem der $\hat{l}$ Punkte des Liniensegments. Die Diskretisierung einer Hough-Gerade wird in Abbildung 4.2 gezeigt.

Die Konfidenz eines Punktes auf dem Liniensegment bestimmt sich aus Gradientenbetrag und Gradientenorientierung. Da der Gradientenbetrag nur schlecht als

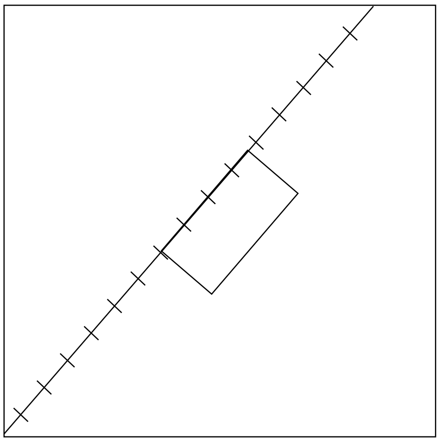


Abbildung 4.2: Diskretisierung einer Hough-Geraden.

Maß zu verwenden ist, wird wie schon bei der orientierungsselektiven, gewichteten Hough-Transformation der normalisierte Gradient E (siehe Abschnitt 3.5.1) zu Hilfe genommen. Die Konfidenz eines Punktes läßt sich genau wie in den Gleichungen (3.29) und (3.30) berechnen.

Abbildung 3.17 zeigt die Funktion f_θ, die geringen Winkelabweichungen eine hohe Konfidenz, größeren Abweichungen aber eine stark fallende Konfidenz zuweist.

4.1.2 Liniensegmente in 3D

Zur Darstellung eines Liniensegments in 3D werden wie in 2D Informationen über Lage, Orientierung und Länge des Segments benötigt. Dazu kommt ebenfalls eine Konfidenz. Sei also $\mathcal{S}$ die Menge aller 3D-Liniensegmente $\mathbf{s} = (x, y, z, \phi, \psi, l; c)$, wobei (x, y, z) der Mittelpunkt ist. Die Orientierung des Liniensegments wird in sphärischen Koodinaten (ϕ, ψ) dargestellt (siehe Abschnitt 2.3 und Abbildung 2.7). Die Länge ist in 3D nun nicht mehr konstant, sie wird vielmehr aus der Länge in 2D berechnet und ist abhängig vom Blickwinkel. Die folgenden Abschnitte beschreiben die Bestimmung von 3D-Liniensegmenten aus zwei Mengen von 2D-Liniensegmenten.

4.2 Finden korrespondierender Liniensegmente

Ein Liniensegment in 3D wird aus dessen Projektionen auf beide Kamerabilder (2D-Liniensegmente). Zu einem 2D-Liniensegment im linken Kamerabild muß also das korrespondierende 2D-Liniensegment im rechten Kamerabild gefunden werden. Diese Korrespondenzfindung basiert im wesentlichen auf der Signalähnlichkeit (photometrisches Matching), also auf der Annahme, daß in zwei Bildern korrespondierende Punkte in ihrer Umgebung eine sehr ähnliche Grauwertstruktur besitzen. Ein Fenster von $w \times h$ Pixeln (w, h ungerade) mit dem Mittelpunkt des Liniensegments des linken Bildes im Zentrum wird mit allen möglichen $w \times h$-Fenstern im rechten Bild verglichen, indem die Grauwertdifferenz für jeden Punkt des Fensters bestimmt und die Quadratsumme der Differenzen gebildet wird [BB82]. Derjenige Punkt im rechten Bild mit der kleinsten Quadratsumme weist die größte Ähnlichkeit in der Umgebung auf und ist der korrespondierende Punkt. Die Korrespondenzfindung beschränkt sich also auf die Mittelpunkte der Liniensegmente und wird durch Nutzung der Epipolargeometrie von einem zweidimensionalen auf ein eindimensionales Problem reduziert. Die Suche beschränkt sich auf eine Gerade im rechten Bild.

Für die Korrespondenzfindung wird ein photometrisches Standardverfahren (siehe auch [Fa93]) verwendet, das für diesen Spezialfall des konfidenzbasierten Matchens zweier Liniensegmente durch einige semantische Überlegungen ergänzt und leistungsfähiger gemacht wurde. Ausgehend vom linken Kamerabild werden zuerst alle Liniensegmente mit geringer Konfidenz gelöscht, da davon ausgegangen werden kann, daß das Segment der Hough-Gerade nicht entlang einer Objektkante verläuft. Ein Liniensegment mit Konfidenz kleiner als ein Schwellwert $c_{\min}$ wird im rechten Bild also höchstwahrscheinlich keine Korrespondenz haben. In dieser Arbeit wird $c_{\min} = \frac{1}{2}$ verwendet.

Strategie dieses Vorgehens ist es nicht, die Bildrepräsentation so gut wie möglich zu extrahieren, sondern vielmehr die Beseitigung der Unsicherheit der Akkumulation (Kapitel 5) zu überlassen. An dieser Stelle wird die Unsicherheit durch die Konfidenz ausgedrückt. Fehlkorrespondenzen sind erlaubt und werden bewußt nicht immer vermieden. Vielmehr werden mehrere potentielle Korrespondenzen zugelassen. Die Fehlkorrespondenz fällt bei der Akkumulation heraus, da sie nicht in anderen Ansichten bestätigt wird, während an dieser Stelle eine Fehlkorrespondenz ggfs. sogar eine höhere Konfidenz als die tatsächliche Korrespondenz haben kann. Umgekehrt ist es auch kein Nachteil, wenn die tatsächliche Korrespondenz in einer Bildrepräsentation nicht gefunden wird. Diese fehlende Korrespondenz kann wettgemacht werden. Wird das Liniensegment in mehreren anderen Ansichten erkannt, so hat die fehlende Bestätigung in dieser Bildrepräsentation bei der Akkumulation nur ein geringes Gewicht.

Um die Zahl der Korrespondenzen nicht unnötig ansteigen zu lassen, werden diejenigen gestrichen, die mit großer Sicherheit Fehlkorrespondenzen darstellen. Dieser Kompromiß zwischen Laufzeit und Güte des Verfahrens erweist sich als gute Lösung, denn diejenigen Korrespondenzen, die aufgrund von Schwellwerten ausgeschlossen werden, werden nur in verschwindend geringer Zahl fälschlich gelöscht. Die Einschränkung der Kandidatenliste auf semantischer Ebene beschränkt sich an dieser Stelle auf einige grundlegende Überlegungen.

Die Berechnung der Korrelation für jedes Liniensegment im linken Bild und jeden Punkte im rechten Bild ist sehr aufwendig und kann wie bereits erwähnt durch Nutzung der Epipolargeometrie von einem zweidimensionalen auf ein eindimensionales Problem reduziert werden [Fa93]. Die Suche beschränkt sich dann für jedes Liniensegment auf eine Gerade, die sogenannte Epipolargerade, im rechten Bild. Auf dieser Geraden (oder aufgrund von Ungenauigkeiten in der bei der Kamerakalibrierung bestimmten Projektionsmatrix (siehe Abschnitt 2.2.1) knapp über oder unter der Geraden) befindet sich der zu einem gegebenen Punkt (x_1, y_1) im linken Bild korrespondierende Punkt (x_2, y_2) im rechten Bild. Darüber hinaus muß das korrespondierende Liniensegment zu einer Hough-Geraden gehören, daher sind potentielle Kandidaten im rechten Bild alle Schnittpunkte von Hough-Geraden mit der Epipolargeraden. Abbildung 4.3 verdeutlicht die Kandidatenbestimmung unter Verwendung der Epipolargeometrie. Siehe auch [Ha99] für eine ausführliche Beschreibung.

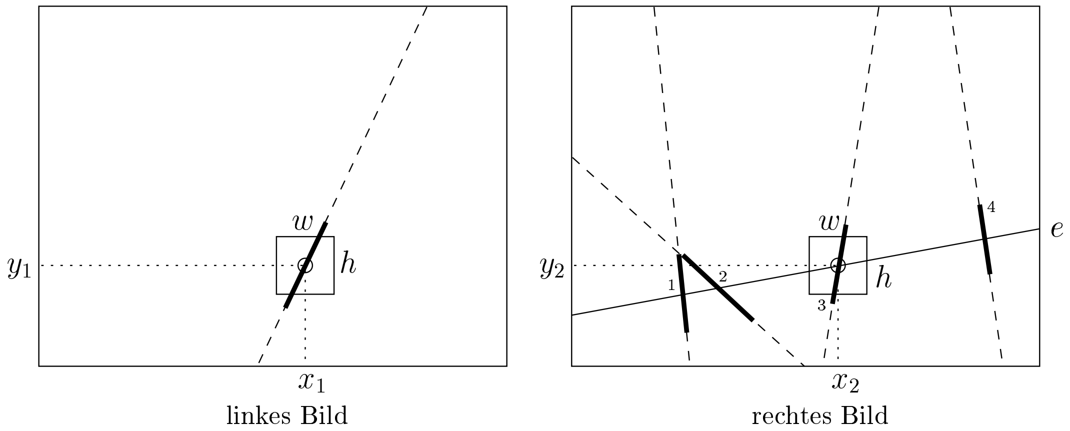


Abbildung 4.3: Kandidatenbestimmung unter Nutzung der Epipolargerade e zu (x_1, y_1), gestrichelte Linien sind Hough-Geraden, die dickeren Liniensegmente im rechten Bild sind die Kandidaten.

Mit einigen weiteren Überlegungen kann die Kandidatenmenge weiter eingeschränkt werden, um offenbar falsche Korrespondenzen, die sich zwangsläufig aus der alleinigen Nutzung der Signalähnlichkeitsfunktion ergeben, vorab zu eliminieren. Es werden nur diejenigen Kandidaten betrachtet, deren Orientierung sich nur gering von der des Liniensegments im linken Bild unterscheidet, die Differenz der Winkel beider Liniensegmente einen Schwellwert $\theta_{\max}$ also nicht überschreitet. Der Schnittpunkt von Hough-Gerade und Epipolargerade darf auch nur in begrenztem Maße von der Position des Punktes im linken Bild abweichen. Diese Abweichung wird als Disparität bezeichnet.

Eine einfache Heuristik mit dem Strahlensatz unter der vereinfachenden Annahme paralleler Kameras zeigt, inwiefern die Position im rechten Bild eingeschränkt werden kann (Abbildung 4.4). Seien f die Bildweite und d der horizontale Abstand der Kameras, sei ferner s die Entfernung des fokussierten Punktes zu den Kameras. Dann ist $x + y$ die Differenz der horizontalen Koordinaten der Projektionen auf beide Bildebenen. Tatsächlich sind die Kameras nicht parallel, dennoch liefert diese Heuristik einen Schätzwert für die maximale Disparität beider Projektionen. Der Strahlensatz liefert $\frac{x+y+d}{s+f} = \frac{d}{s}$, aufgelöst nach $x + y$ erhält man $x + y = \frac{df}{s}$. In Abbildung 4.3 würde Liniensegment 1 aufgrund der zu großen Disparität und Liniensegment 2 aufgrund der zu starken Winkeldifferenz von vornherein gestrichen. [Fa93] führt diese Überlegungen weiter aus.

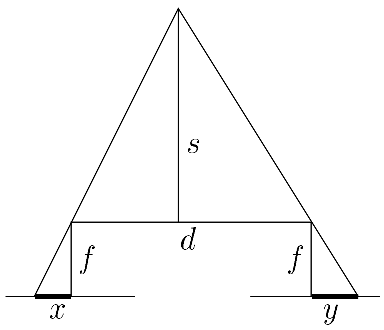


Abbildung 4.4: Schätzung der maximalen Disparität $x + y$.

Für jeden übriggebliebenen Kandidaten im rechten Bild wird nun ein 2D-Liniensegment bestimmt, dessen Mittelpunkt gerade der Schnittpunkt von Hough-Gerade und Epipolargerade ist. Unterschreitet die Konfidenz dieses Liniensegments die minimale Konfidenz $c_{\min}$, so wird auch dieser Kandidat noch verworfen. Dies wird begründet durch die Annahme, daß ein Kantenstück, das im linken Bild mit hoher Konfidenz erkannt wird, auch im rechten Bild eine hohe Konfidenz besitzt.

Nun braucht die Signalähnlicheitsfunktion nur noch auf die Kandidatenliste, d.h. die Menge potentieller Mittelpunkte von Liniensegmenten im rechten Bild, angewendet zu werden. Verwendet wird die normalisierte Kreuzkorrelation $k(x_1, y_1, x_2, y_2)$ von (x_1, y_1) im linken und (x_2, y_2) im rechten Bild [BB82, Fa93]. Seien w und h Breite und Höhe des Template-Matching-Fensters, dann seien $w' := \lfloor \frac{w}{2} \rfloor$ und $h := \lfloor \frac{h}{2} \rfloor$.

$$k(x_1, y_1, x_2, y_2) := \frac{1}{K} \sum_{x_0=-w'}^{+w'} \sum_{y_0=-h'}^{+h'} B_1^{\text{diff}}(x_1, y_1) B_2^{\text{diff}}(x_2, y_2) \tag{4.1}$$

mit

$$\begin{aligned} B_1^{\text{diff}}(x_1, y_1) &:= B_1(x_1 + x_0, y_1 + y_0) - \bar{B}_1(x_1, y_1) && (4.2) \\ B_2^{\text{diff}}(x_2, y_2) &:= B_2(x_2 + x_0, y_2 + y_0) - \bar{B}_2(x_2, y_2) && (4.3) \\ K &:= w \cdot h \cdot \sigma_1(x_1, y_1) \cdot \sigma_2(x_2, y_2) && (4.4) \end{aligned}$$

mit der Standardabweichung im Bild B definiert durch

$$\sigma(x, y) := \sqrt{\frac{1}{wh} \sum_{x_0=-w'}^{+w'} \sum_{y_0=-h'}^{+h'} (B(x + x_0, y + y_0) - \bar{B}(x, y))^2} \tag{4.5}$$

und der durchschnittlichen Intensität $\bar{B}$ an der Position (x, y) definiert durch

$$\bar{B}(x, y) := \frac{1}{wh} \sum_{x_0=-w'}^{+w'} \sum_{y_0=-h'}^{+h'} B(x + x_0, y + y_0) \tag{4.6}$$

Durch die Normalisierung der Kreuzkorrelation werden Unterschiede in der Grauwertverteilung in beiden Bildern durch abweichende Reflektanzeigenschaften und

Blendeneinstellungen teilweise ausgeglichen. Darüber hinaus liefert k einen Wert aus dem Intervall $[-1,+1]$ und damit durch $c_k := \frac{k+1}{2} \in [0,1]$ eine Konfidenz für die Korrespondenz von (x_1, y_1) und (x_2, y_2). Somit können mehrere potentielle Korrespondenzen sinnvoll in die Bildrepräsentation aufgenommen werden. Sei c_k die Korrespondenzkonfidenz der besten Übereinstimmung, dann werden alle potentiellen Korrespondenzen übernommen, deren Korrespondenzkonfidenz $\zeta \cdot c_k$, $\zeta \in [0,1]$ nicht unterschreitet. Der folgende Abschnitt zeigt, wie aus einem Korrespondenzpaar ein 3D-Liniensegment bestimmt wird.

4.3 Bestimmung eines 3D-Liniensegments aus zwei 2D-Liniensegmenten

Die Parameter des 3D-Liniensegments $(x, y, z, \phi, \psi, l; c)$ werden aus zwei korrespondierenden 2D-Liniensegmenten $(\hat{x}_1, \hat{y}_1, \theta_1, \hat{l}_1; \hat{c}_1)$ und $(\hat{x}_2, \hat{y}_2, \theta_2, \hat{l}_2; \hat{c}_2)$ in vier Schritten berechnet. Der Weltpunkt (x, y, z) ergibt sich aus den Mittelpunkten $(\hat{x}_1, \hat{y}_1)$ und $(\hat{x}_2, \hat{y}_2)$ der 2D-Liniensegmente. Die Orientierung (ϕ, ψ) läßt sich aus dem Weltpunkt (x, y, z) sowie den beiden Orientierungen θ_1 und θ_2 berechnen. Die Länge l wird approximiert durch den Abstand des 3D-Mittelpunkts zu den 3D-Koordinaten eines Endpunkts des Liniensegments. Die Konfidenz ergibt sich schließlich aus den Konfidenzen der 2D-Liniensegmente und der Korrespondenzkonfidenz gegeben durch die Kreuzkorrelation.

Sind beide Projektionsmatrizen gegeben, so lassen sich die Weltkoordinaten eines Punktes aus den beiden Projektionen berechnen [Fa93]. Auf eine ausführliche mathematische Erläuterung wird an dieser Stelle verzichtet und auf [Ha99] verwiesen, der diese Berechnung im Rahmen eines gemeinsamen Projekts implementiert hat. Dies gilt ebenso für die Orientierungsbestimmung aus den gegebenen Orientierungen zweier 2D-Liniensegmente. An dieser Stelle sei allerdings besonders hervorgehoben, daß die Orientierungsbestimmung mathematisch nicht immer möglich ist. Liegt das 3D-Liniensegment in der durch die Vektoren vom Mittelpunkt zu den optischen Zentren der Kameras aufgespannten Ebene, so ist das Verfahren instabil. In diesem Fall wird auf die Orientierungsberechnung verzichtet und das 3D-Liniensegment außer acht gelassen. Dies hat keine Auswirkung auf das Verfahren an sich, da dieses Liniensegment durch die selbst gesteuerte Bewegeung des Roboters in anderen Ansichten extrahiert werden kann.

Die Länge eines 3D-Liniensegments kann nicht exakt bestimmt werden, da die Längen beider 2D-Liniensegmente als konstant festgelegt sind. Die Projektionen eines 3D-Liniensegments auf beide Kamerabilder können jedoch unterschiedlicher Länge sein. Während die Mittelpunkte der beiden 2D-Liniensegmente korrespondieren – sie sind die Projektionen eines Weltpunktes – korrespondieren zwei Endpunkte der 2D-Liniensegmente nicht zwangsläufig. Daher wird zu einem Endpunkt des Liniensegments im linken Bild der Schnittpunkt der Epipolargeraden mit der Hough-Geraden als der korrespondierende Endpunkt im rechten Bild bestimmt. Daraus lassen sich wie oben die 3D-Koordinaten dieses Punkts bestimmen, und die Länge wird als die doppelte Entfernung zum Mittelpunkt approximiert. Seien also $\mathbf{p} = (x, y, z)$ die Koordinaten des Mittelpunkts und $\mathbf{p}' = (x', y', z')$ die eines Endpunkts, dann ergibt sich $l = 2 \cdot ||\mathbf{p} - \mathbf{p}'||_2$.

Die Konfidenz des 3D-Liniensegment bestimmt sich schließlich aus dem Durchschnitt der Konfidenzen der 2D-Liniensegmente multipliziert mit der Korrespondenzkonfidenz:

$$c := \frac{c_1 + c_2}{2} c_k = \frac{c_1 + c_2}{2} \frac{k+1}{2} = \frac{1}{4}(c_1 + c_2)(k+1) \tag{4.7}$$

Die Menge $\mathcal{S}$ aller so bestimmten 3D-Liniensegmente ist Teil einer Bildrepräsentation. Im folgenden Kapitel wird für jedes Bildpaar einer Bildsequenz die zugehörige Menge aller 3D-Liniensegmente, wie in diesem Kapitel beschrieben, bestimmt und mit einer Roboterstellung assoziiert. Durch Anwenden der Roboterbewegung auf alle Liniensegmente einer Bildrepräsentation können diese akkumuliert und dadurch stabilisiert werden.

5 Akkumulation im Wahrnehmungs-Handlungs-Zyklus

Um aus Bildern bzw. Bildsequenzen semantische Information zu extrahieren, bedient sich das menschliche visuelle System zusätzlicher Information und globaler Zusammenhänge. Einer dieser globalen Zusammenhänge ist der raum-zeitliche Zusammenhang der starren Körperbewegung, der in diesem Kapitel formalisiert und für das Akkumulieren von Information benutzt wird. Information wird nicht aus einer einzigen Ansicht gewonnen, sondern aus mehreren Ansichten, die das Objekt in einer bekannten Bewegung zeigen. Abbildung 5.1 zeigt das Vorgehen bei der Akkumulation mehrerer Bildrepräsentation zu einer Objektrepräsentation und verdeutlicht die Tatsache, daß nach der ausführlich beschriebenen Vorarbeit noch zwei wesentliche Probleme zu lösen sind: die Bewegung einer Bildrepräsentation und das anschließende Finden korrespondierender Liniensegmente in der Bildrepräsentation und der Objektrepräsentation.

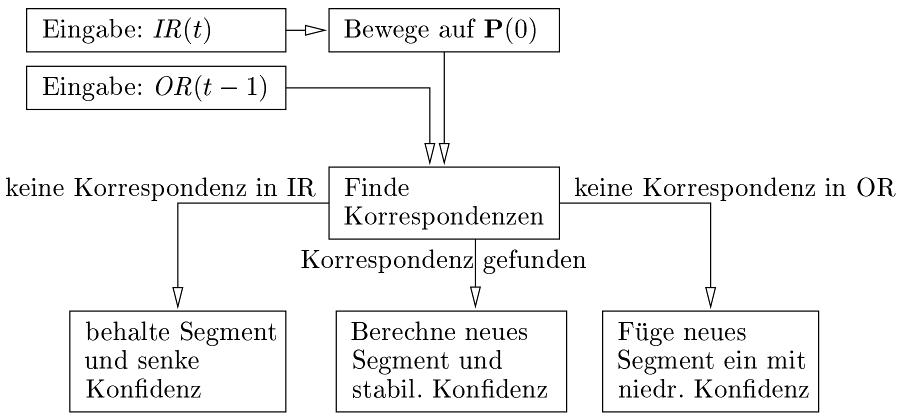


Abbildung 5.1: Vorgehen bei der Akkumulation der Bildrepräsentationen zu einer Objektrepräsentation

Das Verfahren arbeitet iterativ. Aus der ersten Aufnahme wird eine Bildrepräsentation $IR(0)$ gewonnen, die als initiale Objektrepräsentation $OR(0)$ genutzt wird. In den folgenden Schritten wird aus jeder weiteren Aufnahme erneut eine Bildrepräsentation $IR(t)$ gewonnen, die anschließend mit der aktuellen Objektrepräsentation $OR(t-1)$ in einem Aktualisierungsschritt zu einer neuen Objektrepräsentation

$OR(t)$ vereinigt wird.

Für die formale Beschreibung wird die folgende Notation definiert: Aus einem Bildpaar zum Zeitpunkt t wird eine Bildrepräsentation $IR(t)$ gewonnen. Die Objektrepräsentation zum Zeitpunkt t sei $OR(t)$. Damit ist der Akkumulationsvorgang induktiv definiert durch

$$OR(0) := IR(0) \tag{5.1}$$

$$OR(t) := OR(t-1) \oplus \mathbf{M}_t^0(IR(t)), \quad t > 0. \tag{5.2}$$

Die Bewegung der Bildrepräsentation $\mathbf{M}_t^0(IR(t))$ wird in Abschnitt 5.1 erläutert, der Aktualisierungsschritt, beschrieben durch den Operator $\oplus$, in Abschnitt 5.2.

Dieses Verfahren kann nun in jedem beliebigen t abgebrochen werden, $OR(t)$ ist dann die Ausgabe. Durch das Verschmelzen von Liniensegmenten, die nahe beieinander liegen, kann noch eine Verbesserung erzielt werden (Abschnitt 5.2.4).

Eine Entscheidung, die auf Basis einer Objektrepräsentation nach $t > 1$ Schritten getroffen wird, ist durch die t-malige Aktualisierung sicherer als eine Entscheidung, der nur eine einzige Bildrepräsentation zugrunde liegt.

5.1 Bewegung einer Repräsentation

Jeder Bildrepräsentation $IR(t)$ ist die Roboterstellung $\mathbf{P}(t)$ zugeordnet. Allen Objektrepräsentationen ist die Roboterstellung $\mathbf{P}(0)$ zugeordnet, also diejenige der initialen Bildrepräsentation. Die Vereinigung einer Bildrepräsentation mit einer Objektrepräsentation wie in Gleichung 5.2 zu einer aktualisierten Objektrepräsentation wird sinnvoll, wenn beide Repräsentationen bezüglich derselben Roboterstellung ausgerichtet sind. Diese gemeinsame Roboterstellung ist mit $\mathbf{P}(0)$ als die einer jeden Objektrepräsentation festgelegt. Jede Bildrepräsentation wird also so bewegt, daß die ihr zugeordnete Roboterstellung ebenfalls $\mathbf{P}(0)$ ist. Damit zeigen beide Repräsentationen das Objekt in derselben Ansicht. Die Kenntnis der Eigenbewegung ermöglicht damit die Korrespondenzfindung.

Die Bestimmung der Bewegung $\mathbf{M}_t$ von Nullstellung $\mathbf{P}_N = (0,0,0,0,0,0)$ auf $\mathbf{P}(t)$ wurde bereits in Abschnitt 2.2.3 diskutiert. Damit ergibt sich die Bewegung von $\mathbf{P}(t)$ nach $\mathbf{P}(t')$ zu $\mathbf{M}_t^{t'} = \mathbf{M}_{t'}\mathbf{M}_t^{-1}$, also Rückbewegung von $\mathbf{P}(t)$ auf Nullstellung und anschließende Bewegung auf $\mathbf{P}(t')$. Die neuen Mittelpunktkoordinaten eines jeden Liniensegments bestimmen sich somit zu

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \mathbf{M}_t^{t'} \begin{pmatrix} x \\ y \\ z \end{pmatrix}. \tag{5.3}$$

Die neuen Orientierungsparameter (ϕ', ψ') bestimmen sich ebenso nach den Gleichungen 2.10 und 2.11 in Abschnitt 2.3. Die Roboterstellung der initialen Objektrepräsentation $\mathbf{P}(0)$ darf nicht mit der Nullstellung $\mathbf{P}_N$ des Roboters verwechselt werden, die Parameter von $\mathbf{P}(0)$ sind im allgemeinen ungleich Null.

5.2 Akkumulation von Repräsentationen

Die Bildrepräsentation $\mathbf{M}_t^0(IR(t))$ sowie die Objektrepräsentation $OR(t-1)$ aus Gleichung 5.2 repräsentieren beide dasselbe Objekt in derselben Ansicht. Durch Rauschen und Ungenauigkeiten muß an dieser Stelle ein Korrespondenzproblem gelöst werden, um diejenigen Liniensegmente in beiden Repräsentationen zu finden, die dasselbe Kantenstück repräsentieren (Abschnitt 5.2.1). Ferner muß für jeden der drei folgenden Fälle festgelegt werden, wie die betroffenen Liniensegmente in der Objektrepräsentation angepaßt werden sollen (Abschnitt 5.2.2).

1. Korrespondenz wurde gefunden.
2. Ein Liniensegment der Objektrepräsentation besitzt keine Korrespondenz in der Bildrepräsentation.
3. Ein Liniensegment der Bildrepräsentation besitzt keine Korrespondenz in der Objektrepräsentation.

In Abschnitt 5.2.3 wird der Akkumulationalgorithmus noch einmal zusammengefaßt und durch Verschmelzen nahe beieinander liegender Liniensegment, wie in Abschnitt 5.2.4 beschrieben, noch verbessert.

5.2.1 Lösen des Korrespondenzproblems

Die Diskretisierung der Hough-Geraden stellt nicht sicher, daß zwei in zwei Aufnahmen dasselbe Kantenstück beschreibende Liniensegmente exakt denselben Mittelpunkt besitzen. Vielmehr kann das eine Liniensegment bezüglich des anderen um die halbe Länge in Richtung der Kante verschoben sein. Diese Tatsache muß bei der Korrespondenzfindung ebenso berücksichtigt werden wie Ungenauigkeiten bei der Tiefenschätzung und bei der Bewegung.

Als dasselbe Kantenstück beschreibend werden zwei Liniensegmente aus unterschiedlichen Repräsentationen angenommen, wenn ihre Orientierungen *in etwa* übereinstimmen und die Mittelpunkte *etwa* dieselben Weltkoordinaten haben. Um den Begriff *etwa* an dieser Stelle zu formalisieren und den besten Kandidaten für die Korrespondenz zu finden, werden zwei Metriken definiert, d_o für die Orientierungsdifferenz und d_p für die Abweichung der Position im Raum. Beide werden unabhängig voneinander betrachtet.

Seien $\mathbf{s}, \mathbf{s}'$ zwei Liniensegmente und $\mathbf{v}, \mathbf{v}'$ ihre Orientierungseinheitsvektoren, dann ist die Metrik bezüglich der Orientierungsabweichung definiert durch

$$d_o := \arccos(\mathbf{v} \cdot \mathbf{v}'), \tag{5.4}$$

also durch den durch $\mathbf{v}$ und $\mathbf{v}'$ eingeschlossenen Winkel (siehe Abbildung 5.2).

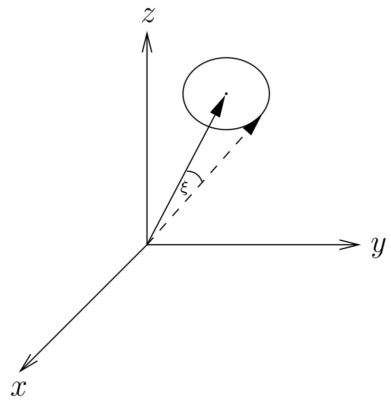


Abbildung 5.2: Bestimmung der Orientierungsdifferenz d_o

Potentielle Korrespondenzen brauchen nun nicht mehr betrachtet werden, wenn die Orientierungen zu stark voneinander abweichen, $d_o(\mathbf{s}, \mathbf{s}')$ also einen Schwellwert ξ überschreitet. ξ sollte klein gewählt werden (in dieser Arbeit ist $\xi = 10$), da zwei Liniensegmente mit zu starker Orientierungsdifferenz nicht sinnvoll zu einem neuen Liniensegment zusammengefügt werden können. Mittelt man die Orientierungsparameter zweier Liniensegmente, deren Orientierung zu stark differiert, so liegen die Mittelpunkte alle auf einer Objektkante, die Liniensegmente sind jedoch nicht in Richtung der Objektkante orientert. Abbildung 5.3 zeigt einen solchen Fall von zwei parallelen Objektkanten, die durch falsch orientierte Liniensegmente repräsentiert werden.

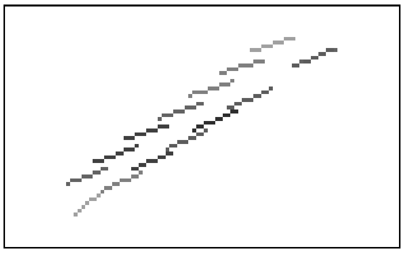


Abbildung 5.3: Kante mit falsch orientierten Liniensegmenten

Für alle Paare von Liniensegmenten $\mathbf{s}, \mathbf{s}'$ mit $d_o(\mathbf{s}, \mathbf{s}') < \xi$ muß noch überprüft werden, ob die Position im Raum übereinstimmt. Der Euklidische Abstand der Mittelpunkte der Liniensegmente ist eine denkbare Definition der Metrik d_p, jedoch wird damit nicht der Nebenbedingung Rechnung getragen, daß korrespondierende

Liniensegmente um die halbe Länge verschoben sein können. Eine Streckung des Suchraumes in Richtung eines Liniensegments ist eine bessere Lösung.

Seien $\mathbf{p}, \mathbf{p}'$ die Mittelpunkte der Liniensegmente $\mathbf{s}, \mathbf{s}'$. Sei $\mathbf{p}$ der Ursprung des Koordinatensystems $\mathcal{K}$, ferner sei die Orientierungsachse des Liniensegments die x-Achse dieses Koordinatensystems. In y- bzw. z-Richtung, also senkrecht zum Liniensegment, darf der Abstand von $\mathbf{p}$ und $\mathbf{p}'$ einen Schwellwert ω nicht überschreiten, in x-Richtung beträgt dieser Wert $\omega + \frac{l}{2}$, wobei l die Länge von $\mathbf{s}$ ist, um eine mögliche Verschiebung zu tolerieren. Der Suchraum ist somit ein Ellipsoid, wie Abbildung 5.4 ihn im Zweidimensionalen veranschaulicht.

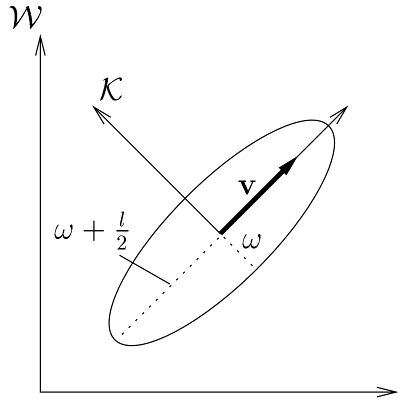


Abbildung 5.4: Elliptischer Suchraum der Positionsdifferenz in 2D

Die Formalisierung $\mathbf{f}_p(\mathbf{s}, \mathbf{s}')$ dieser Überlegungen stellt keine Metrik dar, da die Symmetriebedingung nicht erfüllt ist (Abbildung 5.5). Wählt man jedoch das Minimum

$$d_p(\mathbf{s}, \mathbf{s}') := \min\{\mathbf{f}_p(\mathbf{s}, \mathbf{s}'), \mathbf{f}_p(\mathbf{s}', \mathbf{s})\}, \tag{5.5}$$

so erhält man eine Metrik.

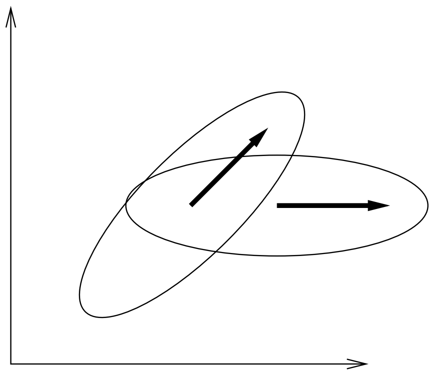


Abbildung 5.5: Die Symmetriebedingung ist bei $\mathbf{f}_p$ nicht erfüllt

Zur Berechnung von $\mathbf{f}_p$ werden die Koordinaten $\hat{\mathbf{p}}'$ von $\mathbf{p}'$ in $\mathcal{K}$ benötigt ($\hat{\mathbf{p}}$ ist der Ursprung von $\mathcal{K}$, $\hat{\mathbf{v}}$ ist die x-Achse). Man erhält $\hat{\mathbf{p}}'$ durch Translation um $-\mathbf{p}$ und anschließende Rotation, so daß $\mathbf{v}$ mit der x-Achse übereinstimmt. Diese Rotation läßt sich sehr gut in Form einer Quaternion beschreiben, da Rotationsachse und -winkel bekannt sind: Die Rotationsachse ergibt sich zu

$$\mathbf{q}' = \frac{\mathbf{v} + \mathbf{e}_x}{||\mathbf{v} + \mathbf{e}_x||} \tag{5.6}$$

als die Mittelachse zwischen $\mathbf{v}$ und x-Achse gegeben durch den Einheitsvektor $\mathbf{e}_x$. Der Rotationswinkel beträgt 180°. Somit erhält man die Rotationsquaternion

$$\mathbf{q} = \cos\frac{\pi}{2} + \mathbf{q}'\sin\frac{\pi}{2} \tag{5.7}$$

und schließlich

$$\hat{\mathbf{p}}' = \mathbf{q}(\mathbf{p}' - \mathbf{p})\mathbf{q}^{-1} = \begin{pmatrix} \hat{x}' \\ \hat{y}' \\ \hat{z}' \end{pmatrix}. \tag{5.8}$$

Gemäß der obigen Forderungen ergibt sich

$$\mathbf{f}_p(\mathbf{s}, \mathbf{s}') := \sqrt{\left(\frac{\omega}{\omega + \frac{l}{2}}\hat{x}'\right)^2 + \hat{y}'^2 + \hat{z}'^2} \tag{5.9}$$

und damit die Metrik $d_p(\mathbf{s}, \mathbf{s}')$ aus Formel 5.5. Ist $d_p(\mathbf{s}, \mathbf{s}') < \omega$, so sind $\mathbf{s}, \mathbf{s}'$ verbleibende potentielle Korrespondenzen. Aus der Menge dieser potentiellen Korrepondenzen wird dasjenige Paar gewählt, dessen Positionsdifferenz d_p minimal ist.

In Abhängigkeit der Schwellwerte ξ und ω kann also mit den oben eingeführten Metriken bestimmt werden, ob die zwei Liniensegmente $\mathbf{s}$ und $\mathbf{s}'$ vermutlich dasselbe Kantenstück beschreiben oder aber verschiedene. Dabei müssen $\mathbf{s}$ und $\mathbf{s}'$ nicht zwangsläufig aus verschiedenen Repräsentationen stammen, wie es bei der Akkumulation der Fall ist, sondern können auch in derselben Repräsentation enthalten sein, wovon beim Verschmelzen einer Objektrepräsentation, wie in Abschnitt 5.2.4 beschrieben, Gebrauch gemacht wird.

5.2.2 Zusammenfügen von zwei Repräsentationen

Sollen die Liniensegmente $\mathcal{S}_{OR}$ einer Objektrepräsentationen und die Liniensegmente $\mathcal{S}_{IR}$ einer Bildrepräsentation zusammengefügt werden, so treten die bereits erwähnten drei Fälle auf. Wird eine Korrespondenz gefunden (Fall 1), seien also $\mathbf{s} \in \mathcal{S}_{OR}$ und $\mathbf{s}' \in \mathcal{S}_{IR}$ und gelte $d_o(\mathbf{s}, \mathbf{s}') < \xi$ sowie $d_p(\mathbf{s}, \mathbf{s}') < \omega$, so wird aus den Liniensegmenten $\mathbf{s}$ und $\mathbf{s}'$ ein neues Liniensegment $\tilde{\mathbf{s}} \in \tilde{\mathcal{S}}_{OR}$ berechnet.

Da sich die Objektrepräsentation im Laufe der Bewegung stabilisieren soll, ist es sinnvoll, daß die Liniensegment der Bildrepräsentation mit der Zeit an Gewicht verlieren, da ein mehrfach aktualisiertes Liniensegment einer Objektrepräsentation sicherere Information repräsentiert als ein Liniensegment einer einzelnen Bildrepräsentation. Daher wird jedem Liniensegment $\mathbf{s}$ die Anzahl der Aktualisierungsschritte $u_{\mathbf{s}}$ zugeordnet. Ist $\mathbf{s}$ Element nur einer Bildrepräsentation, so ist $u_{\mathbf{s}} = 1$.

Auf diese Weise läßt sich $\tilde{\mathbf{s}}$ als gewichteter Mittelwert der Ausgangssegmente $\mathbf{s}$ und $\mathbf{s}'$ berechnen. Es ist

$$\tilde{x} = \frac{u_{\mathbf{s}} x + x'}{u_{\mathbf{s}} + 1}, \tag{5.10}$$

ebenso für die weiteren Parameter $\tilde{y}, \tilde{z}, \tilde{\phi}, \tilde{\psi}, \tilde{l}; \tilde{c}$. Bei der Mittelwertbestimmung der Winkel ist die Periodizität der Wertebereiche geeignet zu berücksichtigen.

Die Fälle 2 und 3 werden auf Fall 1 zurückgeführt, indem das Fehlen einer Korrespondenz so behandelt wird, als sei eine Korrespondenz gefunden worden, die Konfidenz des korrespondierenden Liniensegments sei jedoch Null, die übrigen Parameter stimmen überein. Dies formalisiert die Idee, daß ein Liniensegment einer Objektrepräsentation, das keine Korrespondenz in der Bildrepräsentation besitzt, unverändert bleibt, die Konfidenz jedoch gesenkt wird. Wie stark die Konfidenz gesenkt wird, hängt von der Anzahl der Aktualisierungsschritte ab. Je öfter ein Liniensegment aktualisiert wurde, desto mehr Sicherheit repräsentiert es und desto geringer ist die Absenkung der Konfidenz. Demnach ist die Konfidenz

$$\tilde{c} = \frac{u_{\mathbf{s}}}{u_{\mathbf{s}} + 1} c. \tag{5.11}$$

Besitzt ein Liniensegment einer Bildrepräsentation keine Korrespondenz in der Objektrepräsentation, so wird es mit einer geringeren Konfidenz in die Objektrepräsentation aufgenommen. Wie gering die Konfidenz ist, bestimmt sich aus der Zahl der Aktualisierungsschritte u_{OR} der Objektrepräsentation. Je öfter die Objektrepräsentation bereits aktualisiert wurde, desto unsicherer ist dieses neu hinzugefügte Liniensegment. Somit ergibt sich

$$\tilde{c} = \frac{1}{u_{OR}} c'. \tag{5.12}$$

5.2.3 Iterativer Akkumulationsalgorithmus

Der Akkumulationsalgorithmus kann nun noch einmal wie folgt zusammengefaßt werden (siehe auch die Ablaufdiagramme in den Abbildungen 1.3 und 5.1).

In einer engen Interaktion von Wahrnehmung und Handlung wird in jedem Iterationsschritt ein Bildpaar aufgenommen, aus dem eine Bildrepräsentation gewonnen wird. Anschließend wird das Objekt bewegt. Der Akkumulationsalgorithmus wird mit der ersten Bildrepräsentation initialisiert, die Stellung des Roboters wird gespeichert und mit der initialen Objektrepräsentation verknüpft. Jede weitere Bildrepräsentation wird so transformiert, daß sie der mit der Objektrepräsentation verknüpften Roboterstellung entspricht. Damit lassen sich mit einer einfachen Methode Korrespondenzen finden, so daß die Objektrepräsentation mit der Bildrepräsentation aktualisiert wird.

Die Liniensegmente verändern dabei ihre Parameter und stabilisieren sich schließlich, da der Einfluß jeder neuen Bildrepräsentationen mit wachsendem t sinkt. Daher stabilisiert sich insbesondere die Konfidenz, mit der die Unsicherheit der Information einzelner Bildrepräsentationen modelliert wird. Eine hohe Konfidenz nach mehreren Schritten läßt eine sicherere Aussage zu als eine hohe Konfidenz in einer Bildrepräsentation. Fehler in einzelnen Repräsentationen verlieren während der Akkumulation an Gewicht, so daß schließlich eine stabile Objektrepräsentation gewonnen wird.

5.2.4 Verschmelzen benachbarter Liniensegmente

Aus verschiedenen Gründen kommt es vor, daß ein Kantenstück durch zwei verschiedene Liniensegmente in einer Objektrepräsentation repräsentiert wird. Einerseits findet die Hough-Transformation bei unscharfen, abgerundeten Kanten, wie es bei dem Pappkarton in der Abbildung auch der Fall ist, zwei parallele Hough-Geraden, die beide in die Objektrepräsentation aufgenommen werden. Andererseits wird die Korrespondenz im Akkumulationsschritt gelegentlich durch Ungenauigkeiten in der Tiefenschätzung nicht gefunden und damit ein zusätzliches Liniensegment in die Objektrepräsentation aufgenommen (Abbildung 5.6).

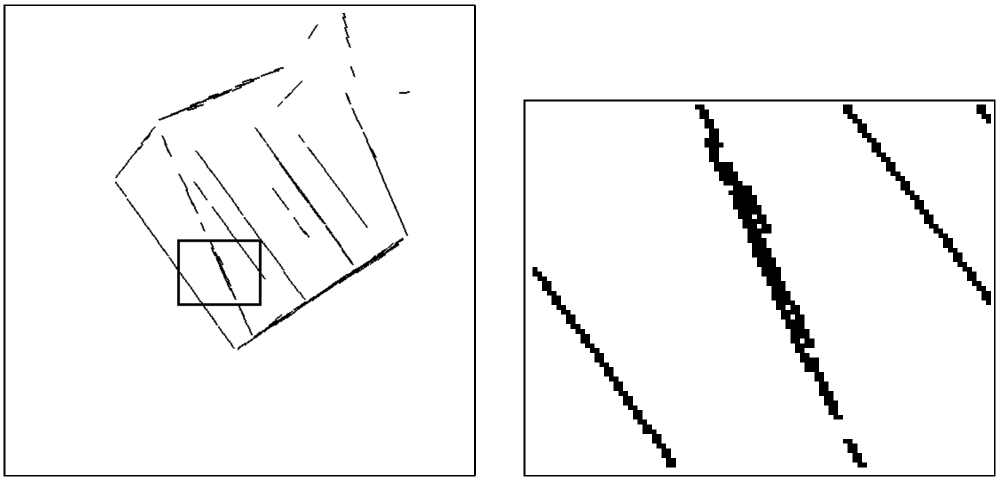


Abbildung 5.6: Beispiel doppelt repräsentierter Kanten

In beiden Fällen wird i.a. in den folgenden Aktualisierungsschritten genau eines der beiden Liniensegmente aktualisiert. Dabei rückt das aktualisierte Liniensegment näher an das nicht aktualisierte heran. Wird nun innerhalb der Objektrepräsentation nach Korrespondenzen gesucht, so können diese Doppelrepräsentationen entdeckt und verschmolzen werden (Abbildung 5.7).

Die zu verschmelzenden Liniensegmente repräsentieren ein und dasselbe Kantenstück. Daher werden die Parameter des verschmolzenen Liniensegments wie in Abschnitt 5.2.2 als gewichteter Mittelwert bestimmt. Da diese Liniensegmente jedoch Elemente einer einzigen Objektrepräsentation sind, es sich hier also um Doppelrepräsentationen eines Kantenstücks handelt, werden die Konfidenzen addiert. Dabei

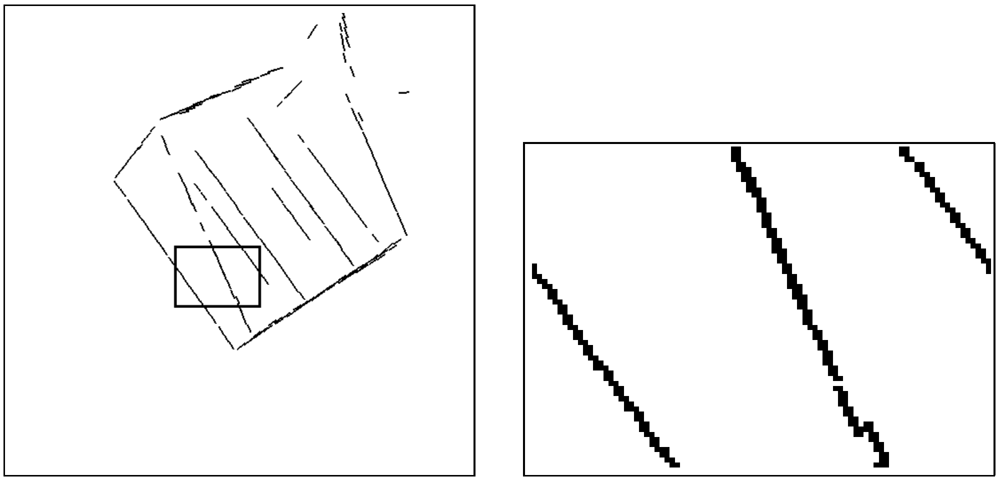


Abbildung 5.7: Verschmelzen doppelt repräsentierter Kanten

darf die Summe den Wert eins nicht überschreiten.

Es gilt also

$$\tilde{c} = \begin{cases} c + c' & c + c' < 1 \\ 1.0 & \text{sonst} \end{cases} \tag{5.13}$$

5.2.5 Vergleich mit anderen Verfahren

Ziel des vorgestellten Verfahrens ist es, Objektrepräsentationen aus Bildsequenzen zu erzeugen. In vielen Systemen sind solche Repräsentationen manuell definiert (z.B. [Yu91, LTC97, Wi97, HH89, Ri97]) bzw. werden unter idealisierten Bedingungen (homogener Hintergrund, kontrollierte Beleuchtung) extrahiert (z.B. [La92]). Ziel dieser Arbeit ist es, Objektrepräsentationen in realistischen Situationen unter Zuhilfenahme der engen Interaktion von Wahrnehmung und Handlung zu extrahieren, ganz im Sinne der Brooksschen Forderung der Körperlichkeit und Saturiertheit [Br91].

Dabei wird von der Möglichkeit profitiert, daß Wahrnehmung kein passiver Vorgang sein muß, sondern in enger Verknüpfung mit Handlungen, hier die Bewegung eines Roboters, formuliert werden kann [Ko92, So97]. Dies eröffnet die Perspektive von Roboteranwendungen, in denen sich visuell gesteuerte Roboter Repräsentationen der zu manipulierenden Objekte vor ihrer Anwendung ohne aufwendige manuelle Informationen aneignen können.

Das hier vorgestellte Verfahren ist eine Erweiterung eines Algorithmus, der Teil des Objekterkennungssystems ORASSYLL [Kr98a] ist. In ORASSYLL werden rein translatorische Bewegungen und ein einfaches Greiferverfolgungsverfahren benutzt, um eine Situation zu erschaffen, in der eine 2D-Ansicht eines Objekts vor variierendem Hintergrund und unter variierender Beleuchtungssituation erscheint (siehe Abbildung 5.8): Mit Hilfe eines Lernalgorithmus wird dann eine stabile 2D-Repräsentation eines Objekts aus einer Menge stark gestörter 2D-Repräsentationen extrahiert. Dieser in [Kr98a] vorgestellte Lernalgorithmus funktioniert im Batchmode, wurde also auf eine bestimmte Anzahl von Bildern angewandt.

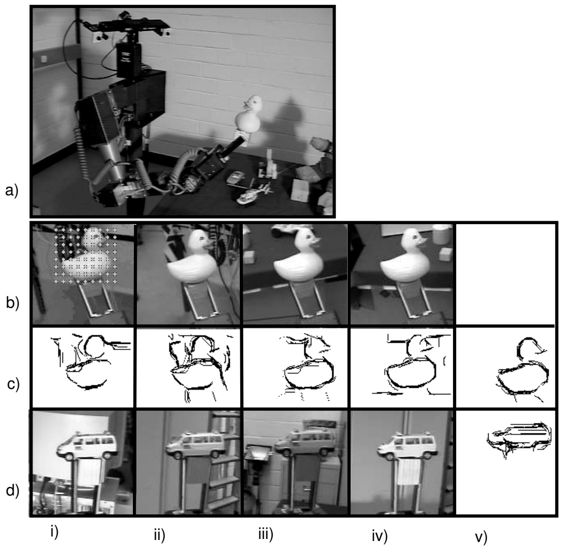


Abbildung 5.8: Beispiel einer rein translatorischen Bewegung. Rechts befindet sich die aus der Menge der gestörten Repräsentationen gewonnene Repräsentation. (aus [Kr98a])

Das in dieser Arbeit vorgestellte Verfahren erweitert den in ORASSYLL verwendeten Algorithmus in dreierlei Hinsicht:

1. Anstatt nur rein translatorische Bewegungen zuzulassen, wird in dieser Arbeit mit allgemeinen starren Körperbewegungen gearbeitet.
2. Anstelle der 2D-Repräsentationen aus Einzelbildern werden 3D-Repräsentationen aus Stereobildern extrahiert.
3. Der hier vorgestellte Algorithmus arbeitet nicht im Batchmode, sondern kann online iterativ angewendet werden.

Koch extrahiert in [Ko95] 3D-Szenen aus einer Stereosequenz. Es werden, ähnlich wie in dieser Arbeit, Tiefenschätzungen aus verschiedenen Stereobildern miteinander zu einer besseren Repräsentation verschmolzen. Das Ziel der Kochschen Arbeit ist dabei weniger die Objektbeschreibung, sondern eine möglichst gute Rekonstruktion der Szene und ein Übergang zu den in der Virtual Reality üblichen texturierten Wire-frames.

Die vorliegende Arbeit unterscheidet sich wie folgt von der Arbeit von Koch:

1. Das Objekt kann durch die Bewegung automatisch vom Hintergrund segmentiert werden. Da die zum Hintergrund gehörenden Liniensegmente einer Bildrepräsentation nach Ausführung der Bewegung im allgemeinen keine Korrespondenz in der Objektrepräsentation haben, verlieren sie – wie auch fehlerhafte Liniensegmente – mit jeder Iteration an Gewicht. Die Konfidenz stabilisiert sich nahe Null.

2. Das Objekt wird durch Liniensegmente und nicht durch Punkte repräsentiert.

3. Die Zielstellung der vorliegenden Arbeit ist nicht die Rekonstruktion eines Objekts, sondern die Extraktion von für potentielle Handlungen notweniger Information.

In Kapitel 6 wird auf Punkt 1 Bezug genommen und die Elimination des Hintergrundes anhand eines einfachen Beispiels verdeutlicht.

6 Experimente und Auswertung

Anhand von fünf Bilderserien mit jeweils elf verschiedenen Ansichten wird in diesem Kapitel die Leistungsfähigkeit des entwickelten Verfahrens unter Beweis gestellt. Zu jeder Bilderserie werden nebeneinander das linke Kamerabild, die Projektion der Bildrepräsentation auf die Bildebene der linken Kamera sowie die Projektion der mit diesem Bild aktualisierten Objektrepräsentation gezeigt. Abgeschlossen wird jede Bilderserie mit der verschmolzenen Objektrepräsentation.

Die erste Bilderserie (Abbildungen 6.1 und 6.2) zeigt einen Pappkarton, dessen Kanten in den meisten Bildrepräsentationen gut extrahiert werden, jedoch enthalten viele Bildrepräsentationen Fehler. Oft fehlt die linke vertikale Kante (Abbildung 6.1, Zeilen 2, 3 und 4), oder die untere horizontale Kante ist falsch orientiert (Abbildung 6.1, Zeile 5 und Abbildung 6.2, Zeile 5). Die kurze Kante in der Mitte der Deckelfläche wird nur in wenigen Ansichten extrahiert, jedoch ist sie wie auch die anderen teilweise fehlenden Kanten in der akkumulierten Objektrepräsentation enthalten. Die in einigen Bildrepräsentationen fehlerhaft repräsentierten Kanten sind jedoch korrekt repräsentiert. Die akkumulierte Repräsentation ist besser als jede einzelne Bildrepräsentation. Die Unsicherheit der Information der Bildrepräsentationen wurde eliminiert.

Die Abbildungen 6.3 und 6.4 zeigen, daß die Leistungsfähigkeit des Verfahrens nicht von der Reihenfolge der Bewegungen abhängt. Die Bilder zeigen den Pappkarton aus der vorigen Bilderserie, der genau in umgekehrter Reihenfolge bewegt wird. Die Bildrepräsentationen sind exakt dieselben, und auch die Objektrepräsentation enthält wieder alle Kanten des Kartons und keine fehlerhafte Information.

Mit einer Batteriekiste wird in den Abbildungen 6.5 und 6.6 in einem weiterem Beispiel gezeigt, daß fehlerhafte Information mittels Akkumulation beseitigt werden kann. Die Fächer der Kiste lassen sich nicht in jeder Ansicht extrahieren, sie fehlen in mehreren Bildrepräsentationen teilweise oder völlig, dennoch sind sie in der Objektrepräsentation nach Durchführung der Akkumulation enthalten. Die letzten drei Ansichten lassen keine gute Bildrepräsentation zu, lediglich einige Außenkanten können extrahiert werden. Darüber hinaus kommt es durch die vielen parallelen Kanten in der Kiste zu Fehlkorrespondenzen, die in der zweiten und dritten Objektrepräsentation in Abbildung 6.5 enthalten sind, jedoch im Verlauf der Akkumulation wieder eliminiert werden.

Auch der Ordner in den Abbildungen 6.7 und 6.8 wird in der Objektrepräsentation gut gezeigt, jedoch fehlt hier die untere Kante. Diese Kante ist zwar in mehreren Bildrepräsentationen enthalten, jedoch nur mit geringer Konfidenz. Darüber hinaus ist sie in der vierten Bildrepräsentation in Abbildung 6.7 und in der ersten in Abbildung 6.8 falsch orientiert, so daß diese Kante im Verlauf der Akkumulation zu wenig Bestätigung findet. Bis zur siebten Objektrepräsentation ist diese Kante jedoch enthalten, sie wird durch einige weniger gute Bildrepräsentationen eliminiert. Das Problem liegt darin begründet, daß das Objekt so ungünstig bewegt wurde, daß die untere Kante fast immer etwa horizontal gezeigt wird. Damit ist die Orientierungsbestimmung instabil (vgl. Abschnitt 4.3). Alternativ könnte man sich eine Bewegungsstrategie überlegen, die das Objekt so zeigt, daß Kanten nicht mehrfach nur instabil extrahierbar sind.

Die Abbildung 6.9 verdeutlicht schließlich in einer kurzen Bilderserie mit fünf Bildern die bereits erwähnte Eigenschaft dieses Verfahrens, daß das Objekt durch die Akkumulation vom Hintergrund segmentiert wird. Das Blatt Papier, das nicht bewegt wird, wird in allen Bildrepräsentationen gut extrahiert. Durch die Bewegung der Bildrepräsentation wird auch das Blatt Papier bewegt und findet in den folgenden Bildrepräsentationen keine Bestätigung, so daß es aus der Objektrepräsentation schon nach dem vierten Schritt vollständig eliminiert ist.

Die beschriebenen Bilderserien zeigen nicht nur die Leistungsfähigkeit des Verfahrens, sondern auch dessen Robustheit. Für keine der Bilderserien war eine Anpassung der Parameter notwendig, alle Akkumulationen wurden mit den im Variablen- und Parameterverzeichnis im Anhang angegebenen Standardparameterwerten durchgeführt.

Die folgende Tabelle zeigt die Rechenzeiten, die auf einer CPU mit 440 MHz für einen Schritt, bestehend aus Extraktion einer Bildrepräsentation und anschließender Aktualisierung der Objektrepräsentation, benötigt werden:

Bildrepräsentation	Bildglättung	10 ms
	Kantendetektion	20 ms
	Kontrastnormalisierung	100 ms
	Hough-Transformation	120 ms
	Maximumsuche	70 ms
	Bestimmung der 3D-Liniensegmente	140 ms
Objektrepräsentation	Akkumulationsschritt	50 ms
Zeitbedarf pro Ansicht		etwa 5 s

Die Akkumulation von elf Bildern wie in den gezeigten Bilderserien benötigt also etwa eine Minute.

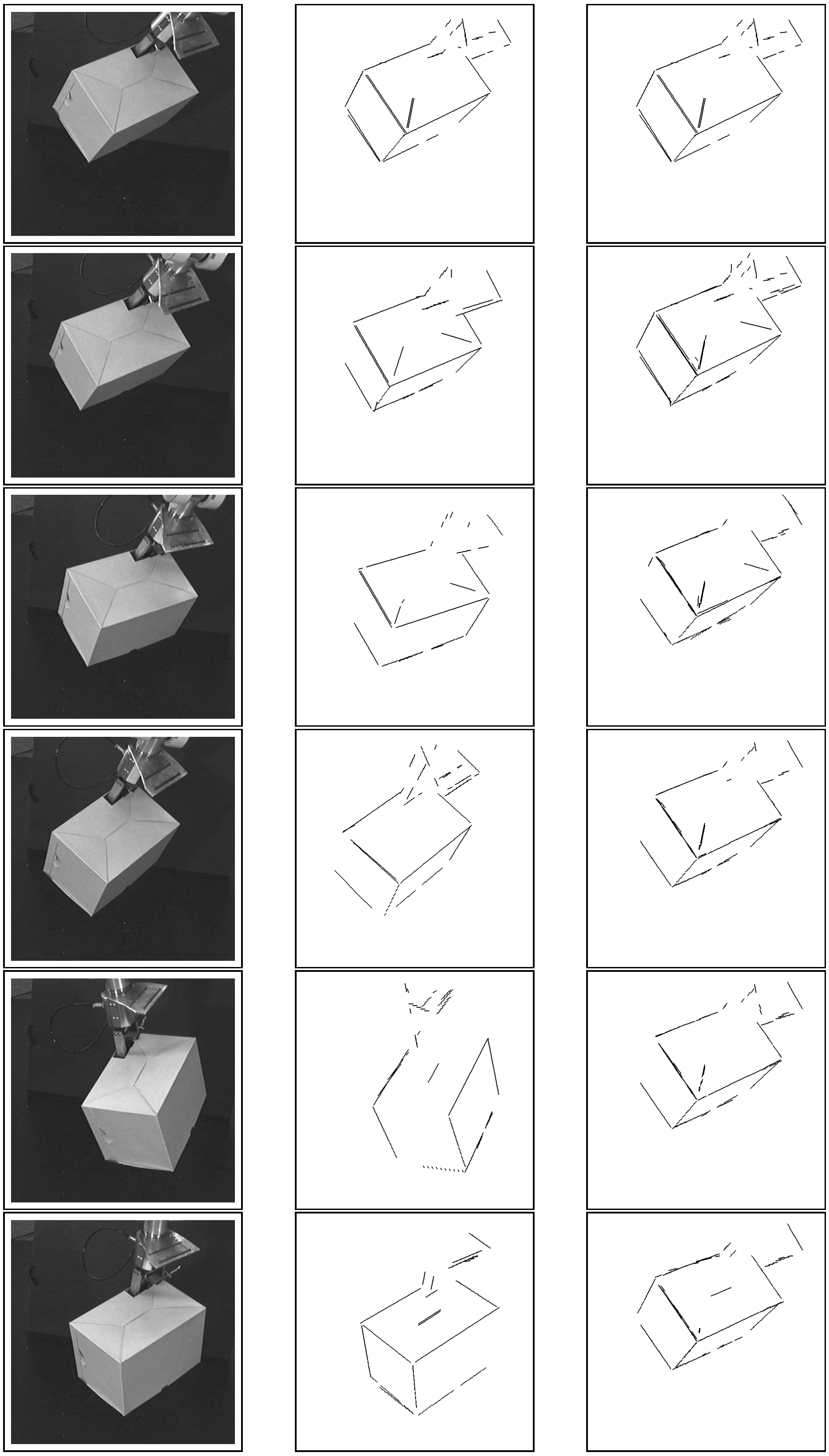


Abbildung 6.1: Bilderserie Karton, links Kamerabild, Mitte Bildrepräsentation, rechts Objektrepräsentation

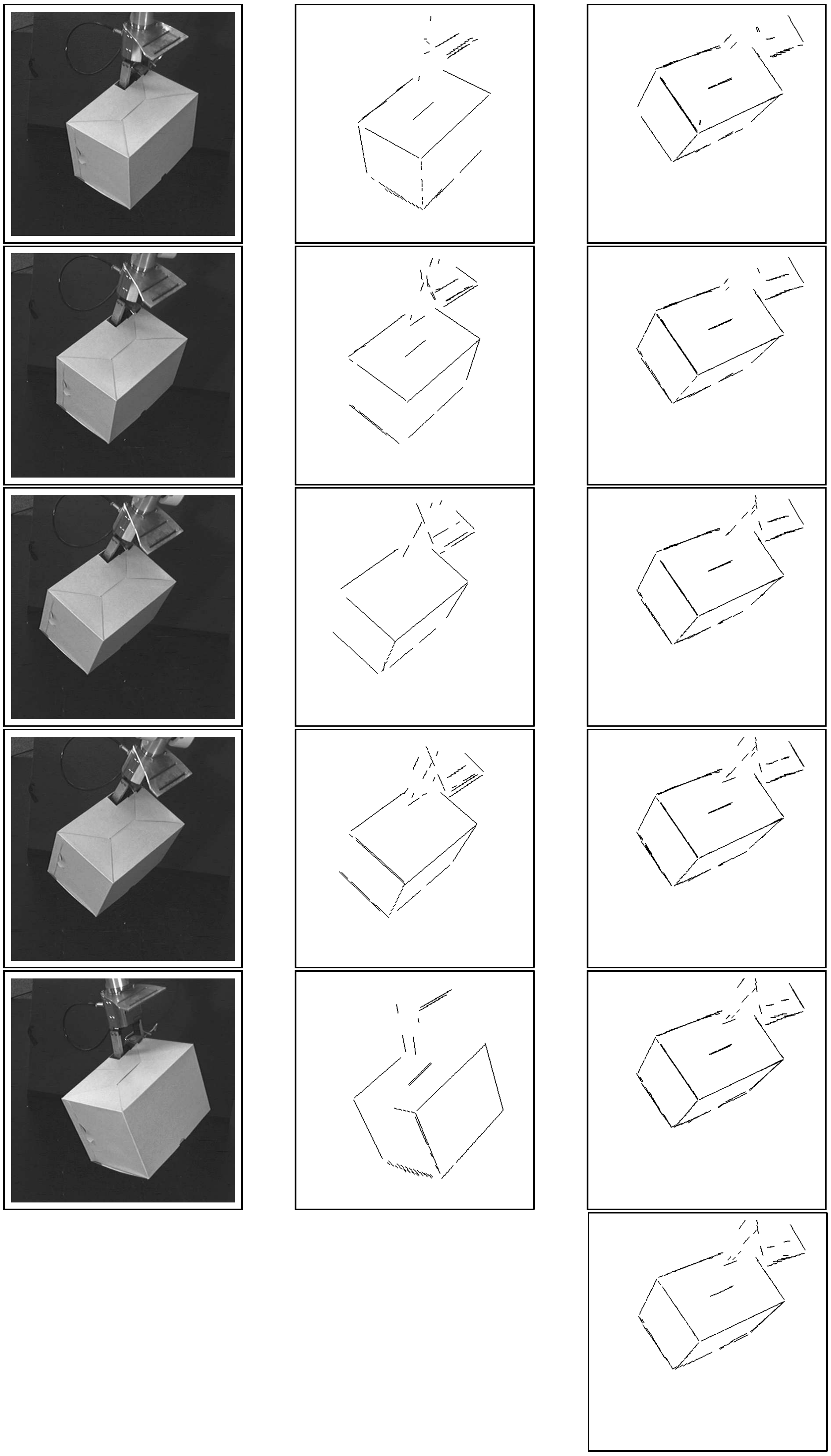


Abbildung 6.2: Bilderserie Karton, links Kamerabild, Mitte Bildrepräsentation, rechts Objektrepräsentation, letzte Zeile verschmolzene Objektrepräsentation

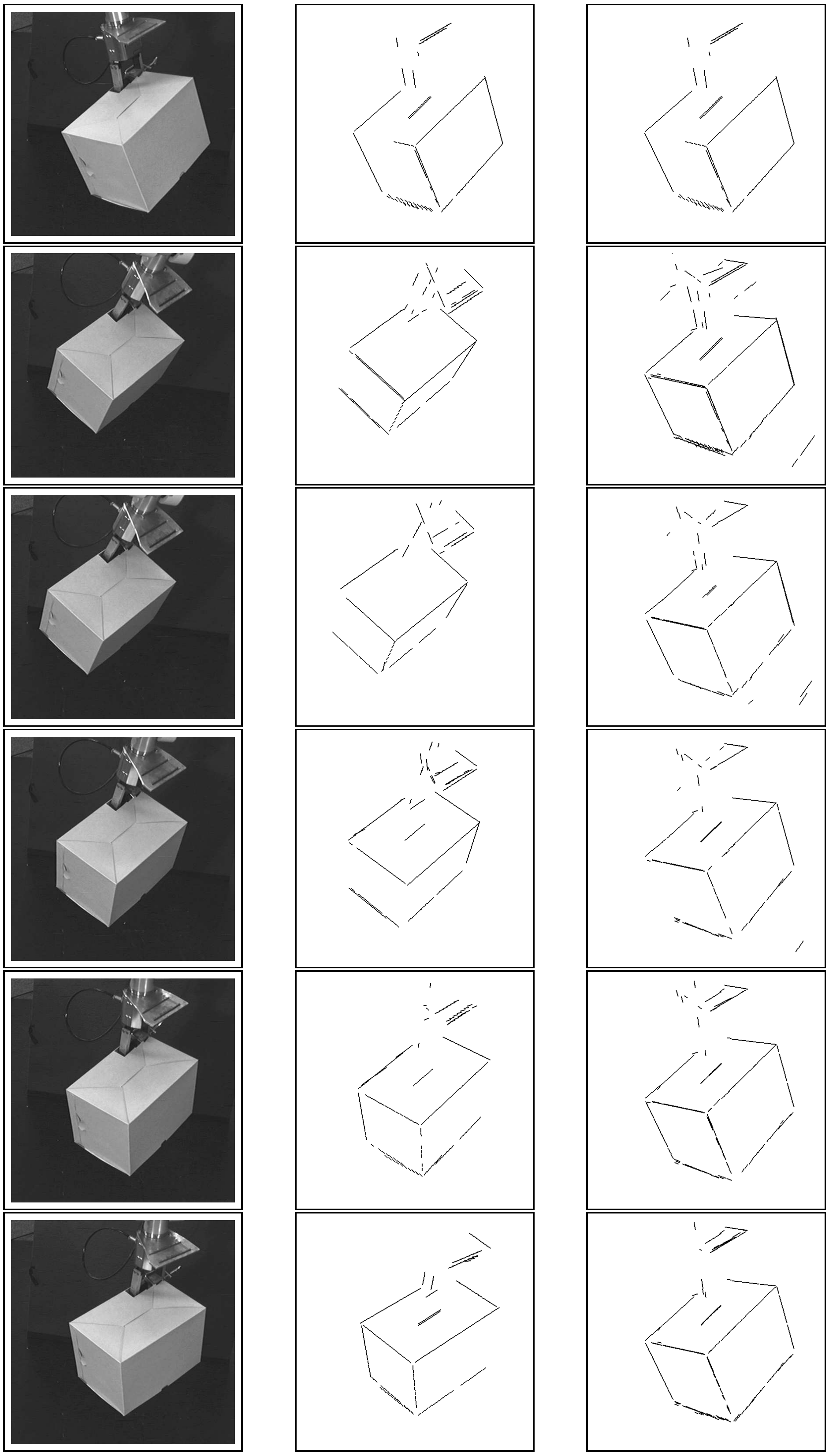


Abbildung 6.3: Bilderserie Karton in umgekehrter Reihenfolge, links Kamerabild, Mitte Bildrepräsentation, rechts Objektrepräsentation

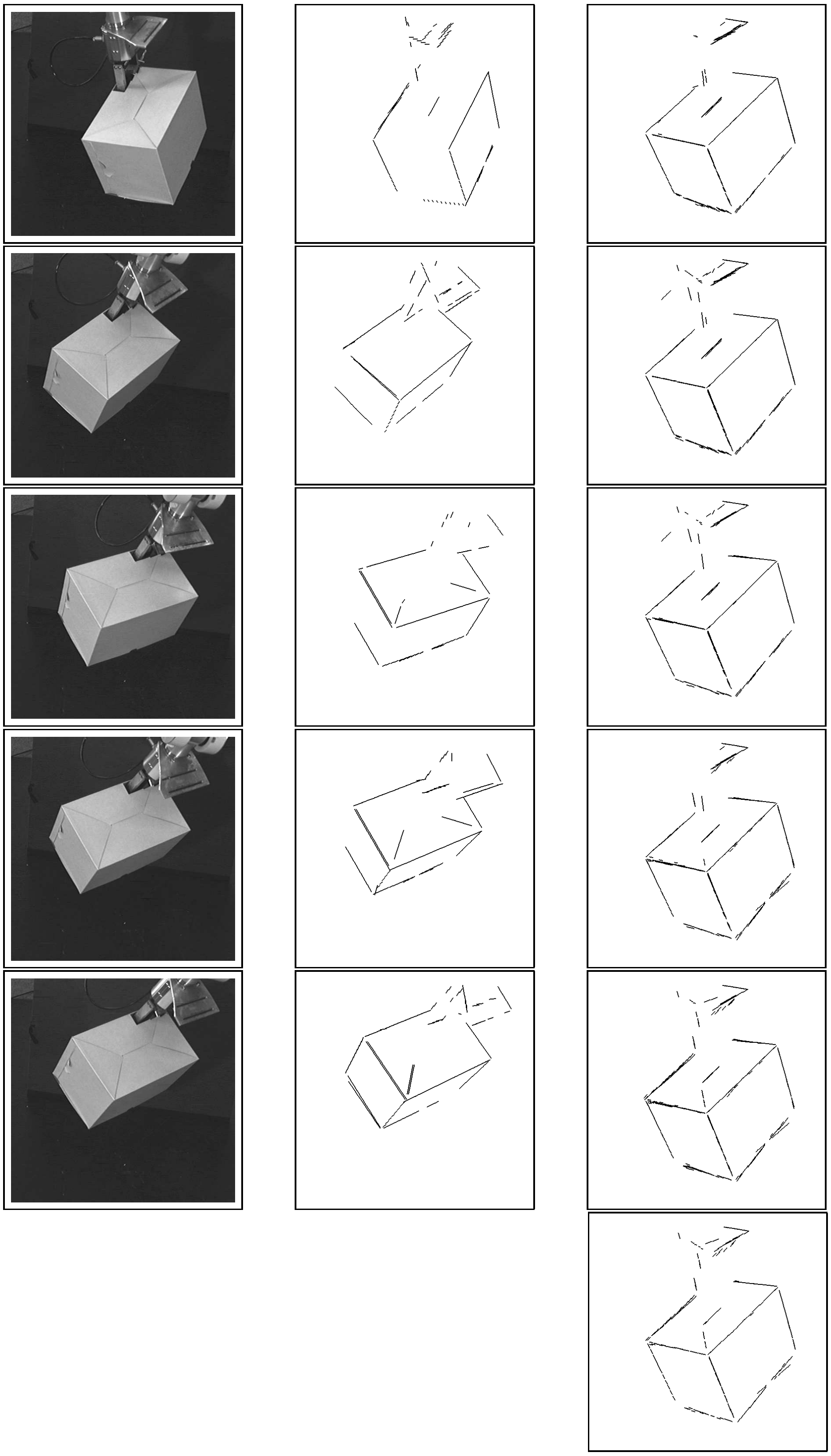


Abbildung 6.4: Bilderserie Karton in umgekehrter Reihenfolge, links Kamerabild, Mitte Bildrepräsentation, rechts Objektrepräsentation, letzte Zeile verschmolzene Objektrepräsentation

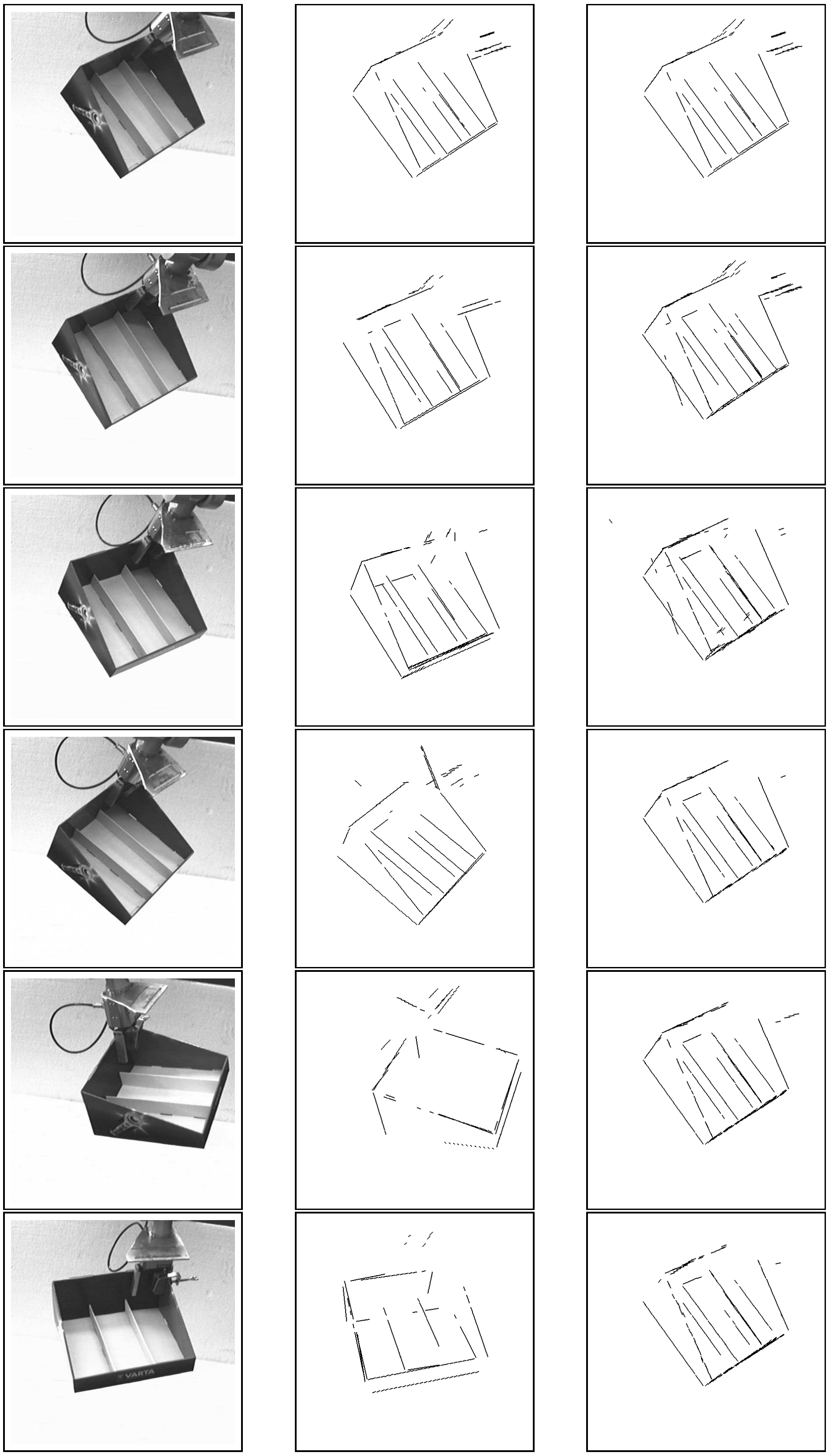


Abbildung 6.5: Bilderserie Batteriekiste, links Kamerabild, Mitte Bildrepräsentation, rechts Objektrepräsentation

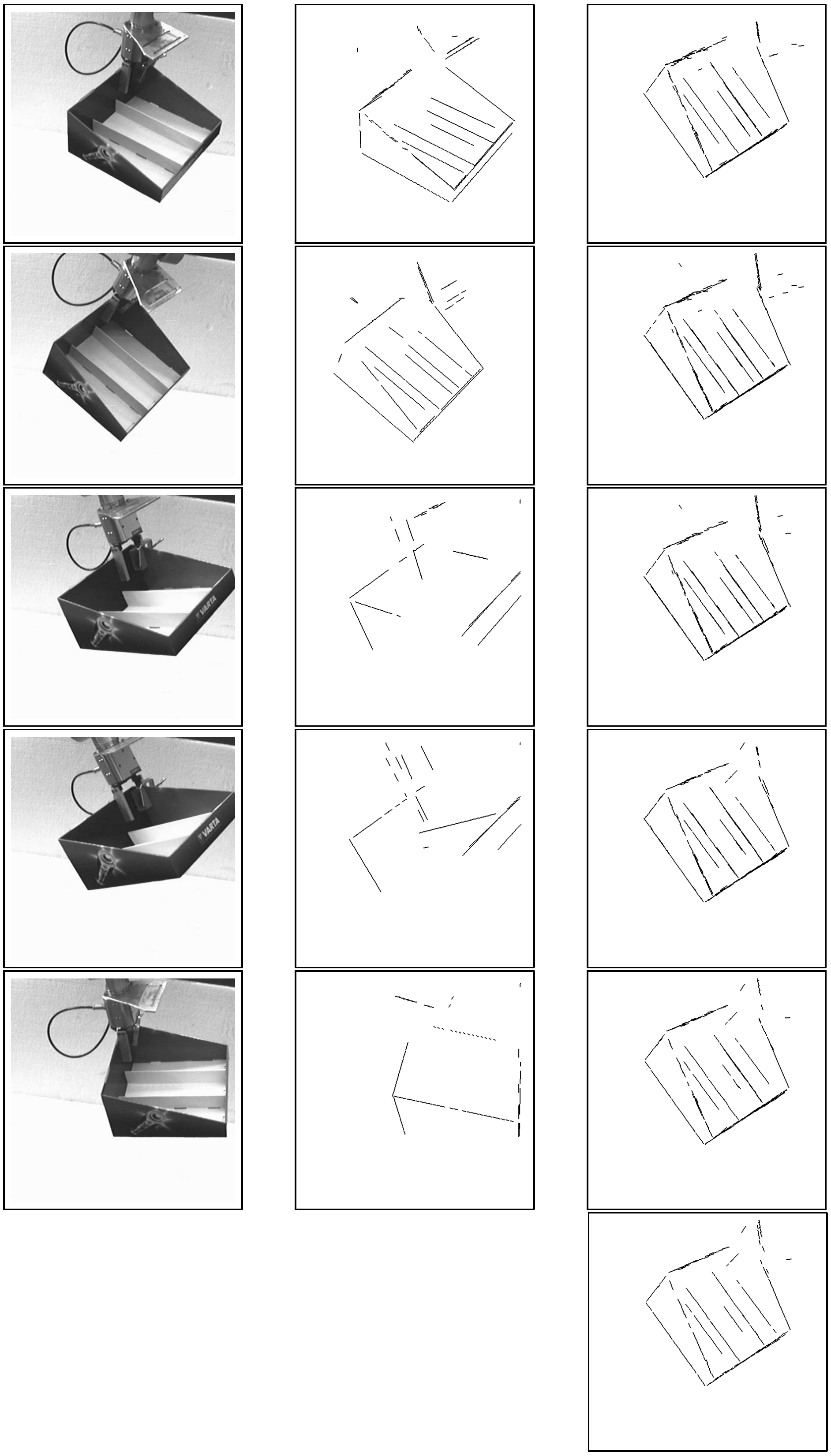


Abbildung 6.6: Bilderserie Batteriekiste, links Kamerabild, Mitte Bildrepräsentation, rechts Objektrepräsentation, letzte Zeile verschmolzene Objektrepräsentation

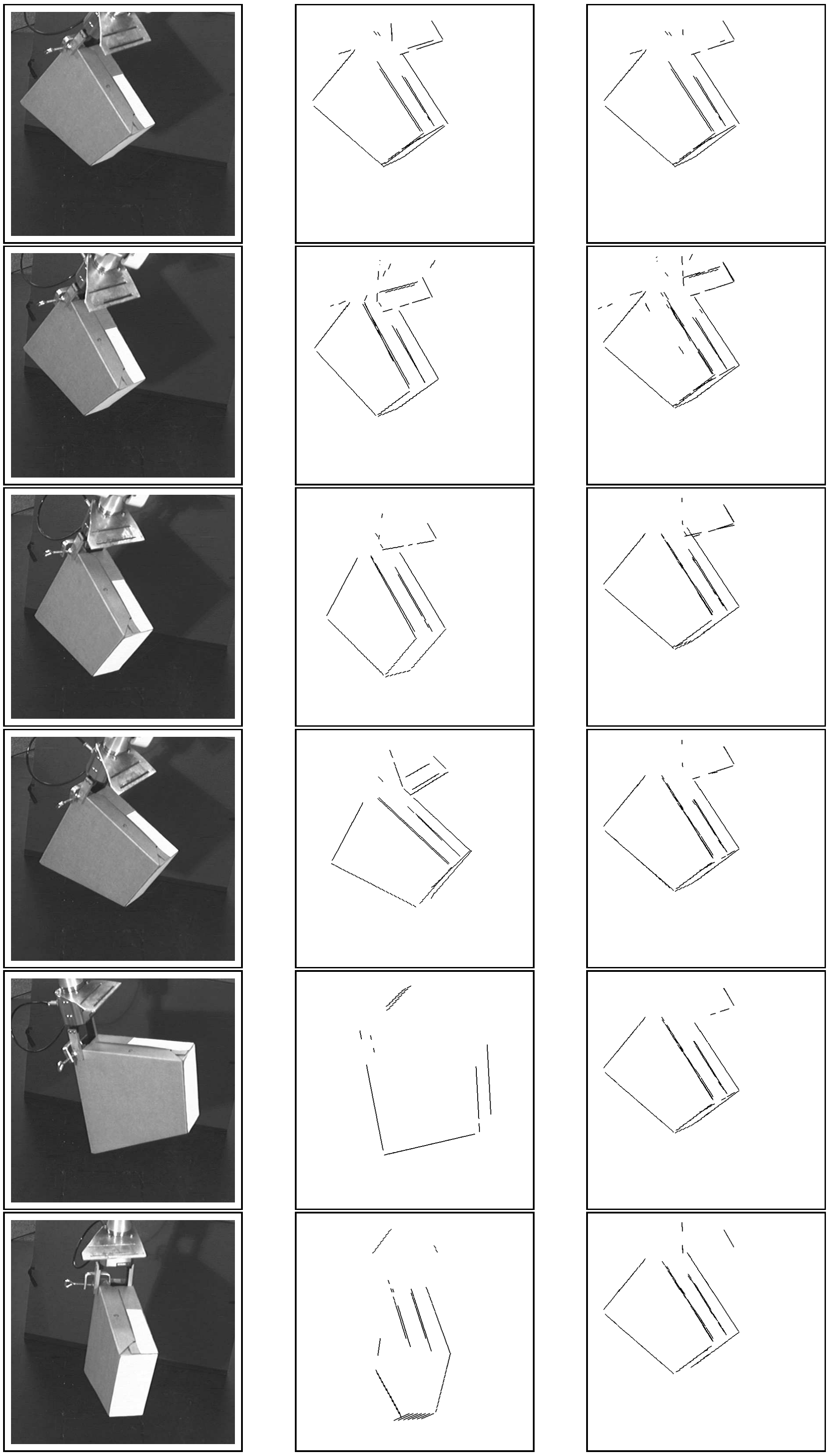


Abbildung 6.7: Bilderserie Ordner, links Kamerabild, Mitte Bildrepräsentation, rechts Objektrepräsentation

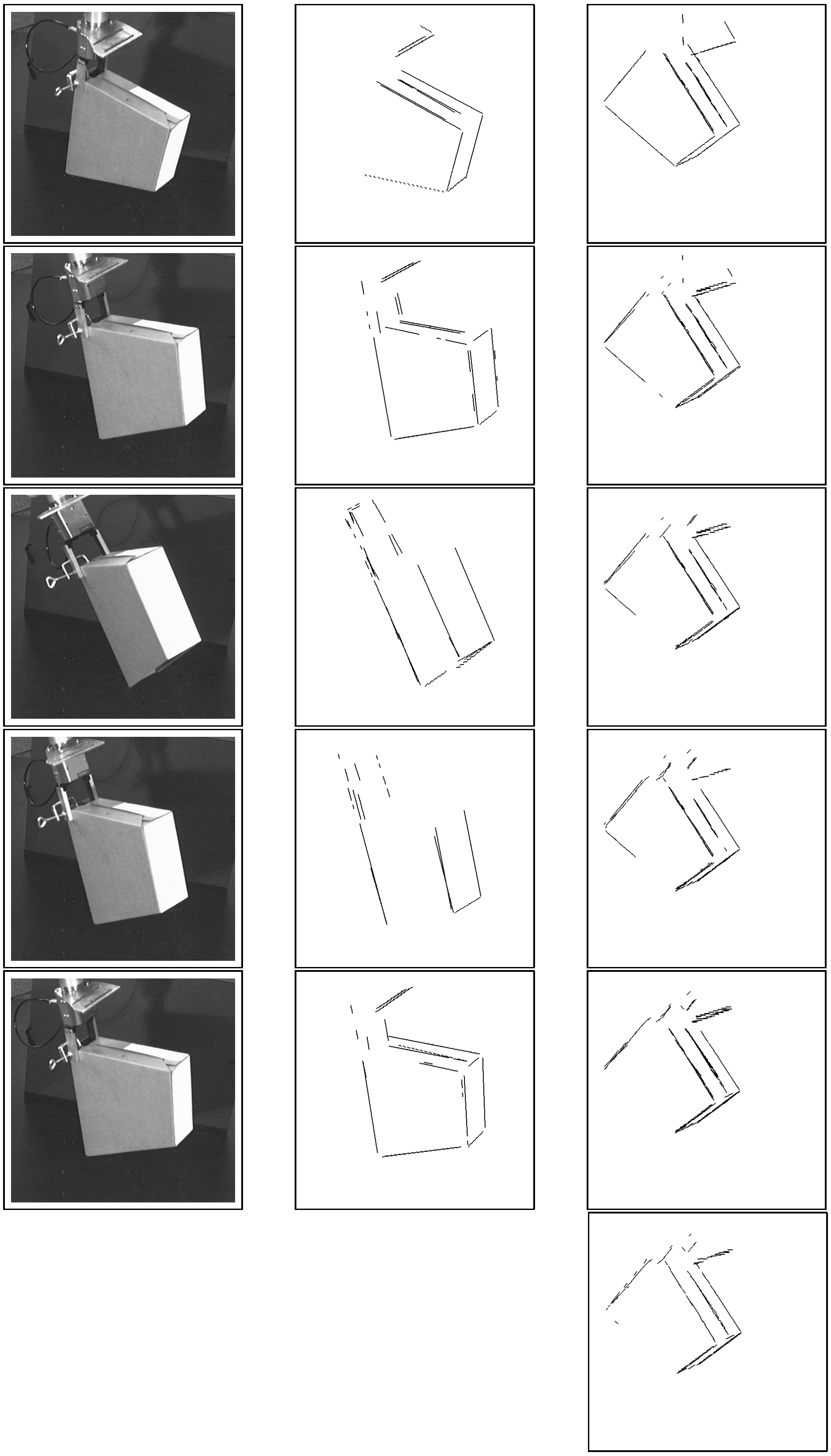


Abbildung 6.8: Bilderserie Ordner, links Kamerabild, Mitte Bildrepräsentation, rechts Objektrepräsentation, letzte Zeile verschmolzene Objektrepräsentation

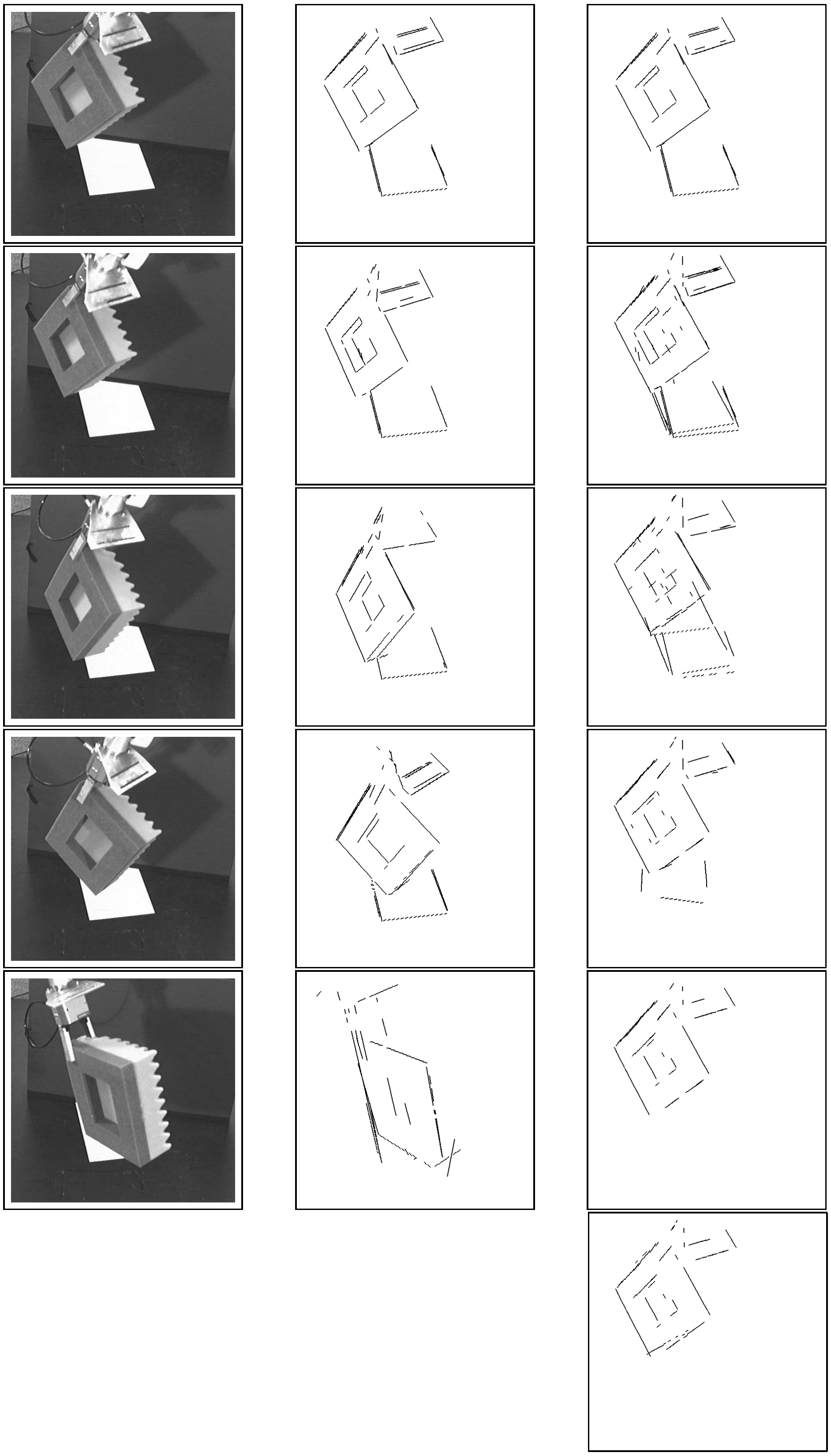


Abbildung 6.9: Bilderserie Schaumstoff, links Kamerabild, Mitte Bildrepräsentation, rechts Objektrepräsentation, letzte Zeile verschmolzene Objektrepräsentation

7 Zusammenfassung und Ausblick

In der vorliegenden Arbeit wurde untersucht, inwieweit die Unsicherheit der aus Bildern extrahierten Information eliminiert werden kann, indem der raum-zeitliche Zusammenhang der starren Körperbewegung genutzt wird. Das Objekt, das sich im Greifer eines Roboters befand, wurde bewegt und mit einem Stereokamerasystem in verschiedenen Ansichten aufgenommen. Alle aus jeder Ansicht extrahierten unsicheren Bildrepräsentationen wurden akkumuliert, was durch die Kenntnis der Roboterbewegung ermöglicht wurde.

Die Unsicherheit der in jeder Bildrepräsentation enthaltenen Information wurde durch eine Konfidenz ausgedrückt, die beschreibt, wie *sicher* das System diese Information extrahiert hat. Aber nicht nur die Konfidenz läßt eine Aussage über die Sicherheit der Information zu, sondern auch die Tatsache, wie oft die Information während der Akkumulation bestätigt wurde. Durch die Akkumulation wurde also die unsichere Information stabilisiert oder eliminiert. Insbesondere konnten Objektrepräsentationen in *schwierigen* Situationen extrahiert werden, in denen auch Hintergrund auftritt. Dieser Hintergrund konnte erfolgreich während des Akkumulationsprozesses eliminiert werden. Hier unterscheidet sich das vorgestellte Verfahren von vielen Objekterkennungsverfahren, die Objektrepräsentationen unter idealen Bedingungen extrahieren oder diese a priori vorgeben.

Ein weiterer Schwerpunkt dieser Arbeit war die Maximumsuche nach der Hough-Transformation. Mehrere der implementierten Verfahren hatten Probleme, in einigen Fällen signifikante von nicht signifikanten Maxima zu unterscheiden, da die Suche ausschließlich im Akkumulatorarray stattfand. Daher wurde ein eigenes Verfahren entworfen, das zusätzlich zum Akkumulatorarray das Ausgangsbild zu Hilfe nimmt und die Bereiche der Hough-Geraden vergleicht, die zur Entstehung des lokalen Maximums beigetragen haben, und so signifikante von nicht signifikanten Maxima unterscheidet.

Sowohl das Verfahren zur Maximumsuche als auch der Akkumulationsalgorithmus wurden erfolgreich implementiert und eingehend getestet.

Der Akkumulationsalgorithmus bietet diverse Ansätze zur Verbesserung und Erweiterung sowie vielfältige Anwendungsmöglichkeiten. Denkbare Erweiterungen sind zum Beispiel die Nutzung zusätzlicher Beschreibungsebenen, wie zum Beispiel Textur, Farbe oder Ecken. Hier sei insbesondere auf die auf derselben Basis gewonnenen Eckpunkthypothesen nach [Ha99] hingewiesen. Das Zusammenfassen der eine Kante beschreibenden Liniensegmente zu einer Kantenrepräsentation mit Hilfe dieser Eckpunkthypothesen bietet die Möglichkeit, Objektkanten durch nur noch ein Liniensegment von der Länge der Kante zu repräsentieren. Durch Summation der Konfidenzen aller dieser zusammengefaßten Liniensegmente kann dann noch mehr Sicherheit erreicht werden.

Bisher werden nur kleine Bewegungen verwendet, so daß das Objekt immer nur von einer Seite betrachtet wurde. Die Akkumulation könnte aber auch so erweitert werden, daß das Objekt von allen Seiten betrachtet wird, so daß eine vollständige Objektrepräsentation gewonnen wird. Eine denkbare Möglichkeit wäre es auch, Repräsentationen verschiedener Aspekte eines Objekts, wie sie in dieser Arbeit extrahiert wurden, in einem Aspektgraphen zu verbinden [KD79].

Wir haben in dieser Arbeit den Ausdruck „Akkumulieren von Objektrepräsentationen“ und nicht „Lernen von Objektrepräsentationen“ benutzt. Lernen bedeutet im allgemeinen die Veränderung eines Verhaltens aufgrund von Erfahrung[1]. Im bisherigen Zustand wird Verhalten noch nicht betrachtet. Um Verhalten mit der extrahierten Repräsentation zu bekommen, müssen Objekte gegriffen bzw. klassifiziert werden. Hierzu ist es notwendig, eine extrahierte Objektrepräsentation mit Objekten in einer realen Szene zu matchen, ein fundamentales und schwieriges Problem (siehe z.B. [JB97]). Einige interessante Ansätze zur Posenschätzung, die eine 3D-Repräsentation von Objekten basierend auf Liniensegmenten benutzen (die jedoch korrekte Korrespondenzen voraussetzen) werden am Lehrstuhl für Kognitive Systeme entwickelt [So00]. Die hier extrahierte Repräsentation zu einem effizienten Matchingverfahren in realen Situationen zu verwenden, ist eine interessante Aufgabe für zukünftige Arbeiten.

[1] „learning, the alteration of behaviour as a result of individual experience. When an organism can perceive and change its behaviour, it is said to learn.“ (aus [EB99])

Abbildungsverzeichnis

Variablen- und Parameterverzeichnis

Es ist schwierig, allen Variablen und Parametern Bezeichner zuzuordnen, die einer Systematik folgen. Es wurde jedoch versucht, die folgenden Regeln weitestgehend einzuhalten. Reelle Werte und Funktionen werden mit Kleinbuchstaben (d, r) bezeichnet, Bildpunkte und markante Funktionen mit fettgedruckten Kleinbuchstaben $(\mathbf{p}, \mathbf{d})$, Winkel und markante reelle Werte mit griechischen Buchstaben (θ, μ), Mengen, Bilder und mehrdimensionale Arrays mit Großbuchstaben (A, G), deren Elemente werden mit eckigen Klammern $(A[x, y])$ oder Indizes $(G_{\mathbf{p}})$ bestimmt. Bei Funktionen wird der Bezug ggfs. ebenfalls über den Index festgelegt (d_r, d_θ).

Die folgenden Variablen finden in dieser Arbeit an verschiedenen Stellen Verwendung und sind hier der Übersichtlichkeit halber alphabetisch sortiert und mit Kurzbeschreibungen zusammengefaßt, griechische Buchstaben folgen am Schluß der Tabelle. Die Parameter des Systems sind in der darauffolgenden Tabelle zusammengestellt.

A	Akkumulatorarray
B	Binärbild
$\bar{B}$	durchschnittliche Intensität im Bild
$\mathcal{B}$	Basiskoordinatensystem des Roboters
c	Konfidenz
c_k	Korrespondenzkonfidenz
d	Distanzfunktion, Metrik
e	Bilddiagonale
e	Epipolargerade
E	normalisierter Gradient
E^{glob}	globale Energie eines Bildes
$E_{\mathbf{p}}^{av}$	Durchschnittsenergie bezüglich eines Punktes
$E_{\mathbf{p}}^{loc}$	lokale Energie eines Punktes
f	generell verwendet für Funktionen
$\mathbf{f}, \mathbf{f}_t$	Hilfsfunktionen für die Distanzfunktion von Hough-Geraden
$\mathbf{f}_{\mathbf{p}}$	Parameterdarstellung der Geraden
f_θ	Winkelabweichungsfunktion
g	Gerade
G	Gradientenbetrag
$I(\theta, \Delta\theta)$	zu bearbeitendes Intervall für die orientierungsselektive Hough-Transformation
IR	Bildrepräsentation

k	normalisierte Kreuzkorrelation
$\mathcal{K}$	Koordinatensystem eines Liniensegments
l	Länge eines Segments
$\mathbf{M}$	Bewegung des Roboters
$\mathcal{M}$	Raum aller Bewegungen
n_r, n_θ	Ausdehnung des Parameterraums
OR	Objektrepräsentation
p	Wahrscheinlichkeit, daß ein Punkt ein Kantenpunkt einer bestimmten Geraden ist
$\mathbf{P}$	Roboterposition
$\mathbf{p}$	Bildpunkt
$\mathbf{q}$	Rotationsquaternion
$\mathbf{R}$	Rotationsoperator
r	Abstand einer Geraden zum Ursprung
$\mathbf{s}$	Schnittpunkt zweier Geraden
S, S'	Sobel- und modifizierter Sobel-Operator
$\mathbf{s}$	Liniensegment, orientierter Punkt
$\mathcal{S}$	Raum aller Liniensegmente
$\mathcal{T}$	Toolkoordinatensystem des Roboters
t	Abstand auf der Geraden zu einem vorgegebenen Punkt
t	Zeitpunkt der Bildaufnahme
$\mathbf{T}_{\mathcal{A}}^{\mathcal{B}}$	Transformationsmatrix von Koordinatensystem $\mathcal{A}$ nach $\mathcal{B}$
$\mathbf{T}$	Translationsoperator
$\mathbf{t}$	Translationsquaternion
u	Zahl der Aktualisierungsschritte eines Liniensegments
$\mathbf{v}$	Orientierungseinheitsvektor eines Liniensegments
$\mathcal{W}$	Weltkoordinatensystem
x, y	generell verwendet für Koordinaten eines Punktes in 2D
X, Y	Ausdehnung des Binärbildes
X	Translationskomponente x des Roboters
Y	Translationskomponente y des Roboters
Z	Translationskomponente z des Roboters
α	Rotationswinkel yaw des Roboters
β	Rotationswinkel pitch des Roboters
γ	Rotationswinkel roll des Roboters
χ	Schnittwinkel zweier Geraden
θ	Gradientenwinkel
Θ	Wahrheitsfunktion, ob die Bedingungen für die Distanz von Hough-Geraden zutreffen
σ	Standardabweichung
ϕ, ψ	Winkel sphärischer Koordinaten

Die folgende Tabelle faßt die Parameter des Systems zusammen. Ferner findet sich zu jedem Parameter noch die Nummer des Kapitels, in dem der Parameter beschrieben ist und ein Standardwert.

Bez.	Kurzbeschreibung	Kapitel	Wert
s_r	Streckfaktor des Akkumulatorarray in r-Richtung	3.2	1
s_θ	Streckfaktor des Akkumulatorarray in θ-Richtung	3.2	1
η	Schwellwert für Binarisierung des Bildes	3.3	8
$\Delta\theta$	Winkelfehler für Hough-Transformation	3.4	10
μ	unterer Wert für Kontrastnormalisierung	3.5.1	1.0
ν	oberer Wert für Kontrastnormalisierung	3.5.1	2.3
n_E	Größe der Binomialmaske für Kontrastnormalisierung	3.5.1	5
ϱ	Gewichtungsfaktor für lokale und globale Energie	3.5.1	0.25
$\mathbf{d}$	Distanzfunktion von Hough-Geraden	3.6.4	40
σ	Schwellwert für Hough-Transformation	3.7	20
$\hat{l}$	Länge eines 2D-Liniensegments	4.1.1	11
$c_{\min}$	Mindestkonfidenz für betrachtete Liniensegmente	4.2	0.5
h	Höhe des Matching-Fensters	4.2	15
w	Breite des Matching-Fensters	4.2	15
ζ	max. Abweichung der Korrespondenzkonfidenz	4.2	1.0
ξ	Schwellwert für die Orientierungsdifferenz	5.2.1	10
ω	Schwellwert für die Positionsdifferenz	5.2.1	10

Literaturverzeichnis

[AF87] N. Ayache, B. Faverjon: *Efficient registration of stereo images by matching graph descriptions of edge segments*, The International Journal of Computer Vision, Vol. 1, No. 2, pp. 107–132, 1987

[Ba81] D.H. Ballard: *Generalizing the Hough transform to detect arbitraty shapes*, Pattern Recognition, Vol. 13, No. 2, pp. 111–122, 1981

[BB82] D.H. Ballard, C.M. Brown: *Computer Vision*, Prentice-Hall, Englewood Cliffs, 1982

[Bl98] W. Blase: *RBF-basierte neuronale Netze zum automatischen Erwerb hindernisfreier Manipulator-Trajektorien, basierend auf monokularen Bildfolgen*, Diplomarbeit am Lehrstuhl für Kognitive Systeme der Christian-Albrechts-Universität zu Kiel, 1998

[Br91] R.A. Brooks: *Intelligence without Reason* in Proceedings of the 1991 International Joint Conference on Artificial Intelligence, pp. 569–595, 1991

[Ca86] J.F. Canny: *A computational approach to edge detection*, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 8, No. 6, pp. 679–698, 1986

[Cr89] J.J. Craig: *Introduction to Robotics – Mechanics and Control*, Addison-Wesley, Reading, 1989

[DH72] R.O. Duda, P.E. Hart: *Use of the Hough transformation to detect lines and curves in pictures*, Communications of the ACM, Vol. 15, No. 1, pp. 11–15, 1972

[ES98] W. Effelsberg, R. Steinmetz: *Video Compression Techniques*, Dpunkt, 1998

[El38] W. Ellis, Hrsg.: *Gestalt Theory, A Source Book for Gestalt Psychology*, Harcourt, Brace & World, 1938

[EB99] Encyclopædia Britannica, 1999

[Fa93] O. Faugeras: *Three-Dimensional Computer Vision – a Geometric Viewpoint*, MIT-Press, Cambridge, 1993

[Gi79] J. Gibson, *The Ecological Approach to Visual Perception*, Boston, Houghton Mifflin, 1979

[GW92] R.C. Gonzalez, R.E. Woods: *Digital Image Processing*, Addison-Wesley, Reading, 1992

[Ha99] M. Hahn: *Semiglobale Verfahren zur Generierung von Eckpunkthypothesen in 2D und 3D*, Diplomarbeit am Lehrstuhl für Kognitive Systeme der Christian-Albrechts-Universität zu Kiel, 1999

[HH89] C. Hansen, T. Henderson: *CAD-based computer vision*, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 11, No. 11, pp. 1181–1193, 1989

[HPB84] R. von der Heydt, E. Peterhans, G. Baumgartner: *Illusory contours and cortical neuron responses*, Science, Vol. 224, pp. 1260–1262, 1984

[Ho62] P.V.C. Hough: *Method and means for recognizing complex patterns*, U.S. Patent 3,069,654, Dec. 18, 1962

[IK88] J. Illingworth, J. Kittler: *A Survey of the Hough transform*, Computer Vision, Graphics and Image Processing (CVGIP), Vol. 44, pp. 87–116, 1988

[Ja89] A.K. Jain: *Fundamentals of Digital Image Processing*, Prentice-Hall, Englewood Cliffs, 1989

[Ja93] B. Jähne: *Digitale Bildverarbeitung*, Springer-Verlag, Berlin, 1993

[Ja95] R. Jain et al.: *Machine Vision*, McGraw-Hill, New York, 1995

[JB97] X. Jiang, H. Bunke: *Dreidimensionales Computersehen – Gewinnung und Analyse von Tiefenbildern*, Springer-Verlag, Berlin, 1997

[KA98] P.J. Kellman, M.E. Artenberry: *The Cradle of Knowledge: Development of Perception in Infancy*, MIT-Press, Cambridge, 1998

[Ko95] R. Koch: *3-D surface reconstruction from stereoscopic image sequences*, International Conference of Computer Vision ICCV '95, Cambridge, MA., USA, 1995

[Ko92] J.J. Koenderink: *Wechsler's vision: An essay review of computational vision by Harry Wechsler*, Ecological Psychology, No. 4, pp. 121–128, 1992

[KD79] J.J. Koenderink, A.J. van Doorn: *The internal representation of solid shape with respect to vision*, Biological Cybernetics, Vol. 32, pp. 211–216

[Kr98a] N. Krüger: *Visual Learning with a priori Constraints*, Dissertation an der technischen Fakultät der Universität Bielefeld, 1998

[Kr98b] N. Krüger: *Collinearity and parallelism are statistically significant second order relations of complex cell responses*, Neural Processing Letters, Vol. 8, pp. 117–129

[Ku99] S. Kunze: *Ein Hand-Auge-System zur visuell basierten Lokalisierung und Identifikation von Objekten*, Diplomarbeit am Lehrstuhl für Kognitive Systeme der Christian-Albrechts-Universität zu Kiel, 1999

[KZ95] R. Klette, P. Zamperoni: *Handbuch der Operatoren für die Bildverarbeitung*, Vieweg, Braunschweig, 1995

[La92] M. Lades et al.: *Distortion invariant object recognition in the dynamic link architecture*, IEEE Transactions on Computers, Vol. 42, No. 3, pp. 300–311, 1992

[LTC97] A. Lanitis, C.J. Taylor, T.F. Cootes: *Automatic interpretation and coding of face images using flexible models*, IEEE Transactions on Pattern Recognition and Machine Intelligence, Vol. 19, No. 7, pp. 743–756, 1997

[LB87] V.F. Leavers, J.F. Boyce: *The radon transform and its application to shape parameterization in Machine Vision*, Image Vision Computing, Vol. 5, No. 2, pp. 161–166, 1987

[Le93] V.F. Leavers: *Which Hough transform?*, CVGIP: Image Understanding, Vol. 58, No. 2, pp. 250–264, 1993

[Ma98] H.A. Mallot: Sehen und die Verarbeitung visueller Informationen, Vieweg, Braunschweig, 1998

[OS84] L. o'Gorman, A.C. Sanderson: *The converging squares algorithm: an efficient method for locating peaks in multidimensions*, IEEE Transactions on Pattern Analysis and Machine Intelligens, Vol. 6, No. 3, pp. 280–288, 1984

[Pa98] J. Pauli: *Geometric/Photometric Consensus and Regular Shape Quasi-Invariants for Object Localization and Boundary Extraction*, Technischer Bericht Nr. 9805 des Instituts für Informatik und Praktische Mathematik der Christian-Albrechts-Universität zu Kiel, 1998

[Pr90] Princen et al.: *An optimizing line finder using a Hough transform algorithm*, Computer Vision, Graphics, and Image Processing, Vol. 52, pp. 57–77, 1990

[Ri97] E.M. Riseman et al.: *Landmark-based navigation and the acquisition of environmental models* in Y. Aloimonos (Hrsg.): *Visual Navigation – From Biological Systems to Unmanned Ground Vehicles*, Lawrence Erlbaum Associates, Mahwah, 1997

[SB97] S.M. Smith, J.M. Brady: *SUSAN – a new approach to low level image processing*, Int. Journal of Computer Vision, Vol. 23, No. 5, pp. 45–78, 1997

[Sch96] H. Scharr: *Digitale Bildverarbeitung und Papier - Texturanalyse mittels Pyramiden und Grauwertstatistiken am Beispiel der Papierformation*, Diplomarbeit an der Fakultät für Physik und Astronomie der Ruprecht-Karls-Universität Heidelberg, 1996

[Sch98] A. Schmidt: *Manipulatorregelung durch fortwährende visuelle Rückkopplung*, Diplomarbeit am Lehrstuhl für Kognitive Systeme der Christian-Albrechts-Universität zu Kiel, 1998

[Se90] H.-P. Seidel: *Quaternionen in Computergraphik und Robotik*, Informationstechnik, Vol. 32, No. 4, pp. 266–275, 1990

[Sh96] F. Shevlin: *Geometric Rectification of Scanned Imagery*, Ph.D.-thesis, Department of Computer Science, University of Dublin, Trinity College, 1996

[So97] G. Sommer: *Algebraic aspects of designing behaviour based systems*, in G. Sommer, J.J. Koenderink (Hrsg.): *Algebraic Frames for the Perception and Action Cycle*, pp. 1–28, Springer, 1997

[So99] G. Sommer: *The global algebraic frame of the perception-action-cycle*, in B. Jähne et al. (Hrsg.): *Handbook of Computer Vision and Applications*, Vol. 3, pp. 221–264, Academic Press, 1999

[So00] G. Sommer, B. Rosenhahn, Y. Zang: *Pose estimation using geometric algebra*, Dagstuhl-Seminar, Multi-Image Search, Filtering, Reasoning, and Visualisation, zu erscheinen 2000

[St94] Bedienungsanleitung V+ zum Steuerschrank CS7 des Roboters Modell RX90 der Firma Stäubli AG, 1994

[Vi92] V.V. Vinod et al.: *A connectionist approach for peak detection in Hough space*, Pattern Recognition, Vol. 25, No. 10, pp. 1253–1264, 1992

[Wi97] L. Wiskott et al.: *Face recognition by elastic bunch graph matching*, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 19, pp. 775–780, 1997

[YU90] A. Yuille, S. Ullman: *Rigidity and smoothness of motion: justifying the smoothness assumption in motion measurement*, in S. Ullman, W. Richards (Hrsg.): *Image Understanding 1989*, pp. 163–184, Ablex Publishing, Norwood, New Jersey, 1990

[Yu91] A. Yuille: *Deformable templates for face recognition*, Journal of Cognitive Neuroscience, Vol. 3, No. 1, pp. 59–70, 1991

[ZK95] R. Zavodnik, H. Kopp: *Graphische Datenverarbeitung, Grundzüge und Anwendungen*, Hanser-Verlag, München, 1995

Danksagungen

Zu dieser Arbeit haben viele Menschen einen Beitrag geleistet, allen gilt mein Dank. Ganz besonders danke ich meinen Eltern, die mich finanziell unterstützt und mir alle Freiheiten gelassen haben. Ich danke Norbert Krüger für eine erstklassige Betreuung, viele fruchtbare Diskussionen und die Idee des KiViGraP, wovon auch die nachfolgenden Studenten profitieren werden. Für eine gute Zusammenarbeit in diesem Projekt danke ich Marco Hahn und zu guter Letzt Alexander Happe, Doris Jentzsch und Birger Streckel für die zahlreichen Tips und für das Korrekturlesen.

Erklärung

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

Kiel, den 12. Januar 2000