KIFS 94

Sensordatenfusion für Anwendungen in der Robotik

A. Knoll, Universität Bielefeld

Inhaltsverzeichnis

1 Motivation

Alle denkbaren Verfahren zur Aufnahme von Umweltdaten weisen spezifische Stärken und Schwächen auf. Es gibt bislang keine universelle Lösung in Form eines Einzelsensors, die für alle denkbaren Anwendungsfälle optimale oder zumindest brauchbare Entfernungsbilder liefert, und es wird wohl auch in Zukunft keine solche Lösung geben. Man ist normalerweise bestrebt, Sensorsysteme zu konzipieren, die unter möglichst allgemeinen Bedingungen brauchbar sind. Es liegt daher nahe, mehr als einen Sensor zu verwenden und die von diesen Sensoren gelieferte, mehr oder weniger redundante Information sinnvoll zu kombinieren. Die dabei verfolgten Ziele sind

- die spezifischen Mängel eines Sensors bei Erfassung eines Umweltmerkmals durch einen anderen (anders aufgebauten) zu kompensieren,
- durch mehrfach durchgeführte Messungen eines Umweltmerkmals mit demselben oder verschiedenen Sensoren die Meßergebnisse von Rauschen zu befreien,
- durch Zusammenwirken mehrerer Sensoren an möglicherweise verschiedenen Standorten ein vollständigeres und glaubwürdigeres Umweltbild zu erhalten, als dies mit einem einzelnen Sensor möglich wäre.

Es können sowohl homogene, d.h. aus gleichartigen Sensoren bestehende Systeme aufgebaut werden (beipielsweise eine Vielzahl baugleicher Kameras an verschiedenen Standorten [Yoshida 92]), wie auch inhomogene Systeme, welche gleichartige Information (etwa über die Entfernung) mit unterschiedlichen Meßprinzipien gewinnen, also beispielsweise ein Ultraschallsensor und eine Kamera (siehe z.B. [Harmon 86; Scholz 88; Beckerman 92]) oder eine Kamera und einen Laserscanner [Bennekers 86]. Solche aus mehreren Einzelsensoren bestehenden Sensorverbunde haben darüber hinaus im allgemeinen den Vorteil erhöhter Fehlertoleranz und Störsicherheit, denn es ist natürlich immer möglich, daß ein Einzelsensor vollständig ausfällt oder vorübergehend falsche Meßdaten liefert. Zusätzlich kann eine größere, statistisch signifikantere Datenmenge gewonnen werden und die Konsistenz der Einzelergebnisse durch gegenseitige Einschränkungen und Überprüfungen verbessert werden.

Die zentrale Frage bei einem solchen Vorgehen ist natürlich die der Art und Weise der Fusion der von den Einzelsensoren gelieferten Daten. Es sollte gewährleistet werden, daß das Fusionsergebnis besser als die Information eines jeden Einzelsensors ist, also genauer, zuverlässiger und reichhaltiger. In der Literatur existieren verschiedene Ansätze für den Entwurf solcher multisensorieller Systeme. In den meisten Fällen handelt es sich dabei um die Verknüpfung spezieller Sensoren für spezielle Einsatzfälle[1], beispielsweise ein Stereokamerasystem und ein taktiler Sensor [Allen 85] oder eine Kamera und ein Infrarot–Temperatursensor [Aggarwal 90]. Die meisten der hierbei für die Fusion verwendeten Verfahren sind ad hoc auf den vorliegenden Einsatzfall zugeschnitten und daher nicht zu verallgemeinern. Im folgenden soll im Gegensatz dazu der Versuch gemacht werden, eine gewisse

[1] Nach dem Wissensstand des Autors gibt es bislang keine diesem Themengebiet gewidmete *umfassende* und allgemein orientierte Darstellung. Es existieren lediglich einige Aufsatzsammlungen [Brady 88; Tou 90] und eine formale Methoden weitgehend meidende Zusammenstellung von Methoden hauptsächlich aus der künstlichen Intelligenz [Raczkowsky 91], sowie die recht speziellen Bücher [Allen 87], [Durrant–Whyte 87b] und [Hager 90]. Außerdem sei das Buch von Clark und Yuille [Clark 90] erwähnt, das sich um eine systematische Herangehensweise bemüht, aber hauptsächlich auf Datenfusion in Sichtsystemen ausgerichtet ist.

Systematik in dieses sehr junge bislang weitgehend ungeordnete Themengebiet zu bringen[1]. Es werden dazu solche Ansätze vorgestellt, die mathematisch fundiert sind und einen Anspruch auf allgemeinere Verwendbarkeit stellen können. Es handelt sich dabei um Verfahren, die auf parametrierten Sensor- und Weltmodellen basieren und statistische Methoden benutzen. Zum einen werden durch statistische Verfahren die unvermeidbaren Ungenauigkeiten der Sensordaten mit in die Berechnung einbezogen und ausgeglichen. Zum anderen können sie prinzipiell sowohl für den Bereich statischer als auch bewegter Systeme eingesetzt werden. Es wird bei der Darstellung nur auf die jeweiligen speziellen Aspekte eingegangen. Die erforderlichen statistischen Grundlagen werden der besseren Übersicht wegen nicht hier, sondern in B dargestellt. Auf Methoden zur Gewinnung redundanter Information durch entsprechende Steuerung von Sensorverbunden wird in Kapitel 6 eingegangen.

Wir betrachten zunächst die drei wesentlichen Aspekte eines Multisensorsystems, das sich (natürlich nicht ausschließlich) für entfernungsmessende Systeme eignet: das *Sensormodell*, das *Weltmodell* und die *Fusionsmethode*. Da Sensor- und Weltmodelle bereits bei Einsensorsystemen Einsatz finden, werden diese nur gestreift. Wesentlich interessanter sind die verschiedenen statistischen Verfahren zur eigentlichen Fusion. Wir gehen auf verschiedene Varianten ein und bewerten diese nach ihrer Fähigkeit, Sensordaten so zusammenzufassen, daß das Ergebnis den Einzelmessungen überlegen ist. Bei Sensordatenintegration sind die folgenden Aspekte von Interesse, die im Verlaufe des Kapitels angesprochen werden:

- *Sensormodell* zur Beschreibung des Sensorverhaltens. Das Sensormodell ist Voraussetzung zur Konsistenzüberprüfung der Daten (Sicherstellung, daß von allen Sensoren das gleiche Objektmerkmal beobachtet wird),
- Allgemeines *Weltmodell* zur Beschreibung der erfaßten Umwelt,
- *Fusionsmethode*,
- Mathematischer und algorithmischer *Aufwand*.

Zur Abschätzung verschiedener Parameter von Fusionsmethoden wurde eine Menge von synthetischen Sensordaten zufällig erzeugt und als Eingabe für die Auswertung benutzt[2] (siehe unten). Wir beschränken uns im folgenden auf Verfahren, die zur Verarbeitung von Daten von Entfernungssensoren verwendet werden oder zumindest prinzipiell in diesem Zusammenhang eingesetzt werden können.

1.1 Einführende Bemerkungen

Ziel der multisensoriellen Datenerfassung ist es, genauere Information über die Umwelt durch verschiedene Sensoren zu gewinnen, als dies mit einem Einzelsensor möglich wäre. Die einfachste Möglichkeit besteht bei einem System aus mehreren Sensoren darin, immer die Daten des Sensors zu wählen, der am „geeignetsten“ erscheint. Bei diesem Auswahlprozeß sind detaillierte Kenntnisse über die Eigenschaften der Sensoren erforderlich. Es muß mit Hilfe möglicherweise bruchstückhafter Information entschieden werden, welcher Sensor in einer gegebenen Konstellation das beste Ergebnis

1 In [Hackett 90] sprechen die Autoren von einem „[…] state of infancy“.

2 Es sei an dieser Stelle bemerkt, daß bei den bislang publizierten Arbeiten regelmäßig keine oder nur rudimentäre experimentelle Ergebnisse angegeben wurden, die keinen Vergleich ermöglichen. Dies hängt sicherlich damit zusammen, daß sich dieses Gebiet noch in den Anfängen befindet. Die Simulation mit identischen Eingangsdaten ist daher die objektivste Vergleichsmöglichkeit.

zu liefern verspricht. Der Hauptnachteil dieser Methode besteht darin, daß die vorhandenen redundanten Informationen nicht genutzt werden, um die Genauigkeit des Meßergebnisses zu erhöhen.

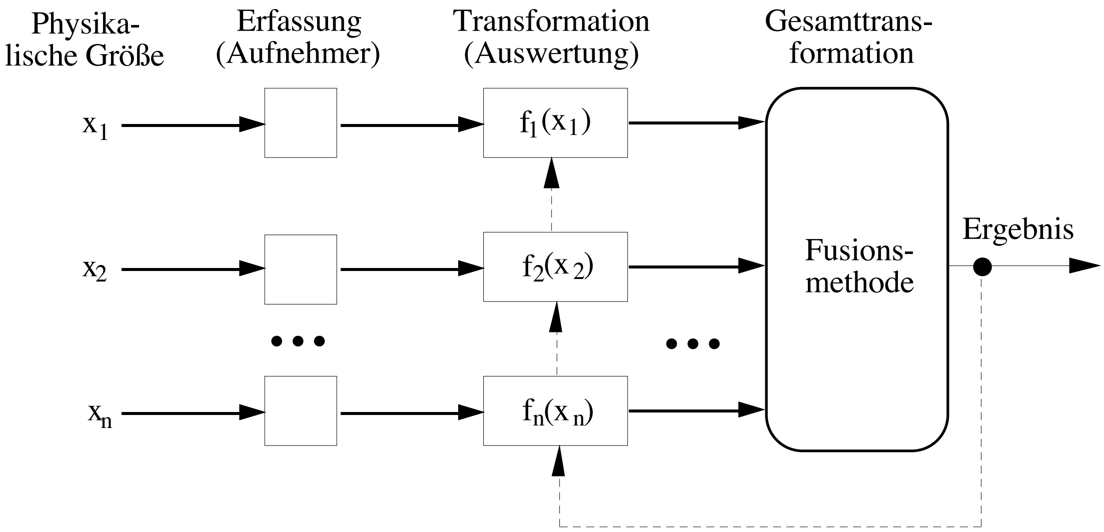

Bild 1: Allgemeines Modell zur Sensordatenfusion

In einem Multisensorsystem sollten deshalb zur Verbesserung der Umweltinformationen die Sensordaten durch eine sogenannte Sensordaten*fusion* geeignet kombiniert werden. Eine solche Fusion findet üblicherweise in zwei Schritten statt:

1. In einem vorbereitenden Schritt müssen die Sensordaten, die von möglicherweise unterschiedlichen, im allgemeinen physikalisch verteilten Sensoren stammen, in eine einheitliche Repräsentation transformiert werden. Diese Transformation basiert auf einer geeigneten Modellierung der Umwelt und der Sensoren (Welt- und Sensormodelle).
2. Im eigentlichen Fusionsschritt muß ein Fusionsergebnis für diese einheitliche Datenrepräsentation gefunden werden.

Bild 1 zeigt eine schematische Darstellung dieses Vorgehens[1]. In dieser Abbildung werden die beobachteten physikalischen Größen mit x_1, x_2, ..., x_n bezeichnet. Zu jedem Sensor gehört eine charakteristische Transformation, die die Beobachtung x_i in Zusammenhang bringt mit dem wahren Gesamtumweltzustand θ (oder einem Ausschnitt davon) aus der – möglicherweise fehlerhaften – Sicht des Einzelsensors. Nach erfolgter Fusion aller Einzelwahrnehmungen wird ein Fusionsergebnis $\hat{\theta}$ ausgegeben, das so weit wie möglich und im Sinne eines Optimalitätskriteriums optimal mit dem wahren Zustand übereinstimmt.

Die Sensordaten liegen in unterschiedlicher Repräsentationsform vor und sind infolge der Verwendung von unterschiedlichen, physikalisch verteilten Sensoren auf verschiedene Koordinatensysteme bezogen. Um den ersten der oben angegebenen Schritte vollziehen zu können, müssen die Sensordaten zunächst unter Zuhilfenahme des *Sensormodells* (das aus der Umwandlungsfunktion des Sensors und seinen Koordinaten besteht) in eine einheitliche Repräsentation transformiert

[1] Siehe dazu auch [Wegner 91]. Es gibt auch andere Modelle für die Fusion, siehe dazu [Krzysztofowicz 90] und [Clark 90, S.74].

werden, eine für alle Sensoren gleichermaßen gültige *Umweltbeschreibung*. Letztere ist ein mathematisches Modell der Umwelt, das geeignet parametrisiert ist. Jeder Sensor trägt seinen Teil zur Festlegung möglichst vieler Parameter dieses Modells bei. Vor der Fusion muß sichergestellt werden, daß von allen Sensoren dasselbe Objekt beobachtet wurde (Sicherstellung der *Konsistenz*). Ferner ist immer mit der Möglichkeit fehlerhafter Daten von einem Sensor zu rechnen. Besteht das System beispielsweise aus zwei Kameras, von denen die eine Kamera den rechten Teil und die andere den linken Teil eines Objektes beobachtet, muß zunächst in beiden Kameras der korrespondierende Bildausschnitt gesucht werden. Die Konsistenz der Sensordaten muß in einem solchen Fall durch ein geeignetes Kriterium überprüft werden, also durch die Berechnung eines *Abstandsmaßes*. Zu dessen Berechnung ist normalerweise die Kenntnis der Meßwerte aller Sensoren und deren Varianzen erforderlich. Verglichen wird der Abstand zwischen jeweils zwei Einzelsensoren des Systems. Je kleiner der Abstand, desto besser ist die Konsistenz zwischen den beiden Sensordaten. Wird ein Schwellwert geeignet festgelegt, können bei einer paarweisen Überprüfung aller Sensorwerte die inkonsistenten Daten herausgefiltert werden.

Bevor wir unsere Betrachtungen zu diesem Problemkreis mit der Untersuchung von Weltmodellen beginnen, wollen wir noch einen kurzen Überblick über wesentliche Forschungslinien und –ursprünge geben, der keinen Anspruch auf Vollständigkeit erhebt. Das Interesse an der Kombination von Meßergebnissen verschiedener Sensoren begann sich in den frühen achtziger Jahren[1] zu entwickeln, als sich einerseits die Erkenntnis durchsetzte, daß ein einzelner Sensor die anstehenden komplexen Aufgaben nicht bewältigen kann und andererseits durch die Fortschritte in der Mikroelektronik leistungsfähige Einzelsensoren so klein und preiswert wurden, daß Abmessungen und Kosten keinen Hinderungsgrund mehr für den Aufbau auch größerer Systeme darstellten. Bei den bearbeiteten Aufgaben lassen sich drei Entwicklungslinien unterscheiden: mobile Roboter, Montageroboter (mit Telerobotik), und theoretische Grundlagen.

Bei den mobilen Robotern besteht das Problem hauptsächlich in der Navigation, d.h. der Feststellung der eigenen Position und der Umfahrung von Hindernissen. Dazu werden intern „Karten“ der Umgebung verwaltet, die die von den Sensoren erfaßte Umwelt möglichst aktuell und genau widerspiegelt. Zu den ersten Arbeiten auf diesem Gebiet gehört [Briot 81; Chatila 85; Freyberger 85], mit Einschränkungen auch [Crowley 85; Ayache 87b]. Die die in diesem Umfeld anzutreffende Problematik der „unsicheren Geometrie“ theoretisch behandelnden Arbeiten sind [Durrant–Whyte 85; Smith 86]. In [Smith 86] und in [Faugeras 86c] werden die Ungenauigkeiten in einer Roboterumgebung in einem nichtlinearen Modell repräsentiert. Zur Fusion der Daten von verschiedenen Sensoren wurde das statische Kalman–Filter (bzw. bei Faugeras das erweiterte Kalman–Filter) vorgeschlagen, welches in diesem Fall äquivalent zur Bayesschen Methode mit dem Kriterium der minimalen Varianz ist (B). Es sei in diesem Zusammenhang auch auf die jüngeren Arbeiten [Dickmanns 91] zur Steuerung eines schnellen Kraftfahrzeuges für Außenanwendungen, [Horn 92] zur Steuerung eines Fahrzeugs mit universeller Sensorik zur Verwendung in Innenräumen, und [Hoppen 90; Stelmaszyk 91; Bonasso 92] hingewiesen.

---

[1] Im militärischen Bereich, wo Informationen aus vielen unterschiedlichen Quellen schnell verarbeitet werden müssen, wurde mit dieser Zielrichtung schon früher geforscht. Auch im Bereich der Rechnersichtsysteme wurden schon früh verschiedenartige Informationen kombiniert (siehe z.B. [Duda 76; Nitzan 77]), ein Beispiel aus jüngster Zeit ist die Kombination von photometrischem Stereo (aktives Verfahren) mit passivem Grauwert–Stereo [Maître 92]; siehe dazu auch [Magee 85; Reeves 85].

Etwa in die gleiche Zeit fallen die Arbeiten von Allen [Allen 85; Allen 87], die dem Gebiet der Objekterfassung für Montagezwecke zuzurechnen und aus den modellbasierten Ansätzen von Bajcsy [Bajcsy 85] entstanden sind. Sein System vereinigt die Informationen eines 3D–Sichtsystems mit denen eines taktilen Sensors. Das Sichtsystem besteht aus zwei starr montierten, auf den Arbeitstisch gerichteten CCD-Kameras. Der taktile Sensor ist auf einem Roboterarm mit sechs Freiheitsgraden angebracht und wird für die flächendeckende Untersuchung von Werkstückoberflächen eingesetzt, die zuvor vom Sichtsystem grob beschrieben wurden. Es handelt sich um ein heterogenes System; die Daten des Sichtsystems und des taktilen Sensors sind komplementär zueinander und sollen zusammen eine robustere Interpretation ermöglichen als die Daten des Sichtsystems allein. Anwendungsbereich war die Werkstückmontage in einer ungeordneten, relativ abgeschirmten Arbeitszelle. Dabei steht das Erkennen und Manipulieren nur graduell unterschiedlicher Werkstücke im Vordergrund. In den von Allen beschriebenen Experimenten wurden als Werkstücke verschiedene Teller, Tassen u.ä. verwendet (siehe auch [Allen 90; Allen 90b; Allen 91]). In die gleiche Richtung zielt auch [Cameron 88] und die Arbeit von [Zytkow 89], in letzerer kommen zusätzlich Lernverfahren aus dem Bereich der künstlichen Intelligenz zur Anwendung. Hochgenaue Multisensorik für Greifaufgaben (insbesondere für Verwendung im Rahmen der Telerobotik) wurde von Hirzinger [Hirzinger 89; Hirzinger 91; Hirzinger 92] entwickelt. Infrarot–Abstandsmesser für den mittleren Entfernungsbereich und eine Stereokamera steuern die Annäherungsvorgänge des Effektors an ein Objekt, der Greifvorgang wird durch das Zusammenwirken von Infrarot–Abstandsmessern für den Nahbereich und einem taktilen Kontaktfeld mit einer Kraftmeßdose kombiniert. Hier wurden Fortschritte in der Miniaturisierung der Elektronik dazu genutzt, ein großes Maß an eigenverantwortlicher Signalverarbeitung direkt am Sensor durchzuführen. Die Sensoren kommunizieren mit der Steuerung lediglich über einen Vierdraht–Bus. Ein erster Ansatz zur Kombination von taktilen Sensoren mit Abstandssensorik findet sich auch in [Dillmann 82]; in [Loughlin 82] wird der Einbau einer Kamera in einen Robotergreifer (zur Kombination mit taktiler Sensorik) beschrieben, ähnliches findet sich in [Holmbom 92]. Ansätze zur Integration eines passiven Sichtsystems mit einem Laser–Scanner finden sich in [Robert 86].

Die ersten theoretischen Arbeiten zur Repräsentation von räumlichen Unsicherheiten und Datenfusion[1] stammen von Durrant–Whyte [Durrant–Whyte 85; Durrant–Whyte 86; Durrant–Whyte 87] und Hager [Hager 87]. Meßungenauigkeiten werden durch statistische Modelle beschrieben. Es werden normalverteilte Meßfehler vorausgesetzt und zur Fusion die Bayessche Methode mit dem Kriterium der minimalen Varianz (siehe 2.1.2.3) für die Fusion der Sensordaten verwendet. Die Modellierung von Meßfehlern ist auch Gegenstand von [Brooks 85], wobei hier eine „Ungenauigkeitsmannigfaltigkeit“ für mobile Roboter definiert wurde und Schlußfolgerungen aufgrund von Markierungsinformationen vorgeschlagen wurden. In [Matthies 87] wurde ein Fehlermodell für ein Stereokamerasystem eines mobilen Roboters verwendet. Als Fusionsmethode wurde hier die Methode der kleinsten Quadrate mit der Inversen der Kovarianzmatrix als Gewichtung eingesetzt, was für statische Probleme ebenfalls der Bayesschen Methode mit dem Kriterium der minimalen Varianz entspricht. Diese Arbeiten waren allerdings nur eine Übertragung bekannter Ergebnisse der Statistik, Entscheidungstheorie und Nachrichtentechnik auf die speziellen Eigenheiten der Multi–Sensor–Datenfusion. Daran hat sich auch lange Zeit nichts geändert, weil die entsprechenden Ergebnisse aus diesen Feldern gut übertragbar sind.

[1] Siehe auch [Grimson 86].

Erst in jüngster Zeit lassen sich Ansätze für spezielle Theorien für die Sensordatenfusion identifizieren: In [Porrill 88] werden geometrische Gegebenheiten durch Listen stochastischer Parameter systematisch erfaßt und eine Schätzung des Umweltzustandes durch eine dem statischen Kalman–Filter äquivalente Gauß–Markow–Schätzung vorgenommen. In [Kamberowa 89] wird die Theorie für eine robuste, gegenüber Ausreißern toleranten Schätzung entwickelt und in [Wen 92] wird Vorwissen über die Szene nahtlos in den Schätzalgorithmus integriert, indem es als Messung ohne Fehler aufgefaßt wird. Als Beispiele für Verfahren, die nicht von Bayes–Schätzung, Maximum–Likelihood–Schätzung oder Kalman–Filter Gebrauch machen seien [Harmon 86] (verteiltes Blackboard zur Sensordatenfusion) und [Shekhar 86] (Fusion von Objektpositionen für taktile Sensoren) angeführt. Noch vor diesen Ansätzen wurde die Verwendung der Dempster–Shafer–Theorie (2.6) zur Datenintegration vorgeschlagen [Garvey 81]; siehe auch [Lowrance 86; Hutchinson 88; Abdulghafour 90; Liu 91]. Dieser Ansatz führt allerdings bis heute eher ein Schattendasein; das gleiche gilt bislang für den Einsatz von Fuzzy–Logic [Wise 86; Huntsberger 87; Weiß 87; Abidi 89; Abdulghafour 90; Raczkowsky 91; Wolfram 92]. Zu den theoretischen Arbeiten gehört auch die Definition eines Formalismus zur einheitlichen Darstellung verschiedener Sensoren. Ein solcher Formalismus ist zwar nicht Voraussetzung zur Sensordatenfusion, erleichtert aber ihre Durchführung. In [Henderson 83; Henderson 84; Henderson 84b] wurde dazu das Konzept der *logischen Sensoren* eingeführt, das die einheitliche und effiziente Behandlung verschiedener Sensoren ermöglichen soll. Wir gehen darauf in Abschnitt 1.3.1 näher ein. Dieses Konzept wurde später zum „Multisensor Kernel System“ weiterentwickelt [Henderson 85; Henderson 85b]. Schließlich seien die Ansätze aus dem Bereich der KI genannt, hier kristallisierte sich schon früh die Verwendung von *blackboard*–Systemen als erfolgversprechendste Variante heraus [Almand 85; Harmon 86]. Ein weiteres Anwendungsbeispiel ist die Korrektur von Fehlern, die bei der Ausführung eines Planes auftreten und die Unterstützung des Planers bei einer eventuell notwendigen Umplanung [Caglioti 88]. Außerdem können auch Sensoren gekoppelt werden, die auf ähnlcihen Wirkungsprinzipien beruhen, beispielsweise Abstand durch Fokusbestimmung und Stereo–Verfahren (siehe [Krotkov 88]; diese Problematik wird auch in [Clark 90] behandelt). Schließlich werden auch immer wieder etwas eigenartig anmutende Kombination vorgeschlagen, beispielsweise die Kombination von Kameras und Mikrowellen–Radar für die Weltraum–Navigation [Shaw 88] oder die Erkennung von Objekten der Größe von Personen [Rozmann 92], Sonar– und Infrarotsensoren in [Flynn 88; Caroll 91] oder Kameras und Temperatursensoren [Jain 90].

Wir schließen damit unsere einleitende Übersicht ab (eine andere, leider korrekturbedürftige Sammlung von Beispielen findet sich in [Hackett 90]; eine fundiertere Sicht vermittelt [Trivedi 90]) und wenden uns kurz dem ersten wichtigen Aspekt beim Entwurf eines Multisensorsystems zu: der Umweltmodellierung.

1.2 Weltmodelle

Umweltmodellierungen werden auch für Einzelsensoren verwendet. In diesem Fall besteht das Modell nur aus Parametern, die der Sensor bestimmen kann. Die Sensordatenrepräsentation entspricht somit der Weltmodellierung. Andere als vom Sensor meßbare Größen kommen im Modell nicht vor. Für ein Multisensorsystem muß eine Unterscheidung zwischen Sensordatenrepräsentation und Weltmodellierung eingeführt werden, da die einzelnen Sensoren nur *Teil*informationen der gesamten Weltinformation ermitteln. Jeder einzelne Sensor liefert Informationen über spezielle

Teilaspekte der Umwelt, die im Weltmodell geeignet repräsentiert werden müssen. Besteht ein System beispielsweise aus einer Kamera und einem Thermometer, muß das Weltmodell die Koordinaten von Objekten und auch Temperaturangaben enthalten. Sinnvollerweise wird dabei an jeder Koordinate innerhalb eines durch die Sensorgenauigkeit vorgegebenen Rasters ein Temperaturwert angebracht. Besonders wichtig bei einer solchen zusammengesetzen Modellierung ist die eben erwähnte Konsistenzüberprüfung zwischen voneinander abhängigen Sensordaten, die von unabhängigen Sensoren ermittelt wurden.

Es existiert eine Vielzahl von Methoden zur Repräsentation von Objekten im Raum, die als Grundlage für ein Weltmodell dienen können und im Einzelfall an die Sensorauflösung oder den gewünschten Detaillierungsgrad angepaßt werden müssen. Einen Überblick über Möglichkeiten zur effizienten Repräsentation von Oberflächen in Rechenautomaten gibt [Meinköhn 92], es wird dazu zweckmäßigerweise auf die bekannten Methoden der Computer–Graphik zurückgegriffen [Barnhill 75; Braid 75; Coons 67; Requicha 80; Foley 82; Encarnaçao 88]. Einfachere, in vielen Fällen aber wenig effiziente Repräsentationen sind Würfelbäume, Gitter und dreidimensionale Positions- und Orientierungsvektoren [Jackins 80; Chen 87b][1]. Bei einer Würfelbaumrepräsentation wird der gesamte Raum durch einen Würfel umschrieben, der in acht gleich große Unterwürfel aufgeteilt wird. Jeder inhomogene Unterwürfel wird erneut aufgeteilt. Homogen bedeutet hier, daß ein Würfel vollständig gleichartig belegt oder vollständig unbelegt ist. Diese Vorgehensweise führt zu einer hierarchischen Repräsentation des Raumes in Form eines Baumes, bei dem von einem Knoten immer acht Kanten zu jeweils homogenen oder inhomogenen Knoten gehen (sogenannte *Octrees*). In [Kent 86] wird diese Methode benutzt, um das Würfelmodell eines gegebenen Objektes in einem Kamerabild wiederzufinden. Gitterrepräsentationen werden von z.B. in [Matthies 88] für die Beschreibung von Sonar–Entfernungsdaten verwendet. Bei diesen Gittern handelt es sich um zweidimensionale Felder, deren Komponenten die Wahrscheinlichkeit für eine Belegung der Zellen angeben. Je größer deren Wert ist, desto größer ist die Wahrscheinlichkeit, daß die einer Komponente entsprechende Raumzelle belegt ist. Diese beiden Gitterrepräsentationen werden zu einem Würfelbaum zusammengefügt. Der Nachteil besteht darin, daß die Feinheit der Auflösung durch die einmal festgelegte Größe der Würfel und Zellen bestimmt wird, daß diese Darstellung nicht rotationsinvariant ist und daß der Speicheraufwand sehr groß ist. Außerdem ist diese Art der Repräsentation nur für raumbezogene Sensorinformationen geeignet. Beispielsweise lassen sich Temperaturen und Farben hier nicht zweckmäßig beschreiben. In [Crowley 86] werden verschiedene Entfernungssensoren verwendet, um ein dreidimensionales Oberflächenmodell von sich berührenden Objekten zu erstellen. Aus den Daten aller Sensoren werden allgemeine Oberflächenstücke erzeugt, deren Zusammensetzung die gesamte Oberfläche beschreibt. Bei [Stansfield 88] findet ein Sichtsystem und ein taktiler Sensor Verwendung[2], um dreidimensionale Merkmale, wie Ecken, und Tastmerkmale (Rauheit oder Rundungen) zu erkennen. Aus diesen Sensordaten wird eine hierarchische Formmodellierung erstellt. Das Kamerasystem wird für die grobe Objektlokalisierung benutzt, während der Tastsensor für die Feinheiten des Objektes eingesetzt wird. Es zeichnet sich hier kein allgemein zu empfehlender Ansatz a2. Die Anforderungen bezüglich zu integrierender Information, verlangter Auflösung, Zugriffs– und Speichereffizienz sind von Anwendung zu Anwendung zu unterschiedlich. Im Zusammenhang mit Anwendungen in der Robotik sollte man jedoch das Weltmodell an der von der Roboter–Steuerung verwendeten Repräsentation ausrichten, um sonst notwendige Transformationen zu sparen.

1 Siehe auch [Leyton 87].
2 Siehe auch [Luo 84; Cameron 88].

Als abschließendes Beispiel seien Weltmodelle angeführt, die auf einfachen und geeignet parametrierten geometrischen Objekten fußen [Ayache 87; Durrant–Whyte 87; Durrant–Whyte 87c; Grandjean 89; Hager 91], siehe auch [Acharya 88] und [Bolle 86]. Solche Modelle sind gut überschaubar und mathematisch einfach zu behandeln. Es werden einzelne geometrische Merkmale im n-dimensionalen euklidischen Raum $\mathbf{R}^n$ durch einen Parametervektor θ mit $\theta \in \mathbf{R}^m$ und eine mehrdimensionale Funktion

$$g(a,\ \theta) = 0$$

mit $a \in \mathbf{R}^n$ dargestellt. Jede Funktion g beschreibt ein spezielles geometrisches Merkmal, (Punkt, Linie, Ebene, Kugel) und jeder Vektor θ modelliert eine bestimmte Instanz dieses geometrischen Merkmals. So können beispielsweise alle Ebenen im $\mathbf{R}^3$ durch

$$g(a,\ \theta) = \theta_x\, x + \theta_y\, y + \theta_z\, z + 1 = 0$$

beschrieben werden mit $a = (x,\ y,\ z)$ und $\theta = (\theta_x,\ \theta_y,\ \theta_z)$. Der Zustandsvektor $\theta = (1,\ 1,\ 1)$ stellt dann eine spezielle Ebene im Raum dar. Ein Punkt kann direkt durch $\theta = (\theta_x,\ \theta_y,\ \theta_z)$ modelliert werden, während eine Kugel mit Mittelpunkt M und Radius r durch den Vektor $(\theta_{Mx},\ \theta_{My},\ \theta_{Mz},\ \theta_r)$ und eine Gerade entsprechend durch einen sechsdimensionalen Parametervektor repräsentiert werden kann.

Der Vorteil dieser Darstellung besteht darin, daß bei nicht genau bekannter Lage eines Objektes (Punkt, Linie, Ebene, Kugel) leicht eine kompakte, statistischer Charakterisierung der Objektlage möglich ist. Dazu wird bei dieser „unsicheren Geometrie“ jedem Parametervektor θ eine (im allgemeinen mehrdimensionale) Dichtefunktion $p(\theta)$ zugeordnet, die ein Maß für die Wahrscheinlichkeit dafür ist, daß eine bestimmte Instanz des Objektes (bzw. eine bestimmte Lage) beobachtet wurde. Da oftmals nur Teilinformationen eines Merkmals beobachtet werden können, sind auch unterbestimmte Dichtefunktionen möglich. Vor allem lassen sich Transformationen des Parametervektors, wie sie zum Beispiel beim Übergang von einem Koordinatensystem auf ein anderes vorkommen (notwendig, wenn die Sensoren an unterschiedlichen Orten plaziert sind), leicht auch für die zugehörige Dichtefunktion vornehmen. Sei θ ein Parametervektor (mit Dichtefunktion $p(\theta)$) in einem Koordinatensystem und sei θ' ein durch die Transformation D daraus hervorgegangener transformierter Vektor mit $\theta' = D(\theta)$ in einem anderen Koordinatensystem. Dann gilt für die transformierte Dichte $p'(\theta')$:

$$p'(\theta') = \frac{p(D^{-1}(\theta))}{\left|\det(J)\right|}$$

wo $J = \partial D / \partial \theta$ die Jacobi–Matrix der Transformation ist. Im Falle einer einfachen, linearen Koordinatentransformation ist dieser Ausdruck leicht zu berechnen. Bei nichtlinearen Transformationen wird die Berechnung aufwendig. Im Falle von durch Erwartungswert und Varianz vollständig charakterisierter normalverteilter Dichtefunktion $p(\cdot)$ kann man sich allerdings auf die Transformation von Erwartungswert und Varianz beschränken. Der Erwartungswert ergibt sich unter der Transformation

D zu: E[θ'] = E[D(θ)] = D(E[θ]). Die Varianz läßt sich über eine Reihenentwicklung des Ausdrucks für p'(θ') näherungsweise bestimmen zu $\Sigma_{\theta\copyright} = J\Sigma_{\theta}J^{T}$.

Die gesamte Umwelt wird bei diesen Modellen zweckmäßig als Graph dargestellt. Seine Knoten stellen die einzelnen Objekte (geometrische Merkmale) dar und seine Verbindungen repräsentieren die topologischen Relationen zwischen den Objekten. In einem solchen Modell lassen sich mit geringem Aufwand zusätzliche lokale oder globale Informationen einfügen. Der Nachteil liegt insbesondere darin, daß es auf die einmal vorgenommene Auswahl an geometrischen Primitiven (hier die Funktionen g(y, θ) = 0) beschränkt ist.

Wir kommen speziell in Abschnitt 1.4.5 noch einmal auf die Fusion bei komplexeren Umweltmodellen, die sich aus den eben aufgeführten Primitiven zusammensetzen, zurück. Es läßt sich aber schon hier festhalten, daß die Bereitstellung von Umweltmodellen, die sich im Kontext der Datenfusion verwenden lassen, eine wichtige Aufgabe für die Zukunft darstellt [Rao 87; Hager 91b; Kamberova 91]. Das gleiche gilt für die Entwicklung von Algorithmen, die Daten fusionieren können, die „in der Sprache“ dieser Umweltmodelle vorliegen.

1.3 Sensormodelle

Ein Sensormodell ist eine Abstraktion des physikalischen Sensorprozesses und beschreibt die Fähigkeit eines Sensors, ausgehend von den physikalischen Meßgrößen Informationen über die Umgebung zu gewinnen. Bei den bisherigen Einzelsensoren wurde die Modellierung oftmals implizit durch die Sensoreinschränkungen bezüglich einer bestimmten vorgeschriebenen Aufgabe vorgenommen. Diese Methode führt bei einem Multisensorsystem zu einer unnötig komplexen Integrationsprozedur und macht es unmöglich, die potentielle Mächtigkeit eines Multisensorsystems zu verstehen und zu nutzen. Grundlage eines Multisensorsystems sollte infolgedessen eine explizite mathematische Modellierung der Einzelsensoren sein (siehe dazu auch [Cowan 88; Durrant–Whyte 88; Hutchinson 88 bzw. Hutchinson 89]). Wir betrachten zunächst die logische Modellierung von Sensoren und danach die physikalische, die die Relation zwischen Umwelt und Sensormeßergebnis beschreibt.

1.3.1 Logische Sensormodelle

Beim Entwurf von Multisensorsystemen und bei der dynamischen Einsatzplanung von Sensorsystemen ist es nützlich, die Sensoren zunächst als abstrakte Gebilde einheitlicher Struktur mit unterschiedlich starker Ausprägung einzelner Eigenschaften aufzufassen. Ein Formalismus zur einheitlichen Darstellung von verschiedenen Sensoren ist zwar nicht Voraussetzung zur Sensordatenfusion, erleichtert aber ihre Durchführung. In [Henderson 84] wurde das Konzept der *logischen Sensoren* eingeführt, das eine solche einheitliche Behandlung verschiedener Sensoren unter Abstraktion von allen physikalischen Details ermöglichen soll. Die graphische Darstellung eines logischen Sensors zeigt Bild 2.

Ein logischer Sensor besteht aus einer Reihe von Komponenten: Der *Name* ist ein global eindeutiger Bezeichner für den logischen Sensor. Der charakteristische *Ausgangsvektor* definiert den Typ der Merkmale, die ein logischer Sensor liefern kann und wird zur Laufzeit vom *Selektor* mit konkreten Werten instantiiert. Der Selektor wählt aus n Programmeinheiten mit einem

Auswahlalgorithmus eine Einheit aus, die die aktuellen Eingaben in den logischen Sensor (also instantiierte Ausgangsvektoren anderer logischer Sensoren) verarbeiten soll. Die n Programmeinheiten stellen alternative Lösungswege dar, zu einer Instantiierung des Ausgangsvektors des logischen Sensors zu gelangen, wobei n konstant ist. Eine Programmeinheit besteht aus einer Menge von Eingabequellen (in Form von Merkmalsvektoren) und dem *Algorithmus*, der diese Eingabequellen bearbeitet. Ist die Menge der Eingabequellen leer, modelliert der logische Sensor einen physikalischen Sensor, wobei die entsprechende Programmeinheit nur den Algorithmus zur Verarbeitung seiner Daten enthält.

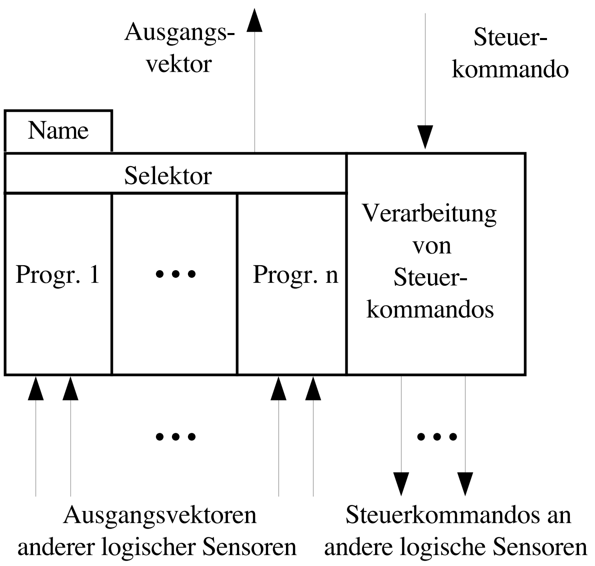

Bild 2: Komponenten eines logischen Sensors

Da beispielsweise eine Kamera kalibriert oder ein Abstandssensor positioniert und ein- oder ausgeschaltet werden muß, ist eine Komponente zur Verarbeitung von Steuerkommandos vorgesehen. Modelliert der logische Sensor einen physikalischen Sensor, wird dieser durch die Steuerkommandos direkt manipuliert; ansonsten werden die Steuerkommandos in angepaßter Form an die logischen Sensoren weitergegeben, deren Ausgangsvektoren als Eingabequellen dienen. Das Konzept logischer Sensoren bietet also zwei Vorteile bei der Spezifikation von Multi-Sensorsystemen:

- Es wird von physikalischen Sensoren abstrahiert, so daß sowohl reale Sensoren als auch nur Algorithmen als (logische) Sensoren aufgefaßt werden können.
- Es lassen sich Netze logischer Sensoren bilden, bei denen ein Knoten, der die Ausgangsvektoren mehrerer anderer logischer Sensoren verwendet, selbst einen komplexen (logischen) Sensor darstellt.

Eine Anwendung findet sich z.B. in [Schoess 88], wo jeder Ausgangsvektor ein Viertupel [r, s, c, t] darstellt mit

r als Sensorwert,
s als Schätzung für den Sensorfehler (Varianz),
c Sicherheitsfaktor, daß in einer bestimmten Situation der richtige Wert gemessen werden kann,
t Zeitstempel der Messung.

Das Konzept wird in [Henderson 84b; Henderson 85] zum „Multisensor Kernel System (MKS)“ weiterentwickelt. Ein Multi-Sensorsystem besteht gemäß diesem Modell aus einem steuernden *zentralen Kernsystem* sowie einer beliebigen Anzahl von logischen Sensorknoten. Das Kernsystem enthält ein Modell der Umwelt und kontrolliert die Anwendung mehrerer Sensoren (z.B. die Ausrichtung einer Kamera auf einen bestimmten Umweltausschnitt oder die zusätzliche Verwendung eines Ultraschallsensors). Ein implementiertes, prototypisches System besteht aus einer CCD-Kamera und einem laserbasierten Abstandssensor [Henderson 85]. Als typische Anwendungen eines solchen Ansatzes werden auch die automatisierte Zielerkennung [Bhanu 83] und die Einbettung in die Steuerung mehrgliedriger Robotergreifer [Jacobsen 84] genannt. Für die einheitliche Beschreibung der Daten verschiedener Sensoren werden die Sensorergebnisse über Objektmerkmale (z.B. Kanten, Punkte, Regionen) zu Merkmalsvektoren zusammengefaßt. Die Elemente eines Merkmalsvektors können von beliebigem Typ sein. Damit ist es möglich, beispielsweise die Daten einer Kamera und die eines Ultraschallsensors einheitlich zu verarbeiten: Der Merkmalsvektor der Kamera und der des Ultraschallsensors werden zu einem neuen Vektor zusammengefaßt, der eine vollständigere Beschreibung des beobachteten Objekts repräsentiert als die beiden ursprünglichen Vektoren allein. Keine Aussage wird über den Fusionsalgorithmus getroffen, der die Information aus beiden Informationsquellen zu einem tatsächlich benutzbaren Ergebnis verrechnet.

1.3.2 Mathematische Modelle der physikalischen Sensoreigenschaften

Die eben beschriebenen Modelle für das Sensorverhalten können nur einen Rahmen für die saubere Strukturierung und möglicherweise die Steuerung eines Multisensorsystems darstellen. Für die eigentliche Fusion ist jedoch ein exaktes mathematisches Modell des Sensors erforderlich. Erst eine solche exakte Beschreibung ermöglicht es, Sensorfähigkeiten zu analysieren und die Eigenschaften des Gesamtsystems zu schätzen. Für reale Sensoren sind statistische Modelle besonders geeignet. Der Grund hierfür liegt in den unvermeidbaren Meßfehlern. Werden Sensorinformationen über dieselbe Objekteigenschaft von verschiedenen Sensoren ermittelt, weichen sie infolge von Meßfehlern generell voneinander ab. Soll unter der Randbedingung solcher Ungenauigkeiten eine effiziente Sensordatenfusion durchgeführt werden, bietet sich die *bedingte Wahrscheinlichkeitsverteilung* als geeignete Beschreibungsform der von den Sensoren gelieferten Daten an. Sie beschreibt die Wahrscheinlichkeit dafür, daß der Sensor einen bestimmten Meßwert liefert unter der Bedingung (bezogen darauf), daß ein beobachtetes Merkmal einen bestimmten wahren Wert aufweist. Einer der Vorteile einer Sensorbeschreibung durch Wahrscheinlichkeitsverteilungen ist die Möglichkeit, eine Menge verschiedener Informationstypen in einer konsistenten Art zu integrieren und einen Fehlerausgleich zu erreichen. Darüberhinaus werden solche statistischen Modelle seit langem in anderen Gebieten verwendet, so daß auf bewährte Methoden zurückgegriffen werden kann, um Optimalität, Ungenauigkeit und Robustheit zu analysieren.

Generell wird für statistische Sensormodelle angenommen, daß der wahre, vom Sensor beobachtete Zustand durch einem M–dimensionalen Vektor θ repräsentiert werde und daß die Menge der vom Sensor gelieferten Meßergebnisse in einem Vektor X der Dimension K zusammengefaßt sei. Diese beiden Vektoren stehen zueinander in folgender expliziter Beziehung:

$$X = \mathrm{f}(\theta, V)$$

oder aber in impliziter Form

$$\mathrm{f}(X,\ \theta,\ V) = 0$$

Dabei ist $\mathrm{f}(\cdot)$ eine Funktion, die die Charakteristik des Sensors beschreibt und V ein Zufallsvektor, der den Einfluß von Fehlern repräsentiert, üblicherweise ein normalverteiltes Rauschen mit Mittelwert Null und Kovarianzmatrix Σ_V. Ein sehr einfaches Beispiel für X könnten die von einem 3–D–Sensor berechneten und abgelieferten Entfernungswerte für ein Punktobjekt mit wahren Koordinaten $\theta = (\theta_x, \theta_y, \theta_z)$ sein. Das Ziel der Modellierung ist eine effiziente Schätzung von θ aufgrund möglichst vieler Meßwerte X (von eventuell unterschiedlichen Sensoren mit unterschiedlichen Sensormodellen). Um die Betrachtungen zu vereinfachen, wird oftmals angenommen, daß es sich um lineare Fehlereinflüsse handelt, sich der Fehler also zu der Transformation der wirklichen Sensordaten addiert. Dann gilt

$$X = \mathrm{f}(\theta) + V$$

Sind θ und V darüberhinaus unabhängig voneinander, was häufig angenommen werden kann und immer angenommen wird, dann ist die Verteilung des Meßwertes (unter der Voraussetzung, daß θ der wahre Wert ist) gleich der Verteilung des Fehlers. Schreibt man dies als bedingte Wahrscheinlichkeitsverteilung, so ist

$$\mathrm{p}_X(x \mid \theta) = \mathrm{p}_V(v \mid \theta) = \mathrm{p}_V(x - \mathrm{f}(\theta) \mid \theta)$$

$\mathrm{p}_X(x \mid \theta)$ beschreibt also die Wahrscheinlichkeitsdichte für den Meßwert x unter der Voraussetzung, daß der wirkliche Wert θ ist. Eine Messung hängt von einer Vielzahl von Geräteeinflüssen ab und die exakte Form von $\mathrm{p}_X(x \mid \theta)$ beruht auf diesen Einflüssen. Eine exakte Beschreibung des Wahrscheinlichkeitsverhaltens der Sensorbeobachtungen läßt sich nur in den einfachsten Fällen finden. Vielfach ist eine exakte Modellierung allerdings auch gar nicht erwünscht. Die Gründe liegen in der Berechnungskomplexität solcher exakter Modelle. Mit der Aussage des zentralen Grenzwertsatzes der Wahrscheinlichkeitsrechnung kann der Fehler der Sensordaten in der Praxis als normalverteilt angenommen werden: Da jedem der verschiedenen Einflüsse eines Sensors eine (unter Umständen unterschiedliche) Wahrscheinlichkeitsverteilung zugrunde liegt, besteht die Wahrscheinlichkeitsverteilung der Sensordaten aus der Summe unterschiedlicher Verteilungen und kann folglich als näherungsweise normalverteilt angenommen werden. Für einen multidimensionalen, normalverteilten Fehler eines Sensorsystems mit Mittelwert Null und Kovarianzmatrix Σ_V ergibt sich für die (mehrdimensionale) Dichte des Meßwertes mit den gerade angestellten Überlegungen für Unabhängigkeit zwischen Fehler und wahrem Zustand θ: [1]

$$\mathrm{p}_X(x|\theta) = \frac{1}{\sqrt{2\pi|\Sigma_V|}}\, e^{-\frac{1}{2}(x-\mathrm{f}(\theta))^T \Sigma_V^{-1} (x-\mathrm{f}(\theta))}$$

---

1 Aus Gründen der einfacheren Notation wird für den Rest von Abschnitt 1 die Unterscheidung zwischen einer Zufallsvariable X und ihrer Realisierung x aufgegeben. Wir schreiben X, wenn wir abstrakt über die Variable sprechen und x, wenn wir ihre Dichtefunktion angeben.

wo $|\Sigma_V|$ die Determinante der Kovarianzmatrix Σ_V kennzeichnet (siehe auch 2.2). Eine solche Modellierung eines Sensorsystems ermöglicht die Entwicklung von einfachen Fusionsmethoden. Allerdings werden durch diese Modellierung keine nicht auf Rauscheinflüsse zurückzuführenden Ausreißer einbezogen, wodurch es zu erheblichen Fehlern kommen kann. Eine gute Konsistenzüberprüfung der Sensordaten ist deshalb bei dieser Modellierung zum Ausschluß von Ausreißern[1] unbedingt notwendig. Eine Berücksichtigung solcher Ausreißer direkt im Sensormodell ist mit Hilfe der sogenannten verunreinigten Gaußverteilung möglich, wie sie in [Huber 81] eingeführt und in [Durrant–Whyte 87; Durrant–Whyte 88] propagiert wird. Dabei wird davon ausgegangen, daß ein Sensor mit einer hohen Wahrscheinlichkeit Messungen im Rahmen einer relativ stark um den wahren Wert zentrierten Normalverteilung N_1 liefert, gelegentlich aber auch Ausreißer produziert, deren Wert weitgehend unvorhersehbar ist und deshalb durch eine sehr flache Normalverteilung repräsentiert wird. Im Falle eines M–dimensionalen Meß– und eines entsprechenden Zustandsvektors (der besseren Übersicht wird für die Transformation $\mathrm{f}(\cdot)$ die Identität angenommen) nimmt eine solche Verteilung die folgende Form an:

$$\mathrm{p}(x \mid \theta) = (1-\varepsilon)\frac{1}{\sqrt{(2\pi)^m |\Sigma_1|}} e^{-\frac{1}{2}(x-\theta)^T \Sigma_1^{-1}(x-\theta)} + \varepsilon \frac{1}{\sqrt{(2\pi)^m |\Sigma_2|}} e^{-\frac{1}{2}(x-\theta)^T \Sigma_2^{-1}(x-\theta)}$$

Hierbei gilt $|\Sigma_2| >> |\Sigma_1|$ und $0.05 < \varepsilon < 0.1$. Korrekte Werte (also keine Ausreißer) werden durch die multidimensionale Normalverteilung $N\sim(\theta, \Sigma_1)$ modelliert. Die Ausreißer (mit erheblich größerer Varianz Σ_2) werden durch die multidimensionale Normalverteilung $N\sim(\theta, \Sigma_2)$ repräsentiert. In [Durrant–Whyte 88] wird Σ_1 als bekannt vorausgesetzt, aber es existieren keine expliziten Informationen über Σ_2, und der genaue Wert von ε ist ebenfalls unbekannt.

Ein einfaches, gebräuchliches Beispiel für eine Sensormodellierung (z.B. [Durrant–Whyte 85; Luo 87b]) ist die lineare Transformation $\mathrm{f}(\theta) = A\theta$ mit additivem Fehler, also $X = A\theta + V$. Dabei ist θ wieder ein M–dimensionaler Zustandsvektor, X ein K–dimensionaler Beobachtungsvektor, V ein ebenfalls K–dimensionaler Störvektor und A eine $K \times M$–Matrix.

Schwieriger werden die Verhältnisse, wenn es sich bei der Transformation $\mathrm{f}(\cdot)$ um eine nichtlineare Funktion handelt, wie sie beispielsweise bei Stereo–Sichtsystemen auftritt (siehe z.B. [Matthies 87; Ayache 88; Ayache 88b]). In solchen Fällen wird zweckmäßigerweise eine Linearisierung um einen guten Schätzwert $\hat{\theta}$ für θ vorgenommen (siehe dazu auch das Vorgehen beim erweiterten Kalman–Filter). In erster Näherung erhält man dann

$$X = \mathrm{f}(\hat{\theta}) + J(\hat{\theta})\left[\theta - \hat{\theta}\right]$$

als lineare Näherung für die Transformationsfunktion um $\hat{\theta}$ herum. Dabei ist $J(\hat{\theta})$ die $N \times M$ Jacobi–Matrix $[j_{ik}]$ mit den Elementen

$$j_{ik} = \left.\frac{\partial \mathrm{f}_i(\theta)}{\partial \theta_k}\right|_{\theta = \hat{\theta}}$$

[1] Solche Ausreißer könnten zum Beispiel durch Programmierungsfehler zustande kommen.

wo f_i eine Komponente des Funktionenvektors $f(\cdot)$ und θ_j eine Komponente des Zustandsvektors darstellt. Wir betrachten auch hierzu ein Beispiel: Es seien eine bestimmte Anzahl von Kameras gegeben. Aus den Beobachtungen (u_i, v_i) eines Punktes $\theta = (x,y,z)$ in den jeweiligen Kamerakoordinatensystemen soll dessen Position möglichst genau geschätzt werden. Um uns auf das Wesentliche zu beschränken, nehmen wir hier an, daß es sich nur um zwei Kameras handle, die in Form eines Standard–Stereokamerasystem aufgebaut seien. Die beiden Kameras sind im Abstand d exakt achsenparallel ausgerichtet und verfügen jeweils über eine Brennweite ϕ. Aus den Beobachtungen $X_1 = (u_l,v_l,)$ und $X_2 = (u_r,v_r)$ eines Punktes $\theta = (x,y,z)$ im linken bzw. rechten Bild sind die wahren Koordinaten (x,y,z) des Punktes möglichst exakt zu bestimmen. Der allgemeine Fall mit beliebig vielen Kameras (oder Kamerastandorten) ist entsprechend zu behandeln, allerdings müssen dabei immer die Kalibrationsmatrizen mitgeführt werden, worauf wir hier verzichten. Die nichtlinearen Transformationsbeziehungen für den Zusammenhang zwischen den wahren Koordinaten des Punktes und den Koordinaten u, v auf dem Bildaufnehmer der Kamera lauten für die linke bzw. rechte Kamera:

$$X_1 = \begin{pmatrix} u_l \\ v_l \end{pmatrix} = \begin{pmatrix} \phi \dfrac{x + d/2}{\phi - z} \\ \phi \dfrac{y}{\phi - z} \end{pmatrix} \text{ bzw. } X_2 = \begin{pmatrix} u_r \\ v_r \end{pmatrix} = \begin{pmatrix} \phi \dfrac{x - d/2}{\phi - z} \\ \phi \dfrac{y}{\phi - z} \end{pmatrix}$$

Für die Jacobi–Matrix erhält man hier (wir betrachten nur noch die linke Kamera):

$$J(\theta) = \begin{bmatrix} \dfrac{\partial u_l}{\partial \theta_j} \\ \dfrac{\partial u_r}{\partial \theta_j} \end{bmatrix} = \begin{bmatrix} \dfrac{\phi}{\phi - z} & 0 & -\phi \dfrac{x + d/2}{(\phi - z)^2} \\ 0 & \dfrac{\phi}{\phi - z} & -\phi \dfrac{y}{(\phi - z)^2} \end{bmatrix}$$

Mit einem geeignet gewählten Schätzwert[1] für den wahren Wert von θ, den wir mit $\hat{\theta}_k = (\hat{x}_k, \hat{y}_k, \hat{z}_k)^T$ bezeichnen wollen, erhält man als lineare Transformation für diesen Fall:

$$X = \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} \phi \dfrac{\hat{x}_k + d/2}{\phi - \hat{z}_k} \\ \phi \dfrac{\hat{y}_k}{\phi - \hat{z}_k} \end{pmatrix} + \begin{bmatrix} \dfrac{\phi}{\phi - \hat{z}_k} & 0 & -\phi \dfrac{(\hat{x}_k + d/2)}{(\phi - \hat{z}_k)^2} \\ 0 & \dfrac{\phi}{\phi - \hat{z}_k} & -\phi \dfrac{\hat{y}_k}{(\phi - \hat{z}_k)^2} \end{bmatrix} \begin{pmatrix} x - \hat{x}_k \\ y - \hat{y}_k \\ z - \hat{z}_k \end{pmatrix}$$

was zu einem vollständigen Sensormodell noch beispielsweise durch eine additive Fehlerkomponente V zu erweitern ist. Dies entspricht einem Meßmodell, wie es sich für die lineare Kalman–Filterung

---

1 Im Falle mehrerer Kameras, die alle über das gleiche Sensormodell (mit unterschiedlichen Parametern) verfügen, nimmt man für den initialen Schätzwert zweckmäßigerweise die von der ersten betrachteten Kamera gelieferte Schätzung. Verfügt man über einen genaueren, aber möglicherweise langsamen Sensor, wie z.B. einen Laserscanner, kann dessen Meßwert als Initialwert Verwendung finden.

anbietet (dieses Meßmodell wird in 2.1.5, Beispiel 4 aufgegriffen und die Berechnung des Schätzwertes angegeben).

Entsprechend kann man auch im impliziten Fall vorgehen (wenn $f(X, \theta) = 0$) und eine nichtlineare Transformationsfunktion um einen Schätzwert $\hat{\theta}_k$ herum linearisieren (siehe dazu [Grandjean 89; Ayache 88]). Es ergibt sich:

$$0 = f(\overline{X},\theta) \approx f(X,\hat{\theta}_k) + \left(\left.\frac{\partial f(\xi,\theta)}{\partial \theta}\right|_{\substack{\xi=X \\ \theta=\hat{\theta}_k}} (\theta - \hat{\theta}_k) \right) + \left(\left.\frac{\partial f(\xi,\theta)}{\partial \xi}\right|_{\substack{\xi=X \\ \theta=\hat{\theta}_k}} (X - \overline{X}) \right)$$

Der ganz rechts stehende Term kann direkt als Fehlerterm interpretiert werden: Wenn $\overline{X} = E[X]$ als korrekte Messung interpretiert wird und X als tatsächliche (verrauschte), so gibt $\Delta = X - \overline{X}$ die Größe der Störung an (wie sie beispielsweise durch Quantisierungsfehler in der Bildaufnehmermatrix der Kamera hervorgerufen wird). Für Δ wird wieder Normalverteilung mit Mittelwert Null und bekannter Kovarianzmatrix Σ_Δ vorausgesetzt. Nach Umformung der letzten Gleichung erhält man das vollständige Sensormodell zu

$$-f(X,\hat{\theta}_k) + \left(\left.\frac{\partial f(\xi,\theta)}{\partial \theta}\right|_{\substack{\xi=X \\ \theta=\hat{\theta}_k}} \hat{\theta}_k \right) = \left(\left.\frac{\partial f(\xi,\theta)}{\partial \theta}\right|_{\substack{\xi=X \\ \theta=\hat{\theta}_k}} \theta \right) + \left(\left.\frac{\partial f(\xi,\theta)}{\partial \xi}\right|_{\substack{\xi=X \\ \theta=\hat{\theta}_k}} \Delta \right)$$

oder nach Einführung von J als Bezeichner für die Jacobi–Matrizen:

$$-f(X,\hat{\theta}_k) + J_\theta \hat{\theta}_k = J_\theta \theta + J_X \Delta$$

was wieder einem linearen Systemmodell $\Xi = A\theta + V$, wie oben angegeben, entspricht. Dabei ist $\Xi = -f(X,\hat{\theta}_k) + J_\theta \hat{\theta}_k$, also eine Funktion des Meßwertes X; A entspricht J_θ und $V = J_X\Delta$, ist also mit $N\sim(0, J_X \Sigma_\Delta J_X^T)$ verteilt. Damit können auch hier die üblichen Techniken (Maximum–Likelihood–Schätzung, Maximum–a–posteriori–Schätzung, Kalman–Filter) angesetzt werden.

Gelegentlich wird die implizite Form $f(X, \theta)$ auch nach θ aufgelöst, so daß sich eine Transformationsfunktion der Form $\theta = f(X)$ ergibt. Dies erscheint auf den ersten Blick naheliegender, weil der Sensor schließlich aus seinen Beobachtungen ein Bild der Natur liefert. Hat man ein Sichtsystem, so interessiert man sich hauptsächlich für $\theta = (x, y, z)$, also die Koordinaten eines Weltpunktes aufgrund der Beobachtung $X = (u, v)$. Man ist nicht in erster Linie daran interessiert, welche Koordinaten (u, v) bei einem gegebenen Weltpunkt ausgegeben werden. Bild 3 verdeutlicht die Situation.

Die Modellierung $\theta = f(X)$ ist allerdings nur dann sinnvoll, wenn Zwischendaten (*low–level–measurements*) verfügbar sind und wenn der Sensor in der Lage ist, den Zustand θ vollständig zu erfassen (also nicht nur einen Teil des gesamten Umweltausschnittes). Es zeigt sich darüber hinaus, daß für die meisten Schätzverfahren die oben angegebene Version $X = f(\theta)$ günstiger ist, weil sie durch Variation von θ als unabhängige Variable leicht die Erfüllung eines bestimmten Optimalitätskriteriums gestattet und damit den über alle Sensoren besten Zustand $\hat{\theta}$ berechnet.

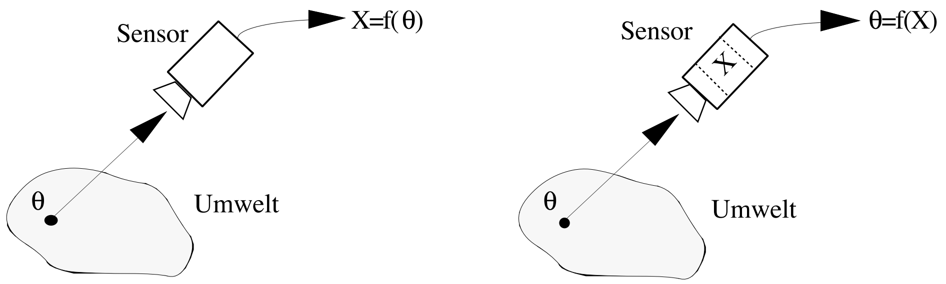

Bild 3: Unterschiedliche Sensormodellierung. Links: Standardmodellierung, ein Umweltmerkmal θ verursacht ein Ausgangssignal X. Rechts: Inverse Modellierung, ein Zustand führt zu internen Sensordaten X, die beobachtet werden und von höheren Verarbeitungsstufen auf den Umweltzustand θ zurückgerechnet werden.

Wir geben die Auflösung nach θ deshalb nur der Vollständigkeit halber an. Es wird hier von Meßwerten X ausgegangen, die direkt durch eine additive Rauschgröße Δ gestört sind, also $X = \overline{X} + \Delta$. Dabei sei Δ gemäß N~(0, Σ_Δ) verteilt. Setzt man dies in $\theta = \mathrm{f}(X)$ ein, so wird dies zu $\theta = \mathrm{f}(\overline{X} + \Delta)$. Ist f(·) nichtlinear, ergibt sich in erster Näherung der Wert von θ (wie ihn der Sensor ausgibt) abermals über eine Reihenentwicklung mit der Jacobi–Matrix der nichtlinearen Funktion

$$\theta \approx \mathrm{f}(\overline{X}) + J_X(\overline{X})\Delta$$

wobei E[θ] = $\mathrm{f}(\overline{X})$ und die Kovarianz $\Sigma_\theta = J_X(\overline{X})\Sigma_\Delta J_X(\overline{X})^T$. Liegt ein solches Sensormodell vor, können also direkt aus dem (bekannten) Meßfehler die statistischen Eigenschaften von dem vom Sensor abgelieferten (aus seiner Sicht richtigen, in Wahrheit aber durch verrauschte Messungen verfälschten) Weltzustand θ berechnet werden. Es liegen aber als Ergebnis bei *N* Sensoren nicht *N* Sensorergebnisse mit einem implizit gegebenen, wenn auch nicht genau bekannten wahren Zustand θ vor, sondern *N* Beziehungen mit *N* verschiedenen, als wahr ausgegebenen Zuständen θ. Dies ist für die Fusion ungünstiger, weil nur ein Zustand θ. gesucht und zu bestimmen ist.

1.4 Fusionsverfahren

Wie bereits mehrfach erwähnt, existieren für die Fusion von Sensordaten abhängig von der Problemstellung verschiedene Verfahren. Die einfachste Situation liegt bei einer *Integration* vor, wenn die Sensoren verschiedene Aspekte der Umwelt beschreiben, beispielsweise Temperatur und Koordinaten eines Gegenstandes. In solchen Fällen können die Sensordaten ohne irgendeine Konfliktsituation in das Weltmodell integriert werden. Diese Situation ist leicht zu behandeln und hier von geringerem Interesse. Die interessanteren Problemstellungen sind die, bei denen Daten von verschiedenen Sensoren zu identischen Umweltmerkmalen geliefert, bei denen also redundante Informationen ausgewertet werden[1]. Bei vollständig redundanten Daten (z.B. im Falle von Mehrfachmessungen) ist die einfachste Methode eine Durchschnittsberechnung. Sind die Genauigkeiten und Fehleranfälligkeiten der Einzelsensoren bekannt, können die verschiedenen Sensordaten entsprechend dieser Exaktheit (oder der Einzelmessung) unterschiedlich gewichtet werden.

Wir behandeln nun die verschiedenen Aspekte, die bei der Fusion und den im allgemeinen verwendeten Methoden zur Schätzung eine Rolle spielen.

1 Schwieriger zu behandeln sind Situationen, bei denen Größen gemessen werden, die nur vordergründig unabhängig sind, wie etwa die von einem Ultraschallsensor ausgegebene Entfernung und die Raumtemperatur.

1.4.1 Fusion mit Bayesscher Schätzung

Viele Methoden zur Fusion beruhen in der einen oder anderen Weise auf dem Satz von Bayes (dazu gehören auch die auf dem statischen Kalman–Filter beruhenden Verfahren). Wir gehen zunächst allgemein auf die zugrunde liegende Problematik ein und untersuchen dann als Anwendungsbeispiel eine Möglichkeit zur Fusion von optischen und akustischen Daten.

Das generelle Problem stellt sich, wie bei allen anderen Verfahren auch, wie folgt: Aufgrund der N Meßwerte X mit $X = [X_1, X_2, \ldots, X_K]^T$ soll eine möglichst gute Schätzung für die unbekannte M–dimensionale gesuchte Größe θ (Zustandsvektor) gefunden werden. Dabei wird angenommen, daß θ und X reelle, kontinuierliche Werte im unendlichen euklidischen Raum der entsprechenden Dimensionen M und K sind. Als Sensormodell wird das in Abschnitt 1.3 vorgestellte Modell gewählt, d.h. der Meßfehler soll sich zu der Transformation der wirklichen Sensordaten addieren:

$$X = \mathrm{f}(\theta) + V$$

Dabei bildet die Funktion $\mathrm{f}(\theta)$ den M–dimensionalen Zustandsvektor in einen Vektor der Dimension K ab.

Ein einfaches, sehr gebräuchliches Beispiel für eine solche Transformation ist die lineare Transformation $\mathrm{f}(\theta) = A\theta$, wobei A eine $K \times M$–Matrix ist. Der Fehler soll normalverteilt mit Mittelwert Null und Kovarianzmatrix Σ_V sein, so daß sich für die Fehlerverteilung (unter angenommener Unabhängigkeit zwischen Fehler V und wahrem Zustand θ) wieder die mehrdimensionale bedingte Normalverteilungsdichte ergibt:

$$\mathrm{p}(x|\theta) = \frac{1}{\sqrt{2\pi|\Sigma_V|}} e^{-\frac{1}{2}(x-\mathrm{f}(\theta))^T \Sigma_V^{-1} (x-\mathrm{f}(\theta))} = \frac{1}{\sqrt{2\pi|\Sigma_V|}} e^{-\frac{1}{2}(x-A\theta)^T \Sigma_V^{-1} (x-A\theta)}$$

Bei Anwendung der Bayesschen Schätzung ist die Erzeugung von a–priori– Information erforderlich. Es wird hierzu davon ausgegangen, daß diese ebenfalls in Form einer Normalverteilung $\pi(\theta)$ für den Vektor θ vorliegt, also

$$\mathrm{p}(x|\theta) = \frac{1}{\sqrt{2\pi|\Sigma_\theta|}} e^{-\frac{1}{2}(\theta-\bar{\theta})^T \Sigma_\theta^{-1} (\theta-\bar{\theta})}$$

Ihre Kennwerte (Erwartungswertsvektor $\bar{\theta}$ und Kovarianzmatrix Σ_θ) werden zweckmäßigerweise durch Experimente ermittelt. Die gesuchte *a–posteriori–Dichtefunktion* $\mathrm{p}(\theta|X)$, die den Stand des Wissens nach Sicht der Daten darstellt, berechnet sich anhand des Satzes von Bayes zu

$$\mathrm{p}(\theta \mid X) = \frac{\mathrm{p}(X \mid \theta)\pi(\theta)}{\mathrm{p}(X)}$$

wo die Bezugsdichte $\mathrm{p}(X)$ durch Integration über den gesamten Wertebereich jeweils aller Komponenten von θ

$$p(X) = \int_\theta p(X \mid \theta)\pi(\theta)\,d\theta$$

bestimmt wird. p(X) hängt nur vom Meßwert ab, der bei der Suche nach dem optimalen Wert für den Zustand θ jedoch fest bleibt, und deshalb muß dieses Integral im allgemeinen nicht ausgewertet werden. Je nach Wahl der *Verlustfunktion* (siehe 2.1.2) läßt sich damit eine Berechnungsvorschrift für den *optimalen Schätzwert* $\hat{\theta}$ für θ bestimmen. Durch entsprechende Wahl der Transformation f(θ) kann mit Bayesscher Schätzung für θ immer eine (verbesserte) Dichtefunktion bestimmt werden, egal ob die Meßwerte von unterschiedlichen Sensortypen, denselben Sensortypen oder zeitsequentiell von einem einzelnen Sensor stammen (für das erforderliche Vorgehen siehe 2.1.3). Interessant ist in diesem Zusammenhang die Frage, ob die Hinzunahme von weiteren Sensoren das Schätzergebnis auf jeden Fall verbessert, oder ob die Gefahr besteht, daß es sich verschlechtert. Diese Frage ist mit der folgenden Argumentation leicht zu beantworten: Nachdem die Menge möglicher Schätzungen durch die Hinzunahme weiterer Sensoren erhöht wird, erhöht sich auch die Anzahl der Minima der Risikofunktion[1] R(θ). Die ursprünglichen Minima (vom ersten Sensor verursacht) bleiben jedoch erhalten. Ausgewählt wird immer das absolute Minimum der Risikofunktion, so daß sich das Schätzergebnis verbessert, wenn die Hinzunahme ein kleineres Minimum verursacht. Das Schätzergebnis bleibt gleich, wenn die zusätzlichen Minima einen größeren Wert annehmen als die ursprünglichen. Im allgemeinen ist die Wahrscheinlichkeit, daß das Bayessche Risiko gleich bleibt, relativ klein im Vergleich zu der Wahrscheinlichkeit, daß das Risiko kleiner wird. Anders ausgedrückt: Die Hinzunahme weiterer Sensoren zur Datenfusion lohnt sich praktisch immer, speziell dann, wenn der Fusionsalgorithmus dafür sorgt, daß die Information entsprechend ihrer Glaubwürdigkeit verwertet wird.

Wir betrachten nun den Einfluß der a–priori–Informationen, die bei kleiner Sensoranzahl einen starken Einfluß auf den Schätzwert haben[2]. Einige Beispiele zeigen die folgenden Bilder 4 und 5. Es wird demonstriert, welches Ergebnis eine Bayes-Schätzung hat, wenn zehn Sensorwerte und ein a-priori-Wert in der oben angedeuteten Weise kombiniert werden. Es wurde von einem wahren Zustand Null ausgegangen und 10 Sensorwerte zufällig aus Normalverteilungen mit Mittelwert Null und den Varianzen 0.02, 0.3, 0.03, 0.09, 0.09, 0.06, 0.06, 0.01, 0.01, 0.02 erzeugt. Dann wurde für diese Stichprobenmenge jeweils der Schätzwert berechnet. Dies wurde zwanzigmal ausgeführt und jeweils der Schätzwert aufgetragen. Die Bilder zeigen das mit a–priori–Information berechnete Ergebnis (helle Säulen) und den ohne diese Information berechneten Schätzwert (dunkle Säulen). In Bild 4a wird die Auswirkung deutlich falscher a–priori–Information N~(0.09, 0.01) gezeigt. Diese Annahme führt zu einer starken Anhebung aller Schätzwerte in Richtung des Mittelwertes 0.09 und folglich im allgemeinen zu schlechteren Schätzwerten, als wenn gar keine Vorannahme gemacht wurde. Bild 4b zeigt die Schätzwerte bei a–priori–Information gemäß N~(0.09, 0.09), also einer wesentlich größeren Varianz (Unsicherheit) als in Bild 4a. Infolge der größeren Varianz wirkt sich diese a–priori–Information nicht mehr so stark aus. Eine leichte Anhebung der Schätzwerte ist aber auch hier vorhanden.

Interessant ist nun die Untersuchung des Verhaltens bei guter a–priori–Information. Bild 5a zeigt die Auswirkung eines sehr guten a–priori–Mittelwertes 0.000001 bei kleiner Varianz 0.01. Obwohl

---

1 Siehe Abschnitt 2.3.
2 Zu den Berechnungsvorschriften siehe 2.5.2.1

die a–priori-Information für den Mittelwert sehr gut war, werden die Schätzwerte nur geringfügig besser. Dies liegt an der gleichgebliebenen Varianz. Erst bei ebenfalls reduzierter Varianz, wie beispielsweise 0.001 in Bild 5b, zeigt sich eine erhebliche Verbesserung. Um die Auswirkung schlechter a–priori–Informationen zu verhindern, sollte daher eine große Varianz gewählt werden, damit sich schlechte a–priori–Mittelwerte nicht auswirken können. Leider haben in diesem Fall auch gute a–priori–Werte weniger Einfluß auf das Ergebnis. Infolge dieser Schwierigkeiten kann es besonders bei kleiner Sensoranzahl und schlechter a–priori–Informationen zu schlechten Schätzwerten kommen. Der Wahl guter a-priori-Information kommt bei der Bayes-Schätzung daher offensichtlich große Bedeutung zu.

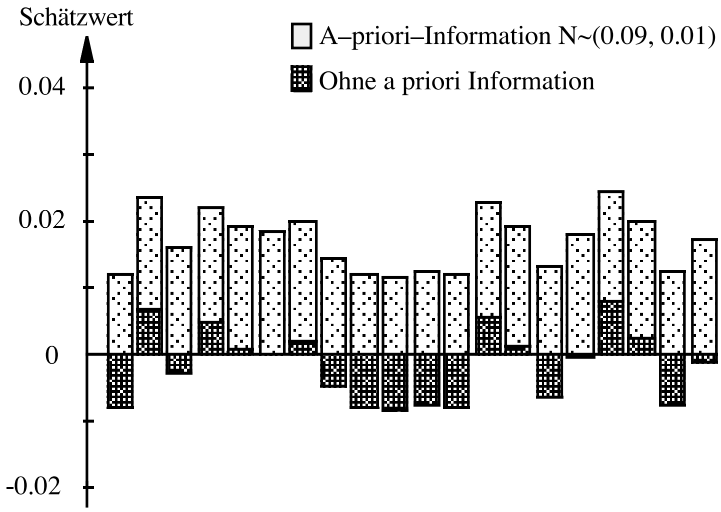

Bild 4a: Schätzwerte ohne und mit a priori Information N~(0.09, 0.01)

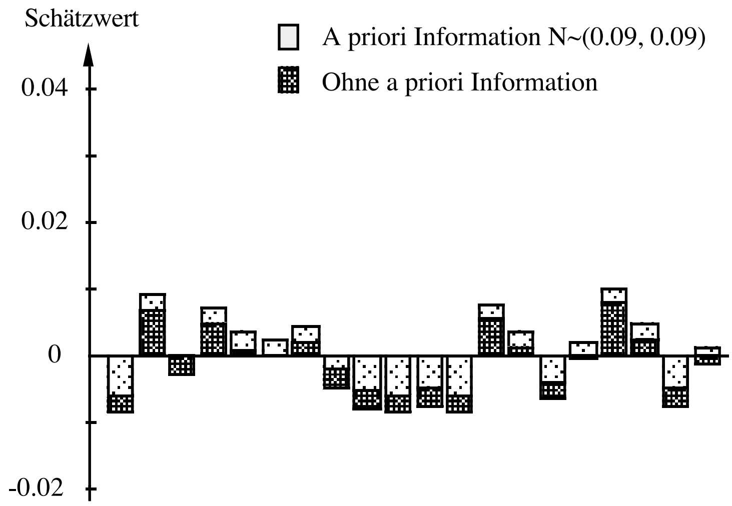

Bild 4b: Schätzwerte ohne und mit a priori Information N~(0.09, 0.09)

Wir gehen nun der Frage nach, unter welchen Voraussetzungen der Fusionsprozeß nach dem beschriebenen Verfahren verteilt ablaufen kann, inwieweit also Berechnungen, die während des Schätzprozesses notwendig werden, von Prozessoren direkt an den Sensoren vorgenommen werden können, bevor sie dem Fusionsalgorithmus zugeleitet werden, um θ zu bestimmen (siehe dazu auch

[Richardson 88]). Dabei ist es offensichtlich wünschenswert, die Verarbeitung soweit wie möglich beim Sensor vorzunehmen und in der zentralen Schätzeinheit, welche das Endergebnis berechnet, nur noch diese vorverarbeiteten Daten zusammenzufassen. Das betrachtete System soll sich aus zwei selbständigen Moduln zusammensetzen. Die Meß*vektoren* X_1 und X_2 sollen unabhängig von zwei Sensorprozessoren vorverarbeitet werden. Das Ergebnis dieser Vorverarbeitung soll in zwei weiteren Vektoren $\theta_1^{\surd}$ und $\theta_2^{\surd}$ stehen. Diese beiden Vektoren müssen genügend Information enthalten, um eine optimale Schätzung unter Berücksichtigung aller Meßwerte zu ermöglichen. Das Problem stellt sich also wie folgt: Welche Voraussetzungen sind nötig, damit eine separate Verarbeitung überhaupt möglich wird und wie können $\theta_1^{\surd}$ und $\theta_2^{\surd}$ charakterisiert werden?

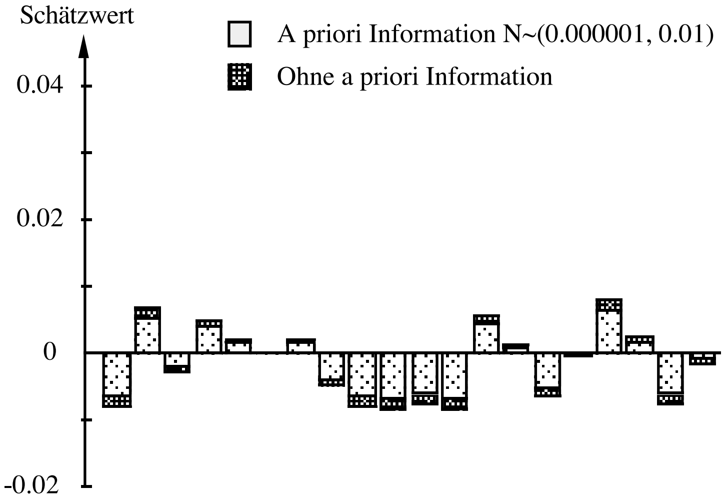

Bild 5a: Schätzwerte ohne und mit a priori Information N~(0.000001, 0.01)

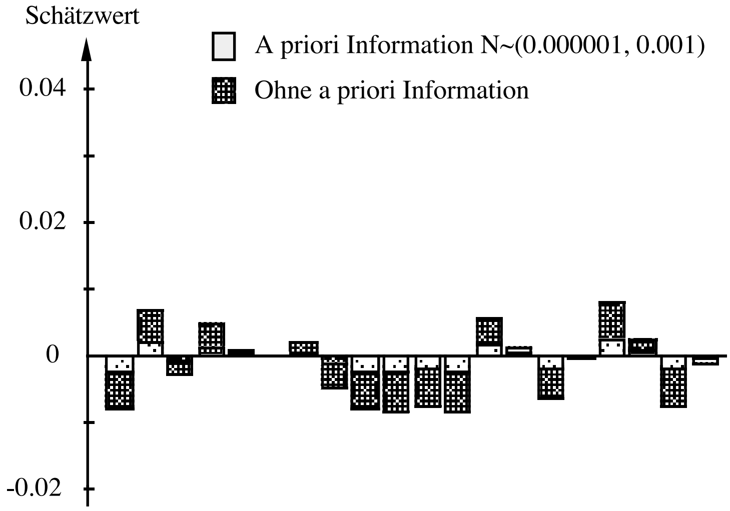

Bild 5b: Schätzwerte ohne und mit a priori Information N~(0.000001, 0.001)

Zunächst ist offensichtlich, daß bei einer separaten Verarbeitung die Meßvektoren X_1 und X_2 für gegebenes θ voneinander unabhängig sein müssen:

$$p(X_1, X_2 \mid \theta) = p(X_1 \mid \theta)\ p(X_2 \mid \theta)$$

Das Sensormodell besteht dann entsprechend aus zwei Teilen (nämlich gerade für beide Moduln getrennt):

$$X_1 = f_1(\theta) + V_1 \text{ und}$$
$$X_2 = f_2(\theta) + V_2$$

wo auch die Fehler V_1 und V_2 unabhängig sind. Wir berechnen zunächst die erforderliche, auf beiden Meßvektoren basierende a–posteriori–Dichte $p(\theta \mid X_1, X_2)$:

$$\begin{aligned} p(\theta \mid X_1, X_2) &= \frac{p(X_1, X_2 \mid \theta)\pi(\theta)}{p(X_1, X_2)} \\ &= \frac{p(X_1 \mid \theta)\, p(X_2 \mid \theta)\pi(\theta)}{p(X_1, X_2)} \\ &= \frac{p(\theta \mid X_1)\, p(X_1)}{\pi(\theta)} \frac{p(\theta \mid X_2)\, p(X_2)}{\pi(\theta)} \frac{\pi(\theta)}{p(X_1, X_2)} \\ &= \frac{p(\theta \mid X_1)\, p(\theta \mid X_2)}{\pi(\theta)} \frac{p(X_1)\, p(X_2)}{\pi(X_1, X_2)} \end{aligned}$$

wo $p(X_1, X_2)$ nur die Rolle eines Normierungsfaktors hat. Diese Beziehung soll nun so weit wie möglich separiert werden, um eine getrennte Berechnung von Teiltermen zu ermöglichen. Wir betrachten dies nun unter drei verschieden stark einschränkenden Voraussetzungen:

1. Wird außer der Unabhängigkeit von X_1 und X_2 keine weitere Voraussetzung gemacht, so können lediglich die bedingten Dichtefunktionen $p(\theta \mid X_1)$ und $p(\theta \mid X_2)$ getrennt ausgewertet werden, das heißt, als Z_1 und Z_2 gewählt werden. Der rechte Bruch in der letzten Gleichung dient nur zur Normierung und kann bei der Berechnung von θ' vernachlässigt werden. Alternativ kann auch die Umformung in der zweiten Zeile der letzten Gleichung verwendet und $p(X_1 \mid \theta)$ sowie $p(X_2 \mid \theta)$ getrennt berechnet werden. Im allgemeinen ist ohne zusätzliche Annahmen eine weiterreichende separate Berechnung für θ bei Messung von X_1 und X_2 nicht möglich.

2. Das betrachtete System besteht aus zwei separaten, nicht zusammenarbeitenden Teilen, die getrennte Bereiche der Umwelt beobachten. Der Zustandsvektor θ setzt sich hier zusammen aus zwei Einzelvektoren θ_1 und θ_2, also:

$$\theta = (\theta_1 \| \theta_2)$$

Die Beobachtung X_1 hängt nur von θ_1 ab, und die Beobachtung X_2 nur von θ_2. Es ist offensichtlich, daß die beiden Schätzungen für die beiden Teilvektoren vollständig getrennt vorgenommen werden können. Wir betrachten dies dennoch unter wahrscheinlichkeitstheoretischen Gesichtspunkten, wobei wir davon ausgehen, daß θ_1 und θ_2 a priori unabhängig sind. Für die a–posteriori–Dichtefunktion folgt mit der Unabhängigkeit von X_1 und X_2 zunächst

$$\mathrm{p}(X_1, X_2) = \mathrm{p}(X_1)\,\mathrm{p}(X_2)$$

sowie

$$\mathrm{p}(X_1, X_2 \mid \theta_1, \theta_2) = \mathrm{p}(X_1 \mid \theta_1, \theta_2)\,\mathrm{p}(X_2 \mid \theta_1, \theta_2)$$

und mit der Unabhängigkeit von θ_1 und θ_2:

$$\mathrm{p}(X_1, X_2 \mid \theta_1, \theta_2) = \mathrm{p}(X_1 \mid \theta_1)\,\mathrm{p}(X_2 \mid \theta_2)$$

Außerdem ist natürlich

$$\pi(\theta_1, \theta_2) = \pi(\theta_1)\pi(\theta_2)$$

Für die gesuchte a–posteriori–Dichtefunktion ergibt sich mit diesen Beziehungen sofort:

$$\begin{aligned}
\mathrm{p}(\theta_1, \theta_2 \mid X_1, X_2) &= \frac{\mathrm{p}(X_1, X_2 \mid \theta_1, \theta_2)\pi(\theta_1, \theta_2)}{\mathrm{p}(X_1, X_2)} \\
&= \frac{\mathrm{p}(X_1 \mid \theta_1)\,\mathrm{p}(X_2 \mid \theta_2)\pi(\theta_1)\pi(\theta_2)}{\mathrm{p}(X_1)\,\mathrm{p}(X_2)} \\
&= \mathrm{p}(\theta_1 \mid X_1)\,\mathrm{p}(\theta_2 \mid X_2)
\end{aligned}$$

Die gemeinsame a–posteriori–Dichte ist also das Produkt der (unabhängig) zu schätzenden a–posteriori–Dichten der beiden Einzelvektoren. Es handelt sich also beim Ergebnis tatsächlich um die direkte Kombination zweier separater Schätzungsprobleme.

3. Neben der Unabhängigkeit von X_1 und X_2 soll hier von normalverteilten a–posteriori–Dichtefunktionen $\mathrm{p}(\theta|X_1)$ sowie $\mathrm{p}(\theta|X_2)$ und den linearen Systemtransformationen

$$\mathrm{f}_1(\theta) = A_1\theta \text{ und}$$
$$\mathrm{f}_2(\theta) = A_2\theta$$

und damit $X_1 = A_1\theta + V_1$ ausgegangen werden, wobei A_1 und A_2 $\mathrm{K} \times \mathrm{M}$–Matrizen sind. Aufgrund der Eigenschaften normalverteilter Zufallsvariablen kann der gesamte Schätzwert in diesem Fall durch die folgende Beziehung berechnet werden (Herleitung siehe Beispiele in 2.5.2.1):

$$\hat{\theta}(X_1, X_2) = \left(\Sigma_{\bar{\theta}}^{-1} + \Sigma_{X_1}^{-1} + \Sigma_{X_2}^{-1}\right)^{-1}\left(\Sigma_{X_1}^{-1}\hat{\theta}(X_1) + \Sigma_{X_2}^{-1}\hat{\theta}(X_2) + \Sigma_{\bar{\theta}}^{-1}\bar{\theta}\right)$$

Dabei ist $\bar{\theta}$ im ersten Schritt der a–priori–Erwartungswert und $\Sigma_{\bar{\theta}}$ die a–priori–Kovarianzmatrix des Systemzustandes. Für alle späteren Beobachtungen sind dies die errechneten neuen Schätzwerte, die als neue a–priori–Werte fungieren. $\hat{\theta}(X_1)$und $\hat{\theta}(X_2)$ sind die auf den Beobachtungen X_1 und X_2 beruhenden Schätzungen von θ mit Σ_{X_1} sowie Σ_{X_2} als zugehörigen (a–posteriori, d.h. hier zu einem Meßwert gehörigen) Kovarianzmatrizen. Hier können also zunächst zwei unabhängige Schätzungen $\hat{\theta}(X_1)$und $\hat{\theta}(X_2)$vorgenommen werden und der Fusionsprozeß muß lediglich die letzte Beziehung auswerten, in der die „vorgeschätzten" Werte $\hat{\theta}(X_2)$und $\hat{\theta}(X_2)$ verrechnet werden. Die Berechnung der letzten beiden Größen geschieht zweckmäßig mit den Beziehungen des statischen Kalman–Filters, denn hier ist optimale Schätzung nach Bayes im Sinne minimaler Varianz gegeben. Mit den einfachen Beziehungen am Ende von 2.5.5 erhalten wir für den hier vorliegenden Fall für die Berechnung von $\hat{\theta}(X_2)$ aus X_1 mit der Hilfsgröße $\hat{K}$(entsprechend für $\hat{\theta}(X_2)$):

$$\hat{K} = \Sigma_{\bar{\theta}} A_1^T \left[A_1 \Sigma_{\bar{\theta}} A_1^T + \Sigma_v \right]^{-1}$$
$$\hat{\theta}(X_1) = \bar{\theta} + \hat{K}[X_1 - A_1 \bar{\theta}]$$
$$\Sigma_{X_1} = \Sigma_{\bar{\theta}} - \hat{K} A_1^T \Sigma_{\bar{\theta}}$$

Zusammenfassend kann bei Unabhängigkeit zwischen zwei Sensordatenmengen eine Modularisierung durchgeführt werden, die zu einer aufgeteilten Berechnung des Schätzwertes führt.

Als Beispiel entwickeln wir ein Modell für die Fusion der Daten von zwei entfernungsmessenden Systemen, also beispielsweise einem akustischen und einem optischen System. Diese Kombination erfreut sich großer Popularität speziell im Bereich mobiler Roboter (siehe z.B. [Ruokangas 86; Matthies 88; Kweon 92]). Das zu untersuchende Objekt soll sich auf einer massiven Oberfläche befinden. Bei beiden Systemen werden die Daten durch einen nach unten auf das Objekt gerichteten Sensor gewonnen. Für die Gewinnung der akustischen Daten (X_1 in der oben benutzten Terminologie) werde eine Messung nach dem Prinzip des Echolots und für die optischen Daten X_2 eine Entfernungsmessung nach dem Stereo–Prinzip vorgenommen. Die statistische Unabhängigkeit der akustischen und optischen Sensordaten bei gegebener Entfernung z (d.h. die statistische Unabhängigkeit der Fehler) wird vorausgesetzt. Die Berechnung von $p(\theta | X_1)$ und $p(\theta | X_2)$ kann also getrennt vorgenommen werden.

Das Ergebnis der Berechnung ist eine Entfernungskarte, d.h. die Auftragung aller Entfernungswerte z über der Meßfläche (x, y). Wir beschränken uns auf den zweidimensionalen Fall $z = f(x)$, wobei die Werte x entlang der Meßlinie entsprechend diskretisiert sind. Jedes der N Rasterelemente trägt einen Index i (siehe Bild 6). Die Größe der Rasterelemente hängt von der lateralen Auflösung der beiden Sensoren ab. Der Einfachheit halber bezeichnen wir mit z_i den Entfernungswert, den der Sensor nach seiner Messung für das Rasterelement i bestimmt hat.

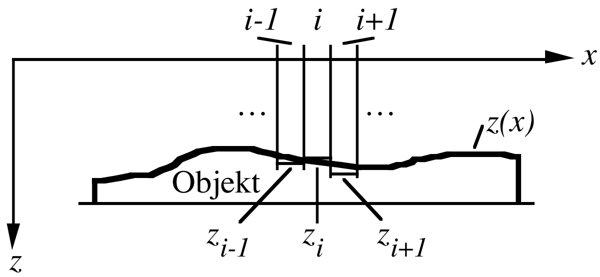

Bild 6: Meßmodell für die Entfernungskartenbestimmung

Zu schätzen ist hier der Vektor $\theta = (Z_1, Z_2, \ldots, Z_N)$ unter der Beobachtung $\zeta_A = (\zeta_{A1}, \zeta_{A2}, \ldots, \zeta_{AN})$ für die akustischen Entfernungswerte und $\zeta_O = (\zeta_{O1}, \zeta_{O2}, \ldots, \zeta_{ON})$ für die optischen Werte. Als Sensormodell setzen wir für den akustischen Sensor voraus:

$$\zeta_A = Z + V_A$$

d.h. die Entfernungsmeßwerte setzen sich zusammen aus der Überlagerung der wahren Werte und einem Fehlerterm, d.h. einem gaußverteilten Zufallsvektor mit Mittelwert Null und Kovarianzmatrix Σ_A. Aufgrund der statistischen Unabhängigkeit zwischen den einzelnen Messungen, die vorausgesetzt werden kann (wenn man einmal von systematischen Fehlern oder globalen Einflüssen, wie der Lufttemperatur, absieht), hat diese Matrix nur Einträge auf der Hauptdiagonalen. Wir nehmen an, daß diese alle gleich groß sind, und zwar σ_A. Verschieden große Einträge sind dann zweckmäßig, wenn die Unsicherheit des Meßergebnisses z.B. zu den Rändern hin bei zentral plaziertem Sensor zunimmt. Die Behandlung eines solchen Falles bereitet keine Schwierigkeiten. Wir erhalten unter dieser Voraussetzung mit den Überlegungen aus Abschnitt 1.3 für die bedingte Wahrscheinlichkeit

$$\mathrm{p}(\zeta_A \mid Z) = \frac{1}{\sqrt{2\pi}^N \sigma_A^N} e^{\frac{1}{2\pi\sigma_A^2}\left(\sum_{i=1}^{N} (\zeta_{A_i} - z_i)^2\right)}$$

Das *optische Modell* setzen wir ganz entsprechend an. Wird ein Laserabstandsmesser verwendet, ist eine mehrfache Durchführung einer Messung unumgänglich. Daraus folgt mit dem zentralen Grenzwertsatz ohnehin, daß es sich bei dem schließlich ausgegebenen Meßergebnis um ein normalverteiltes handelt. Wenn wir für diesen Sensor als Meßmodell

$$\zeta_O = Z + V_O$$

verwenden, so erhalten wir entsprechend für die bedingte Verteilungsdichte

$$\mathrm{p}(\zeta_A \mid Z) = \frac{1}{\sqrt{2\pi}^N \sigma_A^N} e^{\frac{1}{2\pi\sigma_A^2}\left(\sum_{i=1}^{N} (\zeta_{A_i} - z_i)^2\right)}$$

mit der Varianz σ_O, die analog wie σ_A zu behandeln ist.

Die Schätzung der gesuchten Größe mit ihren Komponenten z_i wird durch die Kombination der beiden unterschiedlichen Messungen mit einer Bayesschen Schätzung und dem Maximum–a–posteriori–Kriterium realisiert. Die a–posteriori–Dichtefunktion ergibt sich in diesem Fall zu

$$\mathrm{p}(Z \mid \zeta_A, \zeta_O) = \frac{\mathrm{p}(\zeta_A, \zeta_O \mid Z)\pi(Z)}{\mathrm{p}(\zeta_A, \zeta_O)} = \frac{\mathrm{p}(\zeta_A \mid Z)\,\mathrm{p}(\zeta_O \mid Z)\pi(Z)}{\mathrm{p}(\zeta_A, \zeta_O)}$$

Diese Funktion gilt es nun, in Abhängigkeit von Z zu maximieren. Wie in 2.5.1 beschrieben, wird im Falle von (hier vorliegenden) Normalverteilungen die Gleichung auf beiden Seiten logarithmiert und die so entstehende Gleichung maximiert (durch Gradientenbildung nach Z und Nullsetzen). Es ergibt sich hier unter Vernachlässigung aller für den Maximierungsprozeß durch Variation von Z unerheblichen Konstanten:

$$\begin{aligned}\log(\mathrm{p}(Z \mid \zeta_A, \zeta_O)) &= \mathrm{l}(Z \mid \zeta_A, \zeta_O) \\ &= \mathrm{l}(\zeta_A \mid Z) + \mathrm{l}(\zeta_O \mid Z) + \mathrm{l}(Z) \\ &= -\frac{1}{\sigma_A^2}\left(\sum_{i=1}^{N}(\zeta_{A_i} - z_i)^2\right) - \frac{1}{\sigma_O^2}\left(\sum_{i=1}^{N}(\zeta_{O_i} - z_i)^2\right) + \mathrm{l}(Z)\end{aligned}$$

wo $\mathrm{l}(\cdot)$ den Logarithmus der Dichtefunktion $\log(\mathrm{p}(\cdot))$ bezeichnet. Die a–priori–Dichte $\mathrm{l}(Z)$ läßt sich nur bei Vorliegen eines ausreichend guten Objektmodells oder eines „Objektprototypen" bestimmen, hängt also vom Einzelfall ab. Typischerweise wird für ein gegebenes Objekt der Mittelwert und die Varianz (bzw. die Kovarianzmatrix) berechnet. Damit kann dann eine Normalverteilung auch für die a–priori–Verteilung angesetzt werden. Wenn kein Vorwissen über das Objekt existiert, sollte die a–priori–Varianz zu Unendlich angenommen werden. In diesem Fall fällt der rechts stehende Summand in der obigen Gleichung weg und die Schätzung entspricht einer Maximum–Likelihood–Schätzung. Man erhält dann aus Gradientenbildung und Nullsetzen für die einzelnen Segmente für z_i:

$$0 = \frac{\partial \log(\mathrm{p}(Z \mid \zeta_A, \zeta_O))}{\partial z_i} = \frac{2}{\sigma_A^2}(\zeta_{A_i} - z_i) + \frac{2}{\sigma_O^2}(\zeta_{O_i} - z_i)$$

Daraus folgt sofort für den gesuchten Schätzwert:

$$z_i = \frac{\sigma_A^{-2}\zeta_{A_i} + \sigma_O^{-2}\zeta_{O_i}}{\sigma_A^{-2} + \sigma_O^{-2}}$$

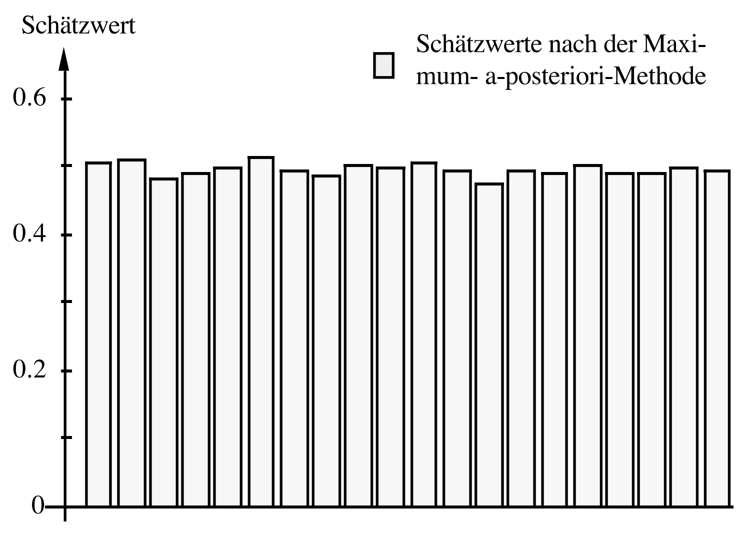

Bild 8: Zwanzig Schätzwerte bestimmt durch die Maximum-a-posteriori-Methode bei jeweils neun konsistenten und einem einen inkonsistenten Sensorwert

gemäß N~(2, 0.02) bei jeder Stichprobe.

Haben beide Sensoren unterschiedliche Auflösungen (d.h. unterschiedlicher Rasterbreiten in *x*–Richtung), so kann das bei diesem Ansatz dadurch Berücksichtigung finden, daß das Meßergebnis des schlechter auflösenden Sensors jeweils für alle dieses Intervall überstreichenden feineren Intervalle des besser auflösenden Sensors[1] konstant gehalten wird.

Bei dem hier betrachteten Vorgehen wird die Leistungsfähigkeit gegenüber einem einzelnen Sensor durch die Fusionsmethode verbessert. Beim gewählten Sensormodell wird das Sensorrauschen einbezogen und damit auch bei der Berechnung der Schätzwerte mit berücksichtigt. Die Bayessche Schätzung liefert einen Schätzwert, der mit geringerer Unsicherheit versehen ist als die einzelnen Sensorwerte für sich genommen. Soll die Methode auf konkret vorliegende Sensoren angewendet werden, müssen für jeden Sensor das typische Sensorrauschen und die a–priori–Informationen ermittelt werden, unter Umständen auch die Transformationen der Sensordaten f(θ). Dies kann mit einer Vielzahl von Durchführungen des gleichen Experiments geschehen. Die Fusion kann allerdings infolge des gewählten Sensormodells bei inkonsistenten Datenmengen oder Ausreißern zu schlechten Schätzwerten führen. Ein Beispiel dafür zeigt Bild 8. Es wurden wieder zwanzig verschiedene Schätzungen mit jeweils zehn Sensorwerten vorgenommen. Neun Sensorwerte wurden zufällig durch Normalverteilungen mit Mittelwert Null und Varianzen (0.02, 0.3, 0.03, 0.06, 0.06, 0.09, 0.01, 0.01, 0.02) generiert. Der jeweils zehnte Wert soll jedoch einen Ausreißer darstellen und wurde deshalb aus einer Normalverteilung N~(2, 0.02) gewonnen. Die Schätzwerte wurden durch eine Maximum-a-posteriori-Schätzung mit a–priori Information N~(0.009, 0.09) berechnet. Man erkennt, daß die Schätzwerte ganz erheblich vom wahren Wert abweichen. Normalerweise sind solche Ausreißer unwahrscheinlich, wenn nur das normale Sensorrauschen einbezogen wird. Algorithmische Fehler (Softwarefehler) und Hardwarefehler können jedoch zu völlig unsinnigen Werten führen. Die Anwendung des Verfahrens sollte deshalb nur mit einer geeigneten Konsistenzüberprüfung stattfinden. Für diese Prüfung bietet sich das im folgenden Abschnitt vorgestellte Kriterium an.

1.4.2 Konsistenzüberprüfung und die Methode von Durrant–Whyte

Basierend auf verschiedenen unvollständigen und unsicheren Sensordaten soll nun eine *robuste* Schätzung der Sensorumwelt vorgenommen werden. Die Methode von Durrant–Whyte zur Sensordatenfusion [Durrant–Whyte 85; Allen 87; Durrant–Whyte 88], deren integraler Bestandteil eine Konsistenzüberprüfung ist, besteht aus zwei Teilen: der Fusion und der Verbreitung der neugewonnenen Information im Weltmodell. Bei der Fusion werden die unsicheren und unterschiedlichen Zustandsbeobachtungen eines Objekts (d.h. eines Auschnittes aus dem aktuellen Umweltmodell) zu einer verbesserten Sicht des betrachteten Objekts zusammengefügt. Bei der Verbreitung werden die gewonnenen Beobachtungen unter Wahrung der Konsistenz in das Weltmodell integriert (und damit jedem Sensor als a–priori–Information zur Verfügung gestellt). Zur Repräsentation der Umwelt wird ein Modell verwendet, wie am Ende des Abschnitts 1.2 angegeben. Einzelne geometrische Merkmale im n-dimensionalen euklidischen Raum $\mathbf{R}^n$ werden durch einen Parametervektor θ mit $\theta \in \mathbf{R}^m$ und eine mehrdimensionale Funktion $g(y, \theta) = 0$ mit $y \in \mathbf{R}^n$ dargestellt. Jede Funktion g beschreibt ein spe-

1 In diesem Fall muß allerdings die Kovarianzmatrix für den schlechter auflösenden Sensor aufgestellt werden, denn die Meßergebnisse sind dann innerhalb des feineren Intervalls jeweils gleich und daher statistisch nicht mehr unabhängig voneinander.

zielles geometrisches Merkmal und jeder Vektor θ modelliert eine bestimmte Instanz dieses geometrischen Merkmals. Jedem Parametervektor θ wird eine Dichtefunktion $p(\theta)$ zugeordnet, die ein Maß für die Wahrscheinlichkeit dafür ist, daß eine bestimmte Instanz des Merkmals beobachtet wurde. Der Vorteil einer solchen „gekoppelten“ Unsicherheitsrepräsentation, die die Geometrie der Umwelt direkt einbezieht, liegt in der einfachen Transformationsmöglichkeit zwischen unterschiedlichen Sensorkoordinatensystemen.

Die Datenfusion baut auf einer Beobachtungssequenz $x_1, x_2, \ldots x_N$ des statischen Weltzustandes auf. Die einzelnen Beobachtungsvektoren x_i werden, wie üblich, als unabhängig angenommen, da nur unter dieser Voraussetzung Berechnungen in vernünftigem Zeitaufwand möglich sind. Unter der Annahme, daß die N verschiedenen Dichtefunktionen $p(x_k|\theta)$ gemäß $N\sim(\theta, \Sigma_k)$ verteilt sind und a–priori–Information $N\sim(\theta_0, \Sigma_0)$ aufgrund früherer Beobachtungen vorliegt, wird als Schätzmethode das diskrete, statische Kalman–Filter (siehe 2.5.5, Beispiel 1) benutzt, also eine Schätzung mit Hilfe der Beziehung

$$\hat{\theta} = \hat{\mu} = \left(\Sigma_0^{-1} + \sum_{k=1}^{N} \Sigma_k^{-1} \right)^{-1} \left(\sum_{k=1}^{N} \Sigma_k^{-1} x_k + \Sigma_0^{-1} \theta_0 \right)$$

für den Erwartungswert und

$$\Sigma = \left(\Sigma_0^{-1} + \sum_{k=1}^{N} \Sigma_k^{-1} \right)^{-1}$$

für die Kovarianzmatrix der normalverteilten a–posteriori Verbundverteilung $p(x_1, x_2, \ldots, x_N)$ vorgenommen. Diese Formel setzt allerdings voraus, daß die N Beobachtungen eine normalverteilte Dichte haben und deshalb kann das Filter nur verwendet werden, wenn zuvor Ausreißer eliminiert wurden. Vor Verwendung dieses Schätzers für den wahren Zustand wird eine Gruppenbildung vorgenommen, die Ausreißer eliminiert. Die in der Gruppe nach Elimination der Ausreißer verbleibenden Beobachtungen können dann als von einer nicht verunreinigten Gaußverteilung abgeleitet verstanden werden und der wahre Zustand mit Hilfe der oben stehenden Beziehung geschätzt werden. In [Durrant–Whyte 87; Shen 90] wird als Kriterium für die Entscheidung, ob eine Sensorbeobachtung noch in eine Gruppe um einen wahren Zustand θ gehört, ein Mahalanobis–Abstand (siehe 2.2) von $r \leq 1$ gewählt. Für eine Beobachtung x_k muß also gelten:

$$(\theta - x_k)^T \Sigma_k^{-1} (\theta - x_k) \leq 1$$

wenn sie zur Gruppe dazugehören soll[1]. Anders ausgedrückt: Wenn aus N Beobachtungen eine Gruppe gebildet werden soll, muß ein Zustand θ gefunden werden, so daß für alle x_k die obige Ungleichung gilt. Dies läßt sich auch so formulieren, daß ein Zustand gefunden wird, bei dem für zwei Beobachtungen x_i und x_j aus der Gruppe gilt:

---

[1] Das gleiche Kriterium läßt sich herleiten, wenn man für die Nutzenfunktion (negative Verlustfunktion, siehe Kapitel 2) für die Zugehörigkeit einer Beobachtung zu einer Gruppe die (mit x_k und Σ_k normalverteilte) Likelihood–Funktion $f(x_k|\theta)$ annimmt.

$$\frac{1}{2}(\theta - x_k)^T \Sigma_k^{-1} (\theta - x_k) + \frac{1}{2}(\theta - x_j)^T \Sigma_j^{-1} (\theta - x_j) \leq 1$$

Wenn diese Bedingung erfüllt ist, dann bilden die zwei Beobachtungen einen *Konsens*. Der Wert für θ, der diese Beziehung minimiert, läßt sich durch Gradientenbildung und anschließendes Nullsetzen finden und ist gerade der obenstehende Schätzwert $\hat{\mu}$ (bei zwei Beobachtungen):

$$\theta = \left(\Sigma_k^{-1} + \Sigma_j^{-1}\right)^{-1} \left(\Sigma_k^{-1} x_k + \Sigma_j^{-1} x_j\right)$$

Diesen Schätzwert kann man schließlich in die Bedingung für die Gruppenbildung einsetzen und erhält sofort als Bedingung für die Zugehörigkeit zu einer Gruppe:

$$\frac{1}{2}(x_k - x_j)^T \left(\Sigma_k^{-1} + \Sigma_j^{-1}\right)^{-1} (x_k - x_j) \leq 1$$

Wenn durch Vergleich aller $N(N-1)/2$ Paare einer Beobachtung jeweils diese Bedingung erfüllt ist, gehören alle Elemente zu der Gruppe. Wenn für eine Paarung die Gleichung nicht erfüllt ist, müssen beide Beobachtungen als mögliche Ausreißer angesehen werden. Der wahre Ausreißer kann dann durch weitergehenden Vergleich ermittelt werden. Die größeren auf diese Weise ermittelten Gruppen können dazu verwendet werden, das Weltmodell zu aktualisieren, während die restlichen Werte außerhalb der Gruppen als Ausreißer ignoriert werden. Diese Konsistenzüberprüfung hat gegenüber den sonst üblichen Konsistenzüberprüfungen den wesentlichen Vorteil, keinen *Schwellwert* zu benötigen.

Bei der Aktualisierung des Weltmodells (Ausbreitungsprozeß) werden die gewonnenen Beobachtungen unter Konsistenzbeibehaltung im Weltmodell verbreitet. Die aus einer neuen Schätzung ($\hat{\theta}$, Σ) resultierende Änderung wird jedoch nicht direkt in das Weltmodell übernommen, sondern zur alten in Beziehung gesetzt. Bezeichne $\hat{\theta}'$ den neuen Schätzwert für den Merkmalsvektor und $\hat{\theta}$ den alten sowie Σ' die neue Kovarianzmatrix, dann wird die Änderung in das Weltmodell übernommen, die die Verlustfunktion $L = (\hat{\theta}' - \hat{\theta})^T \Sigma'^{-1} (\hat{\theta}' - \hat{\theta})$ minimiert. Dies vermindert die Auswirkungen von schlechten Messungen (mit großer Varianz).

Wir untersuchen nun die Wirksamkeit der Konsistenzprüfung anhand einiger Simulationen und betrachten zunächst die Leistungsfähigkeit der Schätzung mit dem statischen Kalman–Filter. Es wurden zwanzig verschiedene Stichproben mit jeweils zehn Sensorwerten zufällig durch Normalverteilungen mit den Varianzen (0.02, 0.3, 0.03, 0.09, 0.09, 0.06, 0.06, 0.01, 0.02, 0.01) ermittelt. Als a–priori–Information wurde ein Mittelwert 0.009 und eine Varianz von 0.09 gewählt. Werden alle zehn Sensordaten aus einer Normalverteilung mit Mittelwert Null und den entsprechenden Varianzen erzeugt, zeigt sich kein Unterschied zwischen den Ergebnissen mit und ohne Konsistenzüberprüfung. Dies ist nicht weiter überraschend und bedeutet, daß die Konsistenzüberprüfung Daten als konsistent erkennt, die im Bereich des Sensorrauschens liegen. Ein solches Beispiel zeigt Bild 9.

Bild 10a zeigt die Ergebnisse von zwanzig Schätzungen mit und ohne Konsistenzüberprüfung. Wieder wurden zwanzig Stichproben zu je zehn Daten erzeugt, wobei die ersten neun Werte aus Normalverteilungen mit Mittelwert Null und der zehnte Wert als Ausreißer mit Mittelwert 0.14

gewonnen wurden. Da der Mittelwert dieses Ausreißers sehr nahe an dem Mittelwert der anderen Sensordaten liegt, kann bei der Konsistenzüberprüfung nicht jeder eigentlich inkonsistente Wert gefunden werden. Die Schätzwerte, die nach einer Konsistenzüberprüfung berechnet wurden, sind bis auf die sechste Stichprobe genauer oder gleich den Schätzwerten ohne Konsistenzüberprüfung. Bei dieser erwähnten Stichprobe lagen die neun „richtigen" Sensordaten zufällig so ungünstig, daß bei der Schätzwertberechnung ohne Konsistenzüberprüfung aufgrund der sich gegenseitig aufhebenden negativen und positiven Abweichungen ein besserer Schätzwert als mit Konsistenzüberprüfung (bei der der inkonsistente Wert entdeckt wurde), ermittelt wurde. Solche ungünstigen Sensordaten wirken sich besonders bei einer kleinen Sensoranzahl stärker aus. Wird statt eines Sensors mit kleiner Varianz der Sensor mit der größten Varianz als inkonsistenter Wert gewählt, werden bei einer Konsistenzüberprüfung erst für einen weiter entfernt gewählten Mittelwert dieses Ausreißers die Daten als inkonsistent erkannt. Das Verhalten in einer solchen Situation ist in Bild 10b angedeutet.

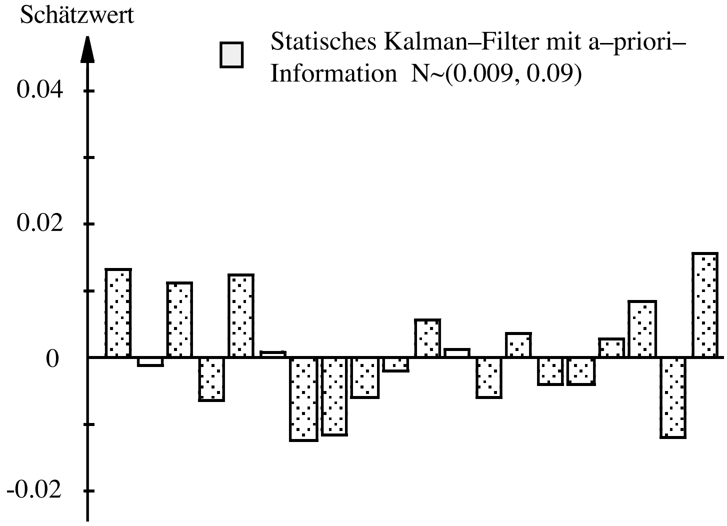

Bild 9: Statisches Kalman–Filter bei konsistenten Daten mit a–priori–Information N~(0.009, 0.09), bei wahrem Zustand Null

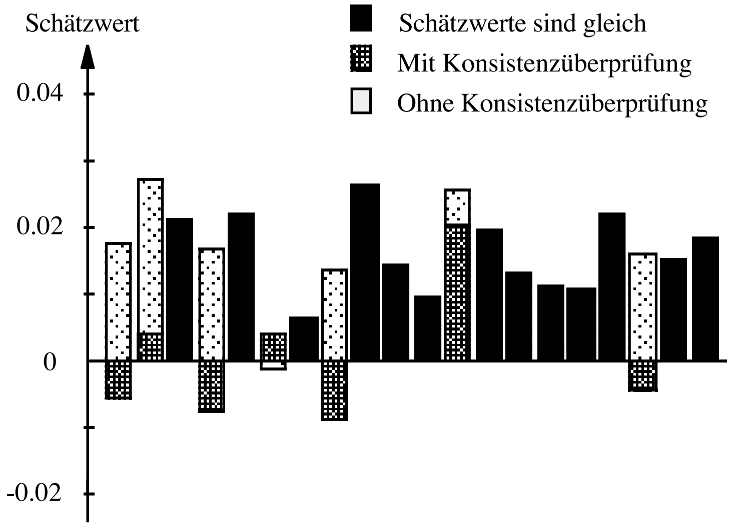

Bild 10a: Schätzwerte mit und ohne Konsistenzüberprüfung bei einem inkonsistenten Sensorwert N~(0.14, 0.01)

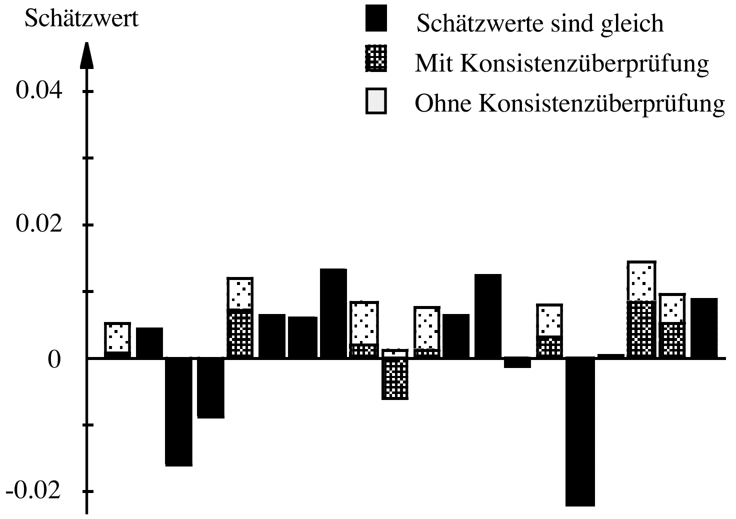

Bild 10b: Schätzwerte mit und ohne Konsistenzüberprüfung bei einem inkonsistenten Sensorwert N~(0.3, 0.3)

Hier wurde als Ausreißer der Sensorwert mit der größten Varianz (also 0.3) mit Mittelwert 0.3 gewählt. Auch unter diesen zwanzig Stichproben war eine ungünstig, nämlich die zehnte. Aufgrund der großen Varianz des Sensors können trotz des weiter entfernten Mittelwertes nicht alle ursprünglichen Ausreißer gefunden werden. Ist hingegen ein inkonsistenter Wert mit Mittelwert 0.3, aber kleiner Varianz 0.01 vorhanden, findet die Konsistenzüberprüfung alle Ausreißer und ist für die 20 zufälligen Stichproben in Bild 11a (gegenüber Bild 10 größere Maßstabseinheit) in allen Fällen deutlich besser. Besonders bei mehreren Ausreißern oder bei einem sehr weit entfernten Mittelwert des Ausreißers kann es ohne Konsistenzüberprüfung zu sehr großen Abweichungen kommen. Ein Beispiel für vier Ausreißer mit Mittelwert 0.3 und den Varianzen (0.3, 0.06, 0.01, 0.02) zeigt Bild 11b. Man erkennt, daß in diesem Fall mit der Methode der Konsistenzüberprüfung eine ganz erhebliche Verbesserung des Schätzwertes durch die Elimination der Ausreißer erfolgt.

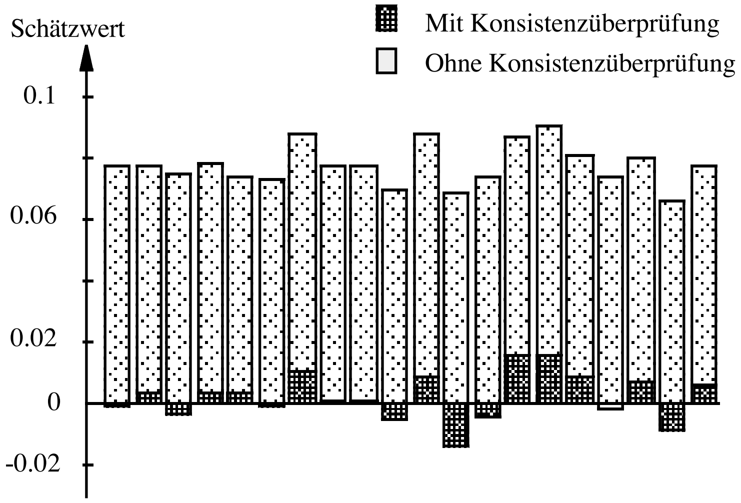

Bild 11a: Schätzwerte mit und ohne Konsistenzüberprüfung bei einem inkonsistenten Sensorwert N~(0.3, 0.01)

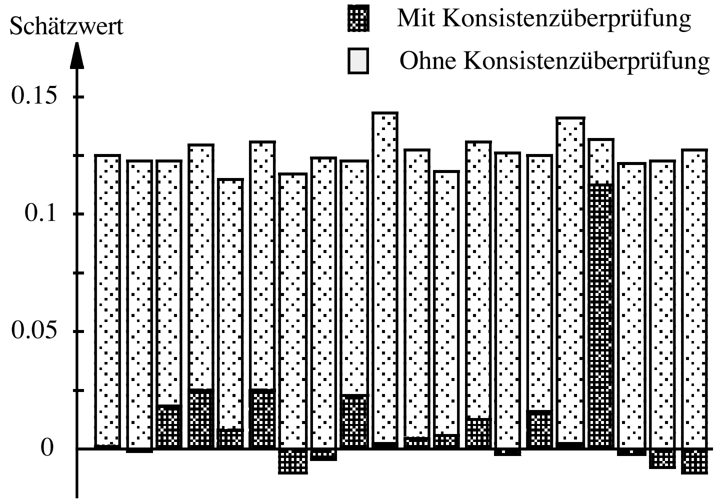

Bild 11b: Schätzwerte mit und ohne Konsistenzüberprüfung mit vier inkonsistenten Sensorwerten N~(0.3, 0.3), N~(0.3, 0.06), N~(0.3, 0.01) und N~(0.3, 0.02)

Zusammenfassend kann festgehalten werden, daß die Konsistenzüberprüfung zu einer wesentlichen Verbesserung des Schätzverhaltens führt, wobei allerdings der Typ der Sensoren auf solche Sensoren eingeschränkt ist, die geometrische Informationen liefern können; Temperaturinformation kann in diesem Weltmodell beispielsweise nicht repräsentiert werden. Bei Verwendung des Kalman–Filters ergeben sich dieselben Probleme bei der Bereitstellung von a–priori–Information wie bei der im vorigen Abschnitt behandelten Bayes–Schätzung. Soll die Schätzmethode auf konkret vorliegende Sensoren angewendet werden, muß zunächst eine Vorverarbeitung durchgeführt werden, bei der die eigentlichen Sensordaten in eine entsprechende geometrische Form transformiert werden (samt zugehörigen Dichtefunktionen). Weiterhin müssen in geeigneter Weise gute a–priori–Informationen und die Varianzen der Sensoren bestimmt werden. Für jede neue Umgebung muß ein neues a–priori–Weltmodell ermittelt werden. Das Verfahren sollte bezüglich der Verwendung allgemeinerer Sensordaten und einer Fusionsmethode möglichst ohne a–priori–Information weiterentwickelt werden. Es ist zu untersuchen, inwieweit die verunreinigte Gaußverteilung für den Störprozeß im Beobachtungsmodell des Kalman–Filters angesetzt werden kann. Für die Umweltmodellierung wäre eine Kombination aus dem topologischen Netz und Datenrahmen zu prüfen. Bei einer solchen Kombination könnte das topologische Netz weiterhin zur Verbesserung der Schätzwerte beitragen, andererseits könnten aber auch alle numerischen Sensordaten mit in dieses Weltmodell integriert werden.

1.4.3 Eine weitere Methode zur Konsistenzüberprüfung

Ein anderes Konzept zur Konsistenzprüfung wurde in [Luo 87; Luo 87b; Luo 88; Luo88b; Luo 90] vorgestellt. Es beruht, im Gegensatz zu dem im vorigen Abschnitt benutzten Vergleich von Meßwertabständen unter Benutzung von Varianzen, auf dem Vergleich von vollständigen Dichtefunktionen, wie diese auch bei bestimmten Entscheidungsproblemen vorgenommen wird (siehe [Duda 73])[1]. Die Sensoren i und j, deren Konsistenz zu überprüfen sei, werden durch skalare Variable $X_i = \theta + V_i$ und $X_j = \theta + V_j$ modelliert, wo V_i und V_j unabhängige Zufallsvariable sind. Diese Zufallsvariablen haben die zum Nullpunkt symmetrischen Dichtefunktionen p_i und p_j. Aufgrund der Messungen werde nun für den wahren Zustand θ jeweils eine Schätzung $\hat{\theta}_i$ bzw. eine Schätzung $\hat{\theta}_j$ vorgenommen, so daß die Dichtefunktionen um diese Schätzwerte zentriert sind. Der Konfidenzabstand d_{ij} zwischen zwei Meßwerten x_i und x_j wird durch die doppelte Fläche definiert, die unter der Dichtefunktion $\mathrm{p}_i(x - \hat{\theta}_i)$ zwischen dem Schätzwert $\hat{\theta}_i$ und x_j aufgespannt wird, also:

$$d_{ij} = 2\left|\int_{\hat{\theta}_i}^{x_j} \mathrm{p}_i(x - \hat{\theta}_i)\,\mathrm{d}x\right| = 2\left|\int_0^{x_j - \hat{\theta}_i} \mathrm{p}_i(x)\,\mathrm{d}x\right|$$

Wenn man den Meßwert zur Schätzung benutzt, ergibt sich:

$$d_{ij} = 2\left|\int_{x_i}^{x_j} \mathrm{p}_i(x - x_i)\,\mathrm{d}x\right| = 2\left|\int_0^{x_j - x_i} \mathrm{p}_i(x)\,\mathrm{d}x\right|$$

Geht man umgekehrt von $\hat{\theta}_i$ aus, so wird

$$d_{ji} = 2\left|\int_{x_j}^{x_i} \mathrm{p}_j(x - x_j)\,\mathrm{d}x\right| = 2\left|\int_0^{x_i - x_j} \mathrm{p}_j(x)\,\mathrm{d}x\right|$$

Bild 12 zeigt ein Beispiel einer solchen Konfidenzabstandsmessung für zwei unterschiedliche normalverteilte Dichtefunktionen p_i und p_j für jeden Sensor. Sei x_i der wahre Wert und zum Beispiel $d_{ij} = 0.6$, dann liegt x_j im 60%–Konfidenzintervall der Dichtefunktion $\mathrm{p}_i(x)$. Im Falle von normalverteilten Dichten kann der Wert für $d_{ij} = \mathrm{F}_i(x_i) - \mathrm{F}_i(x_j)$ direkt aus Tabellen abgelesen werden. Man sieht sofort, daß im Normalfall $d_{ij} \neq d_{ji}$, wobei hier zu untersuchen wäre, inwieweit sich der Ansatz auf vektorielle Messungen ausdehnen ließe.

---

[1] Das in [Luo 90] verwendete Abstandsmaß ist in seiner korrekten Form identisch mit dem Kolmogorow–Abstand zweier bedingter Verteilungen: $D = \max(|F_X(x|C_1) - F_X(x|C_2)|)$. Dies entspricht der gemeinsamen Fläche unter den zugehörigen bedingten Dichtefunktionen (siehe dazu [Flachs 88; Beer 89]).

Mit diesem nichtkommutativen Abstandsmaß läßt sich eine *Abstandsmatrix* aufbauen, die ausdrückt, ob sich Sensoren gegenseitig unterstützen, ob keine oder eine einseitige Unterstützung zu beobachten ist. Das System soll aus N Sensoren, die dieselbe Objekteigenschaft k messen, bestehen. Für diese k-te Objekteigenschaft werden die Konfidenzabstände in Matrixform gebracht, $D_k = [d_{ij}]$ für $i, j = 1, \ldots, N$. In einem zweiten Schritt wird aus der Abstandsmatrix die $N \times N$–*Beziehungsmatrix* $R_k = [r_{ij}]$, mit $i, j = 1, \ldots$, N für die k-te Objekteigenschaft gewonnen. Dazu wird für jedes Element der Matrix D_k eine Schwellwertbildung durchgeführt. Liegt der in die Abstandsmatrix eingetragene Konfidenzabstand für die Messungen zweier Sensoren unterhalb einer gewissen Schwelle, wird an das entsprechende Element der Beziehungsmatrix R_k eine Eins eingetragen, liegt der Wert darüber (unterstützen sich die Meßergebnisse also nicht), wird eine Null eingetragen.

In der Festlegung dieses Schwellwertes liegt natürlich das Hauptproblem für diesen Ansatz, denn dieser Wert kann nur empirisch gewonnen werden. Die Beziehungsmatrix kann anschaulich durch einen gerichteten Graphen repräsentiert werden. Die Sensoren bilden die Knoten, und von Knoten i nach Knoten j befindet sich eine gerichtete Kante, wenn $r_{ij} = 1$ ist. Dabei können zwei Knoten von zwei, von einer oder von keiner Kante verbunden sein.

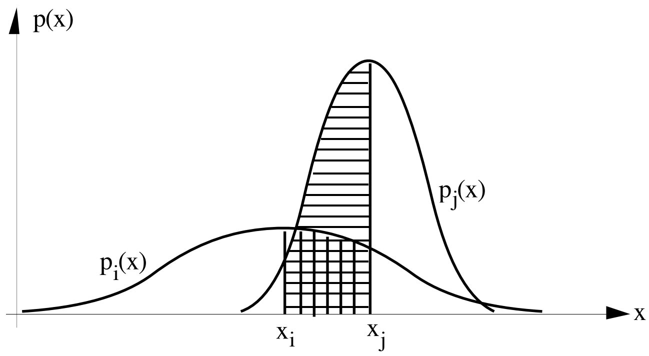

Bild 12: Konfidenzabstandsmessung d_{ij}. Senkrecht schraffierter Bereich: d_{ij}, waagrecht schraffierter Bereich: d_{ji}.

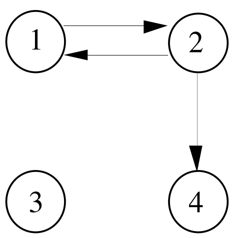

Bild 13: Beispiel für einen Beziehungsgraphen (Erklärung siehe Text)

Wurde beispielsweise aufgrund von vier Sensormessungen die 4×4–Beziehungsmatrix

$$R_k = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

gewonnen, kann diese Matrix durch den in Bild 13 dargestellten *Beziehungsgraphen* repräsentiert werden. Sind zwei Kanten zwischen zwei Knoten vorhanden, unterstützen sich die Knoten gegenseitig. Ist demgegenüber ein Knoten mit keinem anderen Knoten verbunden (wird er also durch keinen anderen Sensor unterstützt), muß angenommen werden, daß die Daten dieses Sensors falsch sind.

Zur Gruppenbildung wird im Beziehungsgraphen der größte Teilgraph der sich gegenseitig unterstützenden Knoten ermittelt. Er repräsentiert die Gruppe der Sensoren, die mit größter Wahrscheinlichkeit die beste Repräsentation des Objektes liefern. Die Knoten dieser Gruppe werden durch eine geeignete Datenfusionsmethode zu einem sogenannten *Superknoten* des Beziehungsgraphen verschmolzen. Falls keiner der Sensoren den anderen bestätigt, wird für weitere Betrachtungen der Sensorwert mit der kleineren Varianz gewählt. Bestätigt Sensor 1 Sensor 2 einseitig, wird der Wert von Sensor 2 als Fusionswert gewählt. Bestätigen sich die beiden Sensoren gegenseitig, wird der Fusionswert bestimmt. Im Gegensatz zur Methode aus dem vorigen Abschnitt führt die Konsistenzüberprüfung hier nicht nur zur Aussonderung fehlerhafter (inkonsistenter) Sensordaten, sondern darüber hinaus zur Feststellung von Unterstützungsbeziehungen zwischen den einzelnen Sensordaten. Die Überprüfung ist genau, aber aufgrund der Benutzung eines a–priori festzulegenden Schwellwertes in der Praxis nur schwierig durchzuführen. Außerdem ist der Aufwand zur Cliquenbestimmung in Graphen im allgemeinen hoch.

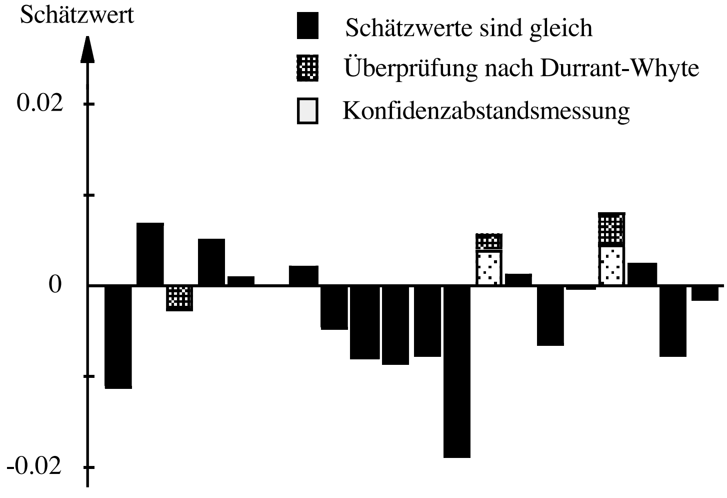

Bild 14a: Schätzwerte nach Konsistenzüberprüfung. Überprüfung nach der Methode von Durrant–Whyte und mit der Methode der Konfidenzabstandsmessung mit Schwellwert 0.55

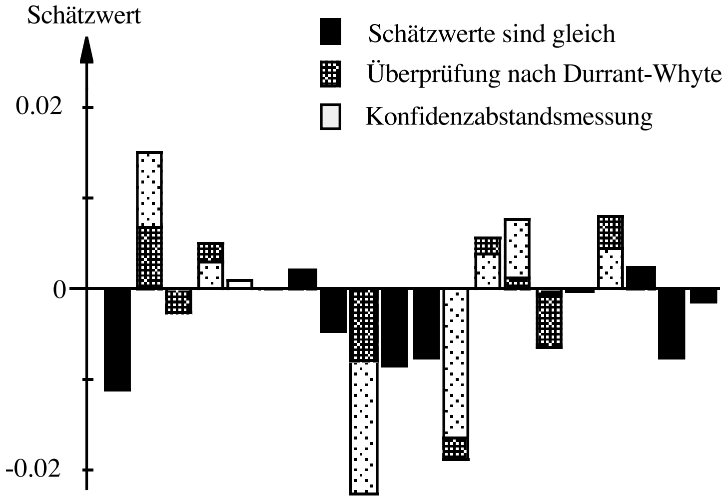

Bild 14b: Wie Bild 14a, aber Konfidenzabstandsmessung mit Schwellwert 0.3

Wir vergleichen diese Methode zur Überprüfung nun mit der aus dem vorigen Abschnitt, indem beide die bereits in den vorigen Abschnitten benutzten Testdaten bearbeiten. Bild 14 zeigt zunächst ein Beispiel für konsistente Sensorwerte. Es wurden wieder zufällig jeweils 10 Sensorwerte durch Normalverteilungen mit Mittelwert Null und Varianzen (0.02, 0.3, 0.03, 0.09, 0.09, 0.06, 0.06, 0.01, 0.02, 0.01) ermittelt. Die Schätzwerte wurden einerseits mit der Methode von Durrant-Whyte (statisches Kalman–Filter) und ohne a–priori Informationen (aber mit Konsistenzüberprüfung) und durch die in [Luo 88b] vorgeschlagene Maximum–Likelihood–Schätzung mit Konfidenzabstandsmessung bei einem Schwellwert von 0.55 berechnet. Die Konsistenzüberprüfung nach Durrant-Whyte fand bei keiner Stichprobe Ausreißer. Die Konfidenzabstandsmessung ermittelte bei drei Stichproben Ausreißer und führte bei diesen tatsächlich zu einem besseren Schätzwert. Wie sich auch hier im Experiment zeigte, ist die Bestimmung eines geeigneten Schwellwertes allerdings sehr problematisch. Wird der Schwellwert zu niedrig gewählt, werden auch Werte eliminiert, die noch im Bereich des Sensorrauschens liegen. Besonders bei kleiner Sensoranzahl kann es dann zu schlechteren Schätzwerten kommen. Dieses Verhalten zeigt Bild 14b.

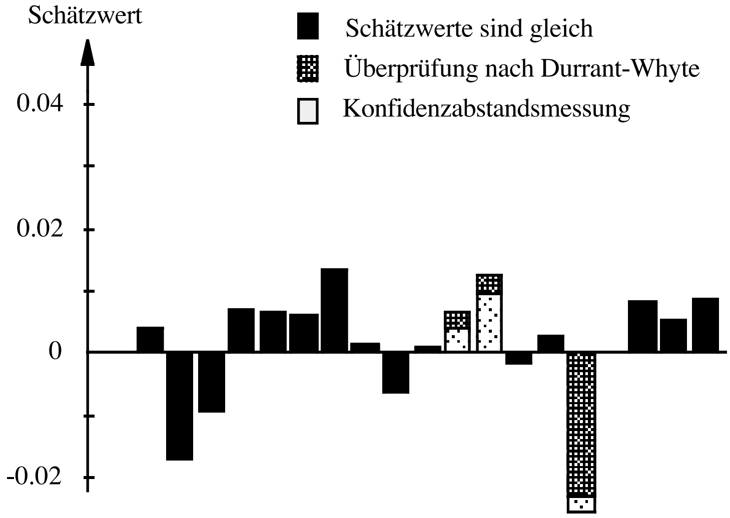

Bild 15a: Schätzwerte mit Konsistenzüberprüfung bei einem Ausreißer N~(0.3, 0.3), ermittelt durch die Methode von Durrant-Whyte und mit Konfidenzabstandsmessung mit Schwellwert 0.55

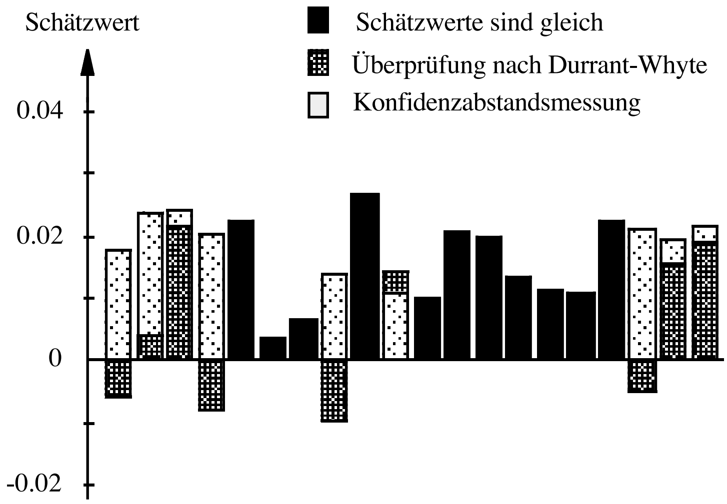

Bild 15b: Schätzwerte mit Konsistenzüberprüfung bei einem Ausreißer N~(0.14, 0.01), ermittelt durch die Methode von Durrant-Whyte und Konfidenzabstandsmessung mit Schwellwert 0.55

Werden die Schätzwerte für die Stichproben aus Bild 14b mit der Konfidenzabstandsmessung bei einem Schwellwert von 0.3 berechnet, werden zwar die Schätzwerte für zwei weitere Stichproben besser, aber andererseits auch die Schätzwerte für drei Stichproben deutlich verschlechtert. Wird demgegenüber ein Schwellwert von 0.8 gewählt, stimmen die beiden Konsistenzüberprüfungen für die Beispiele miteinander überein. Verwendet man die Konsistenzüberprüfung mit Schwellwert 0.55 für die bereits bei Durrant-Whyte untersuchten inkonsistenten Stichproben, zeigt sich, daß dieser für den allgemeinen Fall offensichtlich immer noch zu niedrig gewählt wurde. Für einen Ausreißer mit N~(0.3, 0.3) in Bild 15a erhält man neben einem schlechteren Wert immerhin zwei bessere Werte als der Methode nach Durrant–Whyte. Dagegen kommt es für einen Ausreißer N~(0.14, 0.01) nur bei einem Schätzwert zu einem besseren Ergebnis und bei sieben zu einem schlechteren (Bild 15b). Eine optimale Schwellwertbestimmung, so daß die Schätzwerte besser als die Schätzwerte nach der

Methode von Durrant-Whyte werden, ist für den allgemeinen Fall offensichtlich sehr schwierig. Zusammenfassend läßt sich festhalten, daß diese Konsistenzüberprüfung bei gut gewählten Schwellwerten zu genauen Schätzwerten führen. Leider ist eine sehr genaue Bestimmung dieses Schwellwertes notwendig. An diesem Punkt hätten Weiterentwicklungen auch anzusetzen.

Auf Methoden der statistischen Entscheidungstheorie fußende Verfahren zum Konsistenztest werden in [Kamberova 89; McKendall 90] entwickelt. Sie stützen sich auf Einheitsverlustfunktionen (Abschnitt 2.3) und robusten Schätzern [Huber 81; Berger 88] ab. Diese Methoden sind mathematisch fundiert und setzen im Unterschied zu den oben dargestellten keine speziellen Fehlerverteilungen voraus, allerdings werden keine experimentellen Ergebnisse für ihre Leistungsfähigkeit angegeben. Es bleibt festzuhalten, daß Konfidenzüberprüfungen ein noch weitgehend unbearbeitetes, allerdings sehr lohnenswertes Feld für weitere Forschungen darstellen.

1.4.4 Fusion mit inverser Sensortransformation

Wie in Abschnitt 1.3 bereits erwähnt, besteht bei der Sensormodellierung auch die Möglichkeit, den Sensor durch eine Funktion $\theta = \mathrm{f}(X)$ zu modellieren. Dabei ist θ der gesuchte Umweltzustand und X sind in diesem Fall beobachtbare Eingangsdaten des Auswertungsalgorithmus des Sensors. Betrachtet wird hier nicht der gesamte Sensor, sondern vielmehr nur seine höheren Auswertungsstufen. Zur Erklärung sei an Bild 3 erinnert. Eine solche Modellierung wird z.B. in [Matthies 87] für ein Stereokamerasystem verwendet und in [Nakamura 89] allgemeiner betrachtet.

Wir gehen wieder für jeden Sensor k ($k = 1\ldots N$) von einem (Zwischen–)Meßwert X_k aus, der durch eine additive Rauschgröße Δ_k gestört ist, also $X_k = \overline{X}_k + \Delta_k$, wobei Δ_k gemäß N~(0, Σ_{Δ_k}) verteilt ist. Mit einer nichtlinearen Funktion f(·) und der Jacobi–Matrix J der nichtlinearen Funktion ist der vom Sensor ermittelte Zustand θ_k (mit der Dimension M) näherungsweise

$$\theta_k \approx \mathrm{f}_k(\overline{X}_k) + J_{X_k}(\overline{X}_k)\Delta_k$$

wobei $\mathrm{E}[\theta_k] = \mathrm{f}(\overline{X}_k) = \theta$ und die Kovarianz des Vektors θ_k

$$\begin{aligned}
\Sigma_{\theta_k} &= \mathrm{E}\left[(\theta_k - \mathrm{E}[\theta_k])(\theta_k - \mathrm{E}[\theta_k])^T\right] \\
&= \mathrm{E}\left[(J_{X_k}(\overline{X}_k)\Delta_k)(J_{X_k}(\overline{X}_k)\Delta_k)^T\right] \\
&= J_{X_k}(\overline{X}_k)\mathrm{E}[\Delta_k\Delta_k^T]J_{X_k}^T(\overline{X}_k) \\
&= J_{X_k}(\overline{X}_k)\Sigma_{\Delta_k}J_{X_k}^T(\overline{X}_k)
\end{aligned}$$

was wir der besseren Übersicht wegen mit

$$\Sigma_{\theta_k} = J_k \Sigma_{\Delta_k} J_k^T$$

abkürzen. Diese Kovarianzmatrix ist quadratisch und symmetrisch. Deshalb sind alle M Eigenwerte λ_{i_k}, i = 1, ..., M mit $\lambda_{0_k} \geq \ldots \geq \lambda_{M_k} \geq 0$, reell [Jänich 81; Kowalsky 74]. Die Eigenwerte geben die *skalaren* Varianzen in Richtung der zugehörigen orthonormalen Eigenvektoren e_{i_k} von Σ_{θ_k} an, es ist also $\mathrm{var}(e_{i_k}^T X_k) = \lambda_{i_k}$. Im $\mathbf{R}^{\mathrm{M}}$ spannen die M Eigenvektoren ein *Ellipsoid* auf, dessen Hauptachsen in Richtung von e_{i_k} die Länge der Eigenwerte λ_{i_k} haben.Ein Kriterium für die Güte der Schätzung ist die Minimierung des Volumens dieses Ellipsoids.

Die Fusion der einzelnen Zustände θ_k soll nun durch eine Linearkombination

$$\hat{\theta} = \sum_{k=1}^{N} W_k \theta_k$$

unter Variation der Gewichtungsmatrizen W_k so vorgenommen werden, daß das Volumen des Ellipsoids von $\hat{\theta}$ minimal wird, das also in diesem Sinne die Varianz des geschätzten Zustands minimiert wird. Die Aufgabe besteht also darin, geeignete Gewichtungsmatrizen für jedes *k* zu bestimmen. Die erste Bedingung für die Gewichtungsmatrizen folgt aus der Forderung nach Erwartungstreue für die Schätzung. Für eine gegen unendlich strebende Anzahl von Messungen soll der Erwartungswert des Schätzwerts gegen den wahren Wert von θ streben, also $E[\hat{\theta}] = \theta$. Setzt man den vorstehenden Summenausdruck ein, so ergibt sich

$$\mathrm{E}\left[\sum_{k=1}^{N} W_k \theta_k\right] = \theta$$

und mit $\mathrm{E}\left[\theta_k\right] = \mathrm{f}(\overline{X}_k) = \theta$ sowie der Linearität der Erwartungswertbildung folgt daraus sofort, daß die Summe der Gewichtungsmatrizen die Einheitsmatrix sein muß. Die Kovarianzmatrix $\Sigma_{\hat{\theta}}$ von $\hat{\theta}$, deren assoziiertes Unsicherheitsellipsoid zu minimieren ist, berechnet sich ebenfalls aus dem Summenausdruck zu:

$$
\begin{aligned}
\Sigma_{\hat{\theta}} &= \mathrm{E}\left[(\hat{\theta} - \mathrm{E}[\hat{\theta}])(\hat{\theta} - \mathrm{E}[\hat{\theta}])^T\right] \\
&= \mathrm{E}\left[(\sum_{k=1}^{N} W_k\theta_k - \sum_{k=1}^{N} W_k\,\mathrm{E}[\theta_k])(\sum_{k=1}^{N} W_k\theta_k - \sum_{k=1}^{N} W_k\,\mathrm{E}[\theta_k])^T\right] \\
&= \mathrm{E}\left[(\sum_{k=1}^{N} W_k J_k \Delta_k)(\sum_{k=1}^{N} W_k J_k \Delta_k)^T\right] \\
&= \mathrm{E}\left[(\sum_{k=1}^{N} W_k J_k \Delta_k)(\sum_{k=1}^{N} \Delta_k^T J_k^T W_k^T)\right]
\end{aligned}
$$

$$
= \begin{bmatrix} W_1 & W_2 & \cdots & W_N \end{bmatrix}
\begin{bmatrix}
J_1\,\mathrm{E}[\Delta_1\Delta_1^T]J_1^T & J_1\,\mathrm{E}[\Delta_1\Delta_2^T]J_2^T & \cdots & J_1\,\mathrm{E}[\Delta_1\Delta_N^T]J_N^T \\
J_2\,\mathrm{E}[\Delta_2\Delta_1^T]J_1^T & & & \vdots \\
\vdots & & & \vdots \\
J_N\,\mathrm{E}[\Delta_N\Delta_1^T]J_1^T & \cdots & \cdots & J_N\,\mathrm{E}[\Delta_N\Delta_N^T]J_N^T
\end{bmatrix}
\begin{bmatrix} W_1^T \\ W_2^T \\ \vdots \\ W_N^T \end{bmatrix}
$$

$$
= W \begin{bmatrix}
J_1\Sigma_{\Delta_1}J_1^T & J_1\Sigma_{\Delta_1\Delta_2}J_2^T & \cdots & J_1\Sigma_{\Delta_1\Delta_N}J_N^T \\
J_2\Sigma_{\Delta_2\Delta_1}J_1^T & & & \vdots \\
\vdots & & & \vdots \\
J_N\Sigma_{\Delta_N\Delta_1}J_1^T & \cdots & \cdots & J_N\Sigma_{\Delta_N}J_N^T
\end{bmatrix} W^T
$$

Nimmt man nun an, daß die Kovarianzmatrizen zwischen den einzelnen Störvektoren verschwinden, so wird die obenstehende Matrix zur Diagonalmatrix. Die Berechnung des Volumens des mit der Diagonalmatrix $\Sigma_{\hat{\theta}}$ assoziierten Ellipsoids ist langwierig, es sei deshalb auf [Nakamura 89] verwiesen. Es ergibt sich als Optimum für die einzelnen Gewichtungsmatrizen durch Minimierung des Volumens:

$$
W_k = \left[\sum_{j=1}^{N}\left(J_j\Sigma_{\Delta_j}J_j^T\right)^{-1}\right]^{-1}\left(J_k\Sigma_{\Delta_k}J_k^T\right)^{-1}
$$

Setzt man dies in den Ausdruck für den Schätzwert ein, so wird:

$$
\hat{\theta} = \left[\sum_{j=1}^{N}\left(J_j\Sigma_{\Delta_j}J_j^T\right)^{-1}\right]^{-1}\sum_{k=1}^{N}\left(J_k\Sigma_{\Delta_k}J_k^T\right)^{-1}\theta_k
$$

Vergleicht man dies mit dem Ausdruck für die Maximum–Likelihood–Schätzung (2.5.1) bzw. mit der Maximum–a–posteriori–Methode und dem statischen Kalman–Filter (Beispiel 1 in 2.5.5) ohne a–priori–Information

$$
\hat{\mu} = \left[\sum_{j=1}^{N}\Sigma_{X_j}^{-1}\right]^{-1}\sum_{k=1}^{N}\Sigma_{X_k}^{-1}X_k
$$

so erkennt man vollständige Übereinstimmung.

Die Maximum–Likelihood–Schätzung minimiert also auch das Volumen des Ungenauigkeitsellipsoids der Kovarianzmatrix des Schätzwertes. Es läßt sich aber bei dieser Herleitung gut zeigen, wie nichtlineare und über Jacobi–Matrizen linearisierte Sensormodelle in diese Schätzmethoden einbezogen werden können. Wir erinnern hier noch einmal daran, daß die inverse Sensormodellierung verwendet wurde. Die Sensorinformationen werden nicht durch Addition von transformierten Sensordaten und Fehler berechnet, sondern es wird von einem additiven Fehler hinsichtlich der rohen Sensordaten ausgegangen. Da allerdings danach von einer linearen Näherung für die Fehler ausgegangen wird, erscheint es trotz der im Einzelfall möglicherweise besser mit der Wirklichkeit übereinstimmenden Modellierung fraglich, ob sich diese Fehlerbetrachtung auszahlt. Soll die Methode auf konkret vorliegende Sensoren angewendet werden, muß auch hier für jeden Sensortyp ein Sensormodell definiert werden. Weiterhin muß natürlich für jeden Sensor die Funktion f(·) und das typische Sensorrauschen ermittelt werden.

1.4.5 Fusion unter Einbeziehung komplexerer Umweltmodelle

In den vorhergehenden Abschnitten wurde die Leistungsfähigkeit einzelner Fusionsmethoden bezüglich abstrakter Sensordaten X und einem abstrakten Umweltzustand θ untersucht. In einer konkreten Anwendung stellt sich nun die Frage, wie diese abstrakten Daten beschaffen sein müssen. Im einfachsten Fall werden flächen– oder volumenorientierte Umweltmodelle eingesetzt (siehe Abschnitt 1.2), wie beispielsweise in [Matthies 88; Matthies 88b; Moravec 89; Lim 92]. Dort wird für die Steuerung eines mobilen Roboters der betrachtete Raum in Rasterelemente unterteilt und ein Rasterelement als belegt markiert, wenn sich ein Objekt oder ein Objektteil innerhalb des Elements befindet. Die Sensoren müssen in diesem Fall lediglich entscheiden, ob sich ein Objekt im betrachteten Element befindet oder nicht. Die Menge der Umweltzustände ist also $\theta = \{\mathcal{B}, \mathcal{L}\}$ für „Zelle belegt" und „Zelle leer" und die Sensordaten sind dementsprechend nur $X = \{B, L\}$. Die Fusion mehrerer Beobachtungen solcher binärer Sensoren für eine Zelle erfolgt durch Anwendung des Satzes von Bayes: Sei die a–priori–Wahrscheinlichkeit, daß eine Zelle von einem Objekt belegt werde $\pi(\mathcal{B})$ und die a–priori–Wahrscheinlichkeit, daß eine Zelle leer ist, $\pi(\mathcal{L})$. Die Wahrscheinlichkeit, daß der Sensor eine belegte Zelle als solche erkennt, sei gemäß $P(X|\theta)$ verteilt[1]. Dann ergibt sich die a–posteriori Wahrscheinlichkeit dafür, daß eine Zelle belegt ist mit Hilfe der Messung X eines Sensors nach dem Satz von Bayes sofort zu:

$$P(\mathcal{B} \mid X) = \frac{P(X \mid \mathcal{B})\pi(\mathcal{B})}{P(X \mid \mathcal{B})\pi(\mathcal{B}) + P(X \mid \mathcal{L})\pi(\mathcal{L})}$$

und entsprechend kann dies für weitere Sensoren durchgeführt werden. Eine solche Modellierung mag für die Erzeugung einfacher Entfernungskarten, wie sie etwa zur Umfahrung von Hindernissen benötigt wird, ausreichend sein. Wie in 1.2 ausgeführt, werden jedoch für die meisten anderen Aufgabenstellungen auch genauere Modelle benötigt. Die in 1.2 angegebenen Primitiven sind nur ein erster Schritt zur Repräsentation komplexer Objekte (eine Ecke läßt sich beispielsweise über die

[1] Das heißt hier nichts anderes, als daß die vier Wahrscheinlichkeiten „$P(B|\mathcal{B})$ = Sensor gibt belegt aus, wenn belegt", „$P(B|\mathcal{L})$ = Sensor gibt belegt aus, wenn leer", „$P(L|\mathcal{B})$ = Sensor gibt leer aus, wenn belegt" und „$P(L|\mathcal{L})$ = Sensor gibt leer aus, wenn leer" bekannt sein müssen.

Kombination dreier Geradensegmente im rechten Winkel darstellen). Die Sensoren messen die Parameter dieser Primitiven, wobei jene für eine sinnvolle Fusion nicht nur die gleiche Umwelteigenschaft messen müssen (uns interessiert hauptsächlich die Entfernung), sondern sie müssen ihre Meßergebnisse auch in gleicher oder vergleichbarer Form abliefern, gewissermaßen also „die gleiche Sprache sprechen". Um dies zu ermöglichen, wird gemeinhin davon ausgegangen, daß die Objekte vollständig und ihre Position näherungsweise bekannt ist (siehe unter anderem [Faugeras 86; Grandjean 89; Pollard 87; Porrill 87; Porrill 87b; Porrill 88; Shekhar 86; Shekhar 88]). Das Problem besteht demnach darin, für die Anpassung der Modellvorstellungen an die wahre Umwelt die Position und Orientierung von Objekteigenschaften (z.B von Ecken) durch die Messungen möglichst genau zu ermitteln.

Wir betrachten zunächst eine einfache Methode zur genauen Berechnung der Parameter eines Modells aus verschiedenen Beoachtungen eines Sensors mit der Methode der Minimierung der Fehlerquadrate (2.1.4). Es liege hierzu ein Umweltmodell mit verschiedenen Objekten vor. Das Koordinatensystem dieser Objekte falle jeweils mit dem Ursprung des Sensorkoordinatensystems zusammen. Der Sensor nimmt nun $i = 1 \ldots N$ Beobachtungen vor und ermittle ein Objekt im Raum. Dessen Koordinatensystem sei eine Transformation des korrespondierenden Koordinatensystems dieses Objekts im Umweltmodell über die Beziehung[1]

$$\sigma_i = Ro_i + t$$

Dabei bezeichnet σ_i einen Vektor vom Ursprung des Sensorkoordinatensystems (Koordinatensystem des Umweltmodells) zu einem Punkt auf dem vom Sensor beobachteten Objekt im Raum und o_i einen Vektor vom Ursprung der beiden Koordinatensysteme zum Objektpunkt im Modell. R stellt eine Rotationsmatrix und t einen Translationsvektor dar. Gesucht ist nun unter Benutzung mehrerer Beobachtungen die optimale Matrix $\hat{R}$ und der optimale Vektor $\hat{t}$, die das Objekt im Umweltmodell auf das Objekt in der wahren Umwelt transformieren. Dabei wird vorausgesetzt, daß perspektivische Verzerrungen bereits eliminiert wurden und daß das Modell im gleichen Maßstab wie die Sensorbeobachtung vorliegt. Die Bestimmung von $\hat{R}$ und $\hat{t}$ erfolgt über die Betrachtung mehrerer korrespondierender Punkte von Objektmodell (Punkte o_i) und Sensorbeobachtungen (Punkte σ_i)[2]. Wir betrachten zunächst die Rotation R. Dazu werden mehrere Richtungsvektoren $\delta_i = \sigma_{i_1} - \sigma_{i_2}$ und mehrere dazu korrespondierende Vektoren $d_i = o_{i_1} - o_{i_2}$ im Objektsystem zueinander in Beziehung gesetzt über die Beziehung

$$\delta_i = R_i d_i$$

Gesucht ist nun die Matrix $\hat{R}$, die den Fehler $e(R) = (\delta - Rd)^T(\delta - Rd)$ minimiert. Dabei sind $\delta = [\delta_1\ \delta_2\ \ldots\ \delta_N]^T$ und $d = [d_1\ d_2\ \ldots\ d_N]^T$ die zusammengefaßten Beobachtungs– bzw. Modellvektoren und $R = \mathrm{diag}(\hat{R}, \hat{R}, \ldots \hat{R})$ eine Block–Diagonalmatrix, längs deren Hauptdiagonale die zu schätzende Matrix $\hat{R}$ aufgereiht ist. Die Bestimmung der gesuchten Matrix erfolgt über die Ableitung aller Gleichungen des resultierenden Gleichungssystems nach allen gesuchten Größen, Nullsetzen und

1 Es erscheint wünschenswert, Untersuchungen darüber durchzuführen, wie sich andere Transformationen (mit homogenen Matrizen, mit Quaternionen) auf die Schätzung auswirken.

2 Über die Bestimmung dieser Punkte wird keine Aussage gemacht, obwohl dies in der Praxis (in nicht idealisierten Umwelten) wohl auf erhebliche Probleme stößt. Zweckmäßigerweise verwendet man jedoch markante Punkte, wie beispielsweise Eckpunkte.

Auflösung. Nach Bestimmung der Rotation erfolgt die Bestimmung der Translation in entsprechender Weise. In [Shekhar 86] wird ähnlich vorgegangen, allerdings mit einer zusätzlichen Gewichtungsmatrix, deren Elemente umgekehrt proportional zum erwarteten Fehler der Messung sind. Ein verwandter Ansatz findet sich in [Vemuri 87], siehe auch [Tsai 87; Silverman 87].

Bei dem gerade geschilderten Vorgehen wird vorausgesetzt, daß ein genaues Umweltmodell existiert, welches alle zu erkennenden Objekte enthält (mit allen möglicherweise vorkommenden Ansichten[1]). Dies verringert die Flexibilität erheblich. Einen wesentlich günstigeren Ansatz stellt deshalb die Bereitstellung eines mächtigeren Satzes von Primitiven dar, die auf die beobachtete Szene angepaßt werden [Porrill 88; Crowley 87]. Zu diesen Primitiven gehören z.B. Geradensegmente und Kanten rechtwinkliger Objekte. Wir betrachten zunächst das Vorgehen bei Erkennung einer Kante[2]. Die *Lage* der Kante wird beschrieben durch die Lage der Gerade[3] entlang derer die Kante verläuft. Die gemessene Gerade wird zweckmäßig durch einen Punkt p_0 und einen Richtungsvektor v_0 spezifiziert. Die Abweichung gegenüber der wahren Lage läßt sich nun über eine Abweichung von der Richtung v_0 und eine Abweichung für den Punkt p_0 angeben. Um die Angabe dieser Abweichungen gegenüber Transformationen des Koordinatensystems invariant zu machen, wird am Punkt p_0 eine orthonormale Basis (v_0, v_1, v_2) unter Benutzung des Vektors v_0 aufgebaut. Der neue Punkt p der veränderten Geraden g und ihr Richtungsvektor v wird dann als Linearkombination der Basisvektoren angegeben (siehe Bild 16). Es wird für die Gerade g

$$p = p_0 + \theta_1\, v_1 + \theta_2\, v_2$$

$$v = v_0 + \theta_3\, v_1 + \theta_4\, v_2$$

Damit läßt sich die Abweichung über einen Vektor Π aus den vier skalaren Größen θ_1, θ_2, θ_3 und θ_4 beschreiben.

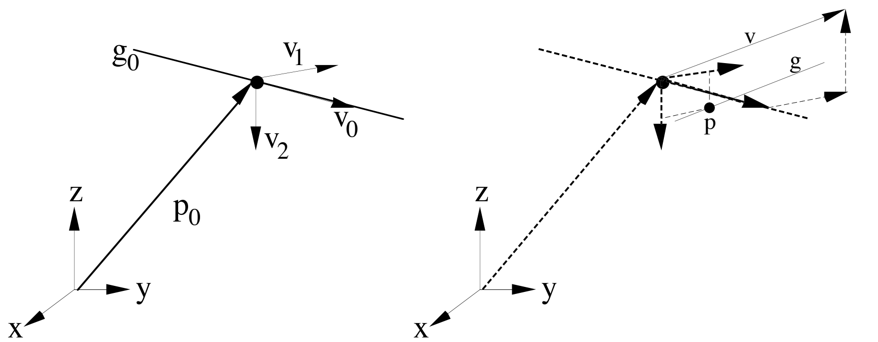

Bild 16: Gemessene Gerade (g_0) und eine ihrer möglichen, wirklichen Lagen (g).
Extremer Fall, nur zur Verdeutlichung der Verhältnisse; normalerweise werden

1 Mit anderen Worten: Es muß ein mehr oder weniger vollständiger Aspektgraph vorhanden sein (siehe [Koenderink 79]).

2 Ein Fusionsansatz, der sich ausschließlich der Kanten in (akustisch und optisch gewonnen) Bildern bedient, findet sich in [Delcroix 88]. Dabei werden die aus den beiden Beobachtungen resultierenden Kantenbilder mit Hilfe der Variationsrechnung so kombiniert, daß das resultierende Kantenbild mehreren Randbedingungen genügt. Deren interessanteste ist das von den Autoren so genannte Konformitätsprinzip: Danach sollte die Feststellung einer Kante an einer Stelle im einen Bild begleitet sein von einer Kante im anderen Bild an der gleichen Stelle oder in deren Nähe. Dies läßt sich auch mathematisch über die Differenz der beiden „Kantenfunktionen“ beschreiben. Diese Differenz wird über die gesamte Bildfläche integriert und durch Variation der Kanten im Ergebnisbild minimiert.

3 Die Lage der Endpunkte wird hier nicht berücksichtigt. Es geht hier nur um die Lage der Gerade entlang derer die Kante läuft. Dazu reicht die Feststellung eines Punktes und eines Richtungsvektors.

g und g_0 nahe beieinander liegen.

Für die statistische Verteilung der Vektorkomponenten läßt sich eine Verbunddichte angeben. Im Falle einer Normalverteilung genügt dazu die Angabe eines Mittelwertvektors $\bar{\theta}$ und einer Kovarianzmatrix Σ_θ. Insgesamt läßt sich für die Kante dann eine vollständige Beschreibung über (p_0, v_0, v_1, v_2, $\bar{\theta}$, Σ_θ) geben. Die beiden letzten Gleichungen lassen sich auch zusammenfassen zu

$$\psi = \psi_0 + F(v_0, v_1, v_2)\theta$$

mit offensichtlicher Wahl der Abkürzungen. Entsprechendes gilt für einen Punkt; dieser wird über eine beliebig zu wählende Basis (v_0, v_1, v_2) (mit drei Parametern (θ_1, θ_2, θ_3)) und wieder über Mittelwertvektor und Kovarianzmatrix beschrieben. Schließlich läßt sich auch eine Ebene über einen Punkt und einen Richtungsvektor beschreiben, so daß hier die gleiche Anzahl Parameter wie bei einem Punkt erfoderlich ist. Alle diese Primitiven lassen sich demnach in die Form $\psi = \psi_0 + F(v_0, v_1, v_2)\theta$ bringen, wo F eine Matrix und θ der Parametervektor ist, der hier die Abweichungen beschreibt. Ein gesamtes Objekt besteht dann aus vielen solcher Primitiven, diese können wieder in einem Vektor zusammengefaßt werden. Ein zusammengesetztes Objekt wird durch eine Liste von Primitiven

$$\Psi = (\psi_1, \psi_2, ..., \psi_\mathrm{N})$$

mit den jeweils zugehörigen orthonormalen Basen modelliert. Jede Deformation des Objektes wird durch die Liste der Darstellungsfehler θ_i im Hinblick auf diese Rahmen ψ_i durch

$$\Theta = (\theta_1, \theta_2, ..., \theta_N)^T$$

charakterisiert. Das Objekt wird damit vollständig durch die aus einer oder mehreren Messungen resultierende Liste Ψ aller Objekteigenschaften und die korrespondierende Liste Θ der Abweichungen beschrieben. Letztere kann entweder durch Mehrfachmessungen und/oder durch Betrachtung des Bildes und schrittweise Einführung geometrischer Randbedingungen (Einschränkungen) gewonnen und verbessert werden. Der Vorteil einer solchen Darstellung ist insbesondere, daß auf diese Art und Weise die Meßergebnisse heterogener Sensorsysteme (Systeme aus unterschiedlichen Sensoren) einheitlich repräsentiert werden können. Im Falle kleinerer Darstellungsfehler können Normalverteilungen angenommen werden. Dann ist der Erwartungswert $\bar{\Theta}$ und die Kovarianzmatrix Σ_Θ des „Systemvektors" Θ eine ausreichende Statistik für die Darstellungsfehler eines zusammengesetzten Objektes. Sind die θ_i voneinander unabhängig, handelt es sich bei der Kovarianzmatrix um eine Matrix, deren Diagonale aus den einzelnen Varianzen besteht und deren übrige Elemente Null sind.

Für ein Sensorsystem, welches auch komplexere Szenen ohne Vorkenntnis über mögliche Objekte und Objektkonfigurationen erfassen soll, ist die Definition mehrerer geometrischer Primitiven (Kante, Ecke, Kreisbögen, …) und Entwicklung korrespondierender Sätze der Abbildungskonfigurationen solcher Primitiven erforderlich. Unter einem Satz solcher Primitiven ist dann beispielsweise ein Katalog aller möglichen Konfigurationen von Ecken zu verstehen, wie sie bei der Bildinterpretation schon lange Zeit verwendet werden. Solche Kataloge können darüber hinaus als „gemeinsame Sprache“ für viele Sensoren verwendet werden, sofern sie alle in der Lage sind, diese Primitiven erkennen zu können[1].

Als Sensortransformation kann im allgemeinsten Fall die in Abschnitt 1.3 angegebene implizite, nichtlineare From Verwendung finden: Sei Ψ_0 eine gute Annäherung an das Gesamtobjekt und $\hat{\Theta}_k$ guter Linearisierungspunkt für den Abweichungsvektor, dann gelte für den die Beobachtung des Gesamtobjekts beschreibenden Vektor X für ein fehlerfreies System f($\overline{X}$, θ, Ψ_0) = 0 mit $\overline{X}$ als korrekter Beobachtung. Nach Linearisierung wird das zu

$$0 = \mathrm{f}(\overline{X},\Theta,\Psi_0) \approx \mathrm{f}(X,\hat{\Theta}_k,\Psi_0) + \left(\left. \frac{\partial \mathrm{f}(\xi,\Theta,\Psi_0)}{\partial \Theta} \right|_{\substack{\xi=X \\ \Theta=\hat{\Theta}_k}} (\Theta - \hat{\Theta}_k) \right) + \left(\left. \frac{\partial \mathrm{f}(\xi,\Theta,\Psi_0)}{\partial \xi} \right|_{\substack{\xi=X \\ \Theta=\hat{\Theta}_k}} (X - \overline{X}) \right)$$

bzw. nach Einführung von Abkürzungen (siehe Abschnitt 1.3):

$$-\mathrm{f}(X,\hat{\Theta}_k,\Psi_0) + J_\Theta \hat{\Theta}_k = J_\Theta \Theta - J_X \Delta$$

Gesucht wird hier der optimale Schätzvektor $\hat{\Theta}$ für die Abweichung[2]. Dies kann wieder über Kalman–Filterung geschehen (wie in 2.5.5, Beispiel 4 geschildert). Es ergeben sich zwei Probleme, die wir abschließend betrachten wollen:

1. **Unvollständige Information**: Insbesondere bei der Fusion der Daten heterogener Sensorsysteme ist es wahrscheinlich, daß ein Sensor, der eine oder mehrere Beobachtungen beisteuert, nicht alle im Vektor Θ zusammengefaßten Objekteigenschaften messen kann. Soll nach dieser unvollständigen Messung gleich geschätzt werden, so kann dies so durchgeführt werden, daß für die Berechnung des neuen Schätzwertes nur die neue Information herangezogen und die restliche als unverändert gültig angesehen wird. Dann ist in $\mathrm{f}(X,\hat{\Theta}_k,\Psi_0)$ im Vektor X an der entsprechenden Stelle die Teilinformation des Sensors einzusetzen, an den restlichen Stellen die Information, die sich unter Berücksichtigung der Sensortransformation aus Ψ_0 und $\hat{\Theta}_k$ ergibt. Die Sensortransformation muß dabei natürlich so beschaffen sein, daß sie die Umsetzung der betrachteten Objekteigenschaft beschreibt und die restlichen, vom Sensor nicht beeinflußten Komponenten von X unverändert läßt.

2. **Randbedingungen**: Vielfach ist von einem Objekt bekannt, daß einige (oder alle) seiner beobachteten Primitiven in bestimmten Relationen zueinander stehen (siehe z.B. [Nishida 87; Malik 87]). Das wohl einfachste Beispiel dafür sind zwei Kanten, von denen bekannt ist, daß sie senkrecht zueinander stehen, ihre Richtungsvektoren also orthogonal sind. Mit einem Kunstgriff können solche Randbedingungen mit in die Berechnung des Schätzwertes einbezogen werden: Sie werden dazu so formuliert, daß

---

1 Zur Erzeugung von Drahtrahmenmodellen einfacher Objekte siehe [Stenstrom 86].

2 Hier könnte man natürlich ebenso den Gesamtvektor Ψ schätzen, dies wird in [Wen 92] auch tatsächlich durchgeführt. In diesem Fall stellt sich aber die Frage, ob die für die kleinen Werte von Θ vorgenommene Annahme von Normalverteilungen auch für Ψ gelten kann.

sie als Beobachtungen ohne Fehler aufgefaßt werden können [Porrill 88; Durrant–Whyte 91; Wen 92]. Diese Einbeziehung wird in [Porrill 88; Pollard 89] als Nebenbedingung bei einer Schätzung nach der Gauß–Markow–Methode realisiert. Bei dieser Methode handelt es sich um eine statische Formulierung des Kalman–Filters (die Methode und ihre Weiterentwicklung zum Kalman–Filter werden in [Sorenson 80] erläutert). Dabei können allerdings nur lineare oder linearisierte Randbedingungen mir einbezogen werden. In [Durrant–Whyte 91; Wen 92] wird demgegenüber das gesamte Meßmodell um die Randbedingung erweitert und mit Hilfe der Methode der kleinsten Quadrate ein Parametervektor Θ bestimmt, der zum aus den Messungen gewonnenen Parametervektor und dem den Randbedingungen vollständig entsprechenden den kleinsten Abstand hat.

Am einfachsten und flexibelsten scheint es jedoch, die Randbedingung als ideale Messung direkt in die Folge der anderen Beobachtungen einzubeziehen. Es sei also beispielsweise vorgegeben, daß eine bestimmte Kante, die im Parametervektor durch die Komponente θ_j beschrieben werde, senkrecht auf einer Kante stehen muß, die im Meßvektor durch die Komponente x_i beschrieben wird. Dann würde eine Formulierung dieser Randbedingung als ideale Beobachtung k die folgende Form annehmen:

$$\begin{pmatrix} -\cdot- \\ x_i \\ -\cdot- \\ -\cdot- \end{pmatrix}_k = F_k \begin{pmatrix} -\cdot- \\ -\cdot- \\ -\cdot- \\ \theta_j \end{pmatrix}_k$$

wo die Matrix F_k dann so beschaffen sein muß, daß sie die Rotation von x_i auf θ_j um 90° beschreibt und die anderen Komponenten des Meßvektors unberührt läßt. Die nach dieser „Beobachtung“ zweckmäßigerweise anzuwendende Vorschrift für die Schätzung ist die des statischen Kalman–Filters mit einem Meßrauschen von Null, also $\Sigma_W = 0$ (siehe dazu das Ende von Beispiel 4 in Abschnitt 2.5.5):

$$\begin{aligned} \hat{K}_k &= P_{k-1} F_k^T \left[F_k P_{k-1} F_k^T \right]^{-1} \\ \hat{\Theta}_k &= \hat{\Theta}_{k-1} + \hat{K}_k [X_k - F_k \hat{\Theta}_{k-1}] \\ P_k &= P_{k-1} - \hat{K}_k F_k^T P_{k-1} \end{aligned}$$

wobei $\hat{K}$ die Zielmatrix, P die Kovarianzmatrix des Parametervektors Θ und X_k der Meßvektor ist. Wenn andere Randbedingungen einzuhalten sind, müssen sie in die Form $X = F\Theta$ gebracht werden. Handelt es sich um nichtlineare Bedingungen, dann müssen sie zunächst linearisiert werden (in gleicher Weise, wie dies in Abschnitt 1.3 für ein einfaches Sichtsystem gezeigt wurde).

Der große Vorteil des in diesem Abschnitt vorgestellten Ansatzes „unsicherer Geometrie“ besteht darin, daß die Meßungenauigkeit verschiedener Sensoren in direkten Zusammenhang mit der Geometrie der von ihnen vermessenen Objektmerkmale gebracht wird. Die Schätzung des richtigen Meßwertes geht mit der Schätzung des „korrekten“ Objektes einher. Ein bestimmtes Objekt oder eine bestimmte Konfiguration wird nicht vorausgesetzt. Damit kann prinzipiell ein Multisensorsystem aufgebaut werden, das sich auf unterer Ebene – abgesehen von den Primitiven – auf kein Modell stützt. Entscheidend ist die Möglichkeit, Zusatzwissen einfach in den Mechanismus der Schätzung einbeziehen zu können, da damit oftmals entscheidende Verbesserungen erzielt werden können.

Bei Anwendung auf konkret vorliegende Sensoren muß ein Transformationsprozeß definiert werden, der die eigentlichen Sensordaten in eine entsprechende geometrische Form bringt. Geeignete Meßprimitiven müssen ausgewählt werden und das Sensorrauschen für jeden Sensor bestimmt werden.

1.4.6 Fusion mit Bayesscher Schätzung bei beliebig verteiltem Meßfehler

In den vorhergehenden Abschnitten wurde zur Fusion der Daten die Maximum–a–posteriori–Schätzung, die Maximum–Likelihood–Schätzung oder deren rekursive Formulierung in Form des statischen Kalman–Filters herangezogen. Implizit oder explizit wurde dabei ein normalverteiltes Sensorrauschen und ein lineares oder linearisiertes Sensormodell vorausgesetzt. Während man gebräuchliche Sensoren häufig durch linearisierte Sensormodelle ausreichend beschreiben kann, schränkt die Annahme von normalverteilten Meßfehlern die Glaubwürdigkeit von Schätzergebnissen insbesondere dann stark ein, wenn mit Ausreißern zu rechnen ist. In [Durrant–Whyte 88] wird zur Modellierung von Ausreißern zwar die verunreinigte Normalverteilung nach [Huber 81] vorgeschlagen, aber daraus keine Fusionsmethode entwickelt, die derart verteilte Daten direkt behandeln könnte. Um eine Schätzung bei beliebiger Verteilung und bei beliebiger Sensortransformation vornehmen zu können, wird zweckmäßigerweise direkt auf den Satz von Bayes zurückgegangen und die a–posteriori–Dichte aus a–priori–Dichte und Beobachtung berechnet. Ist diese zentrale Funktion erst einmal bestimmt, kann die Schätzung des Parametervektors nach Definition einer geeigneten Verlustfunktion (siehe 2.5.2) durchgeführt werden. Die aus a–priori–Verteilungsdichte für den Schätzwert $\pi(\theta)$, auf die Beobachtung X bedingter Verteilungsdichte $\mathrm{p}(X|\theta)$ und zur Normierung verwendeten Dichte $\mathrm{p}(X)$ folgende a–posteriori–Dichte berechnet sich nach dem Satz von Bayes wie folgt:

$$\mathrm{p}(\theta | X) = \frac{\mathrm{p}(X | \theta)\pi(\theta)}{\mathrm{p}(X)}$$

Dabei wird angenommen, daß $\mathrm{p}(X|\theta)$ (Sensorparameter, siehe Abschnitt 1.3) und ebenso $\pi(\theta)$ bekannt sind. Die von θ unabhängige und nur zur Normierung dienende Randdichte $\mathrm{p}(X)$ läßt sich durch Integration des im Zähler stehenden Ausdrucks über den gesamten Wertebereich von θ gewinnen. Damit bleibt als wesentliche Operation die Berechnung des Zählers in obiger Gleichung. Da es sich bei den Dichten nun nicht mehr um Normalverteilungen handeln soll, die sich vollständig durch Mittelwert und Varianz spezifizieren lassen, muß bei jeder auf eine Beobachtung folgende Berechnung der gesamte Funktionsverlauf sowohl von $\mathrm{p}(X|\theta)$ als auch von $\pi(\theta)$ ausgewertet werden.

Die auf N Beobachtungen fußende Schätzung von θ läuft dabei so ab, daß mit einer ersten a–priori–Dichte $\pi(\theta)$ und einer ersten Beobachtung X (mit zugehöriger Dichte $\mathrm{p}(X|\theta)$) eine erste a–posteriori–Dichte $\mathrm{p}(\theta|X)$ berechnet wird. Letztere wird dann, wie üblich, als a–priori–Dichte $\pi(\theta)$ des nächsten Schätzschrittes verwendet. Die Abbildung der a–posteriori–Dichte $\pi(\theta)$ auf die bedingte Dichte $\mathrm{p}(X|\theta)$ erfolgt bei der Beobachtung nach der Sensortransformationsvorschrift $\mathrm{f}(\theta, X) = 0$ bzw. $X = \mathrm{f}(\theta)$. Die vollständige Multiplikation zweier Dichten verbietet sich jedoch aus Aufwandsgründen bei der praktischen Durchführung einer Schätzung insbesondere im Falle vektorieller Größen X und θ. Zur Lösung dieses Problems wird in [Hager 89; Hager 89b; Hager 89c; Hager 90; Hager 91] vorgeschlagen, Wahrscheinlichkeitsdichten zu verwenden, welche sich durch abschnittsweise konstante Funktionen („Treppenkurven") gleichbleibender Intervallbreite ausreichend genau approximieren lassen (siehe Bild 17).

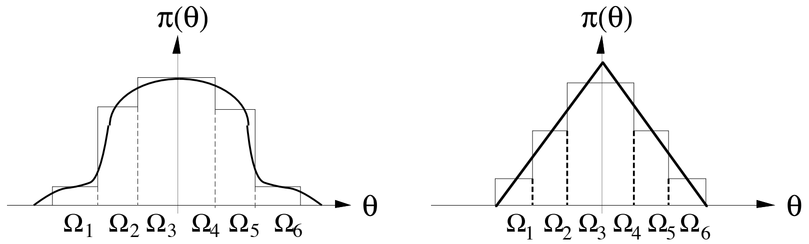

Bild 17: Zwei (a–priori–)Dichtefunktionen $\pi(\theta)$ für den hier skalaren Parameter θ, die sich gut durch die eingezeichneten Treppenkurve mit den Unterteilungen Ω_i annähern lassen.

Innerhalb eines Unterteilungsintervalls Ω_i des Wertebereiches von θ ist der Funktionswert konstant. Die Berechnung des Zählers in obiger Gleichung wird dann mit diesen Näherungen durchgeführt. Dementsprechend erhält man eine ebenfalls treppenförmige a–posteriori–Dichte, die jeweilige Höhe des Treppenstücks ergibt sich durch Integration über das entsprechende Unterteilungsintervall Ω_i, also

$$\mathrm{p}_i(\theta|X) = \frac{\int_{\Omega_i} \mathrm{p}(X|\theta)\pi(\theta)\,\mathrm{d}\theta}{\sum_{i=1}^{K}\int_{\Omega_i} \mathrm{p}(X|\theta)\pi(\theta)\,\mathrm{d}\theta} \qquad (1.4.6\text{–}1)$$

wo K die Gesamtanzahl der Unterteilungsintervalle ist. Berücksichtigt man, daß die Approximation von $\pi(\theta)$, also Ω_i, innerhalb eines Intervalls konstant ist, so wird das zu

$$\mathrm{p}_i(\theta|X) = \frac{\pi(\Omega_i)\int_{\Omega_i} \mathrm{p}(X|\theta)\mathrm{d}\theta}{\sum_{i=1}^{K}\pi(\Omega_i)\int_{\Omega_i} \mathrm{p}(X|\theta)\,\mathrm{d}\theta} \qquad (1.4.6\text{–}2)$$

Zur Berechnung von $\mathrm{p}(\theta|X)$ ist damit statt der Multiplikation zweier Funktionen nur noch die Multiplikation der Werte zweier innerhalb einer Unterteilung konstanten Treppenfunktionen erforderlich. Innerhalb des Intervalls ist der Wert der Funktion $\pi(\Omega_\mathrm{i})$ bekannt, der von $\mathrm{p}(X|\theta)$ muß durch Integration der „Wahrscheinlichkeitsmasse" über das Unterteilungsintervall der a–priori–Dichte ermittelt werden. Wir illustrieren das Vorgehen an einem Beispiel für den skalaren Fall, beim vektoriellen Fall sind die üblichen Erweiterungen vorzunehmen; dies stellt dann kein Problem dar, wenn die Komponenten des Beobachtungsvektors als unabhängig vorausgesetzt werden. Es werde angenommen, daß eine unterteilte a–priori–Dichte wie in Bild 18a vorliege. Die Unterteilungen weisen einen Abstand der Länge $d/3$ auf; innerhalb der Unterteilung ist der Funktionswert, also die Höhe des Treppenteils, $\pi(\Omega_i)$.

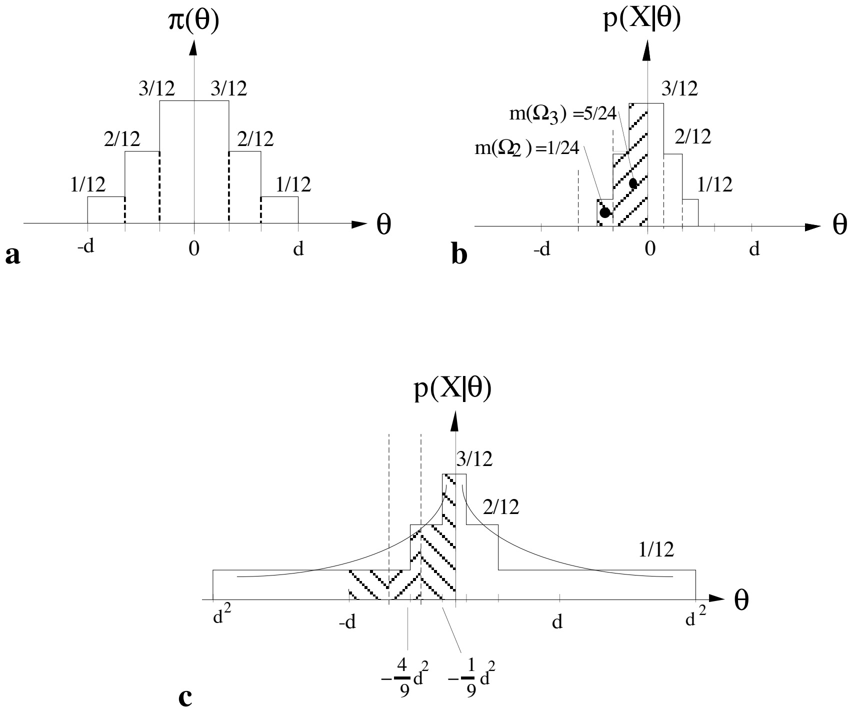

Bild 18: a Beispiel für eine durch eine Treppenfunktion darstellbare a–priori–Funktion. **b** Form der Dichte nach Transformation mit der Funktion $X = \theta/2$. **c** Form der Dichte nach Transformation mit der Funktion $X = \theta^2$. Die eingezeichnete Hyperbel soll den Verlauf nach einer Transformation der nicht diskretisierten Dreiecksverteilung nach a andeuten. Die Funktionswerte wurden nur übernommen; eine Normierung der Höhe erfolgte nicht; dies erfolgt erst in einem zweiten Schritt (siehe Text und Bild 19), da hier nur die Verhältnisse interessieren. Die schraffierten Flächen in b und c bezeichnen den Wert der „Wahrscheinlichkeitsmasse", die sich innerhalb der Unterteilung nach a befinden (also den Wert des Integralausdrucks, siehe Text).

Diese Funktion ist der eine Faktor, der zur Berechnung des Zählers in Gl. (1.4.6–2) eingeht. Der andere Faktor ist die auf die Beobachtung X bedingte Dichte $\mathrm{p}(X|\theta)$, die im Falle eines Sensormodells der Form $X = \mathrm{f}(\theta) + V$ durch Transformation der a–priori–Dichte mit der Sensortransformationsvorschrift gewonnen wird. In Bild 18b ist eine Dichte aufgezeichnet, die sich bei einer Beobachtung $X = 0$ aus der Dichte nach Bild 18a bei einer linearen Transformationsvorschrift der Form $X = \theta/2$ ergeben würde. Bild 18c zeigt dieselbe Dichte nach Transformation durch die nichtlineare Vorschrift $X = \theta^2$, ebenfalls für $X = 0$.

Um eine korrekte Multiplikation der beiden diskretisierten Dichten vornehmen zu können, wird wie folgt vorgegangen: Zunächst wird die Ausgangsdichte (Bild 18a) transformiert, indem die Partitionierungsintervalle Ω_i der Transformationsvorschrift gehorchend gedehnt oder gestreckt werden. Der Funktionswert $\pi(\Omega_i)$ bleibt dabei unverändert (der hierbei entstehende Normierungsfehler wird später beseitigt; im Moment kommt es nur auf die Verhältnisse an). In einem zweiten Schritt wird berechnet, welche „Wahrscheinlichkeitsmasse" $\mathrm{m}(\Omega_i)$ der transformierten Dichte $\mathrm{p}(X|\theta)$ sich innerhalb der ursprünglichen Intervalle Ω_i befindet. Dazu wird der Ausdruck

$$\mathrm{m}(\Omega_i) = \int_{\Omega_i} \mathrm{p}(X \mid \theta)\,\mathrm{d}\theta$$

berechnet (dies entspricht den schraffierten Flächen in Bild 18b und c). Auch hier interessieren nur wieder die Verhältnisse.

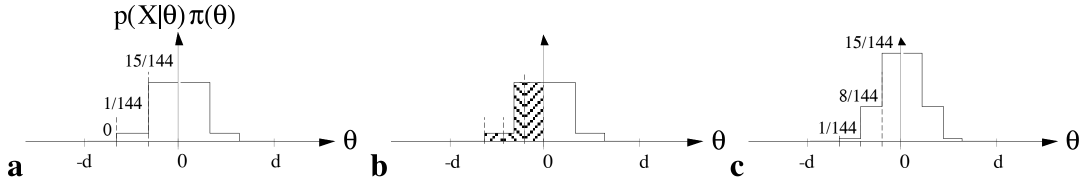

Bild 19: **a** Noch nicht normierter Verlauf der a–posteriori–Dichte p(θ|X) mit Unterteilungen Ω_i.
b Unterteilung des Bereichs, in dem die Dichte nicht verschwindet in sechs Abschnitte und Bestimmung der Wahrscheinlichkeitsmassen für jedes Intervall.
c Endgültige Dichte nach Normierung.

Nachdem damit die Werte der beiden Dichten p(X|θ) und π(θ) innerhalb der Intervalle Ω_i bekannt sind, kann die Berechnung des Zählers von Gl. (1.4.6–2) für eben diese Intervalle erfolgen. Dazu werden die Werte von m(Ω_i) und π(Ω_i) in jedem Intervall miteinander multipliziert. Das Ergebnis zeigt Bild 19a (aus Gründen der Übersichtlichkeit wird hier nur die Transfomation $X = \theta/2$ betrachtet).

In einem letzten Schritt wird dieser Verlauf für den Bereich, innerhalb dessen seine Werte ungleich Null sind, wie die a–priori–Dichte π(θ) unterteilt (in unserem Fall also in sechs Intervalle gleicher Länge). Dann wird bestimmt, welcher Anteil der Wahrscheinlichkeitsmasse sich in jedem dieser Intervalle befindet (Bild 19b) und eine Normierung so vorgenommen, daß gilt $\int_{-\infty}^{\infty} \mathrm{p}(\theta \mid X)\,\mathrm{d}\theta = 1$. Die engültige a–posteriori–Dichte zeigt Bild 19c. Diese Dichte wird im nächsten Schritt zusammen mit der folgenden Beobachtung erneut modifiziert. Mit der nach allen Schritten berechneten a–posteriori–Dichte und einer *für die Aufgabenstellung geeigneten Verlustfunktion* wird der endgültige Schätzwert bestimmt.

Die Fehler, die dieses Verfahren mit sich bringt und die approximative Auswertung des Integralausdrucks

$$\int_{\Omega_i} \mathrm{p}(X \mid \theta)\,\mathrm{d}\theta = \int_{\Omega_i} \mathrm{p}(X - \mathrm{f}(\theta))\,\mathrm{d}\theta$$

werden in [Hager 89b; Hager 91] ausführlich dargestellt. Die Aufgabe der Annahme von Normalverteilungen bringt die Notwendigkeit mit sich, bei jedem Schätzvorgang die Dichte neu berechnen zu müssen (nicht nur zwei Größen, wie bei der Normalverteilung). Dafür können aber zur Darstellung des Fehlers beliebige Dichtefunktionen angenommen werden, was insbesondere für die Modellierung von Ausreißern oder bei nichtlinearen Fehlereinflüssen von Vorteil ist. Generell bieten Verfahren, die möglichst wenige Grundannahmen über die Fehlerverteilungen machen und bedingte Verteilungsfunktionen zur Informationsdarstellung benutzen, den größtmöglichen Spielraum bei der Anpassung an spezifische Aufgabenstellungen. Es bieten sich hier die folgenden Möglichkeiten für weitere Entwicklungen an:

- Spezialisierung auf verunreinigte Gaußverteilung und Optimierung des Algorithmus.
- Vergleich mit Lösungen, die auf Kalman–Filter–Bänken aufbauen und damit ebenfalls die Annahme rein normalverteilter Fehler aufgeben (siehe [Anderson 79], Kap. 8.4).

- Aufgabe der Annahme der bedingten Unabhängigkeit zwischen einzelnen Vektorkomponenten.
- Durchführung praktischer Experimente und Demonstration der Vorteile gegenüber (Bayesschen) Fusionsmethoden, die auf Normalverteilungen basieren.

Bei der angesprochenen Verarbeitung von Daten mit ganz oder teilweise unbestimmter (Fehler–) Dichtefunktion kann auf die Standardverfahren zur Schätzung von Dichten aus Stichproben zurückgegriffen werden. Ein Ansatz dazu findet sich in [Beckerman 92]. Dort wird die a–priori–Dichtefunktion aus den beobachteten Daten geschätzt unter der zusätzlichen Annahme, daß diese Dichte die Entropie der aus ihr folgenden Stichproben maximiert (eine Definition der Entropie im informationstheoretischen Sinne gibt z.B. [Andrews 77]). Wird nur diese Annahme gemacht, folgt daraus sofort, daß eine Gleichverteilung angenommen werden muß. Wird allerdings als zusätzliche Forderung an die Dichte gestellt, daß der Mittelwert der von ihr gezogenen Stichproben gleich dem Mittelwert der beobachteten Daten entspricht, so ergeben sich andere Verteilungen, die mit dem beobachteten Mittelwert parametrisiert sind. Mit diesen Dichten kann dann eine a–posteriori–Dichte mit dem Satz von Bayes berechnet werden.

Wir gehen im folgenden Abschnitt in aller Kürze auf weitere Möglichkeiten ein, Sensordaten zu kombinieren, bevor wir zu einer abschließenden Wertung kommen.

1.4.7 Weitere Aspekte und Ansätze

In den vorigen Abschnitten wurden die wichtigsten grundlegenden Aspekte der statistischen Datenfusion und Konsistenzüberprüfung anhand jüngerer Arbeiten auf diesem Gebiet betrachtet und ihre Leistungsfähigkeit anhand ausgewählter Simulationen illustriert. Wir geben nun noch einen kurzen Überblick über weitere Ansätze, deren Anwendung auf (oder Weiterentwicklung für) eine gegebene Problemstellung geprüft werden sollte.

Wie sich bei der Darstellung gezeigt hat, ist unter der Annahme normalverteilter Fehler das statische Kalman–Filter vorteilhaft einzusetzen, da es die rekursive Berechnung eines Maximum–a–posteriori–Schätzwertes gestattet und vor allem ausführliche Kenntnisse über zeit– und aufwandsoptimierte Algorithmen zu seiner Durchführung vorliegen. Deshalb wird das Kalman–Filter wohl auf absehbare Zukunft der für die meisten Anwendungen zweckmäßigste Schätzer bleiben. Sein Anwendungsbereich ist damit allerdings noch nicht erschöpft. In [Balchen 90] wird der Einsatz des erweiterten Kalman–Filters vorgeschlagen, um aus einer großen Anzahl von Sensoren anhand der momentanen Struktur der Kovarianzmatrix die Menge von Sensoren auszuwählen, die den besten Beitrag zur Verbesserung des Schätzwertes leisten können (zu dieser Problematik siehe Kapitel 1.2). Ferner wird die Verwendung von Kalman–Filtern für die Schätzung und Voraussage der Bewegungsparameter beweglicher und von einer oder mehreren Kameras beobachteten Objekte angeregt (siehe dazu auch [Vemuri 90; Skofteland 91] und [Kay 91; Allen 91; Papanikolopoulos 91]). Bei der Schätzung mit den Meßwerten sehr vieler Sensoren tritt allerdings unter Umständen das Problem auf, daß die Sensoren Meßwerte dem zentralen Zustandsschätzer schneller liefern, als dieser sie verarbeiten kann. Um solche Probleme zu vermeiden, sind schon vor einiger Zeit dezentrale Schätzarchitekturen vorgeschlagen worden [Chang 86; Chiou 86; Hashemipour 88; Fernández 90; Irving 91; Thomopoulos 91]. Bei diesen nimmt jeder Sensor eine lokale Schätzung basierend auf den ihm zur Verfügung stehenden Meßwerten vor. In einem algorithmisch nicht sehr aufwendigen zweiten Schritt

(Assoziation oder Assimilation) werden jedem Sensor die lokalen Schätzwerte aller anderen Sensoren zur Verfügung gestellt und durch Zusammenfassung aller gelieferten Informationen kann der optimale Gesamtschätzwert berechnet werden[1]. In [Durrant–Whyte 90] wird dieses Konzept verwendet, um damit die Schätzung beweglicher Objekte durchzuführen. Der offensichtliche Nachteil ist, daß jeder Knoten jedem anderen Knoten die gesamte Information über seine eigene Schätzung liefern muß, da sonst eine Berechnung der Kovarianzmatrix nicht möglich ist. In [Durrant–Whyte 90] findet der Informationsaustausch über einen Kommunikationsbus statt, an den alle Sensoren angeschlossen sind. Es liegt jedoch auf der Hand, daß hier andere Varianten möglich[2] und sinnvoll sind. Es kann hier auf einen reichen Fundus an Erfahrung aus dem ähnlich gelagerten Bereich der Zielerkennung zurückgegriffen werden (siehe [Han 90; Krzysztofowicz 90; Hoballah 89; Bar–Shalom 88; Chen 87c; Srinivasan 86; Chair 86; Tenney 81]), insbesondere für die Bayes–Schätzung speziell von bewegten Objekte bei verteilten Beobachtern.

Bei den betrachteten statistischen Schätzmethoden wurde davon ausgegangen, daß die Parameter von Meß– oder Prozeßrauschen bekannt sind. Ist dies nicht der Fall, können adaptive Filter eingesetzt werden. Ein Ansatz zu einem solchen Vorgehen findet sich in [Sasiadek 87]. Hier werden die Parameter eines linearen Kalman–Filters in Abhängigkeit vom beobachteten Signal/Rauschverhältnis eingestellt, wobei sich diese Größe natürlich nur langsam ändern darf. Eine solche Adaption ist für praktische Anwendungen sehr nützlich, deshalb sind Forschungen bzw. die Übertragung bekannter Ergebnisse aus der Nachrichtentechnik auf diesem Feld angebracht.

Ein gegenüber den auf Bayes–Schätzungen fußenden Methoden erweitertes Modell des Wahrscheinlichkeitsbegriffes liegt den Verfahren zugrunde, die sich zur Schätzung der Dempster–Shafer–Theorie bedienen (siehe 2.6 und [Garvey 81; Lowrance 86; Bogler 87; Tirumalai 90; Luo 90b; Huntsberger 90]). Wiewohl nicht dem engeren Kreis der Verfahren zur Datenfusion zugehörig, gibt [Sato 92] einen gutes Beispiel dafür, wie die für diese Theorie zentrale Kombinationsregel von Dempster im konkreten Fall eingesetzt wird: Hier wird aus einem Graphen, dessen Knoten innerhalb einer gewissen Diskretisierung alle möglichen Ansichten eines Objektes beinhalten (Aspektgraph), diejenige ausgewählt, die mit der größten Wahrscheinlichkeit vorliegt. Die Schwierigkeit liegt dabei darin, daß bei der Aufnahme Glanzlichter vorliegen. Solche Störungen können bei nicht modellbasierten Systemen den Erkennungsvorgang vollständig unterbinden. Eine dennoch einigermaßen zuverlässige Aussage wird über die Betrachtung von zwei Objekten O_A und O_B (oder Teilobjekten eines größeren Objektes) getroffen. Der Aspektgraph eines gegebenen Objektes wird gewonnen, indem eine das Objekt umgebende gedachte Einheitskugel (deren Mittelpunkt mit dem Ursprung des Objektkoordinatensystem übereinstimmt) abgeschritten wird und für alle Erhebungs– und Umfangswinkel die orthographische Projektion auf die Kugel vorgenommen wird. Da man nicht jede Ansicht bei einer sehr kleinen Änderung der Winkel in den Graphen eintragen kann, führt man eine Änderung nur aus, wenn ein bestimmter Winkel überschritten wurde, und ein „Strukturereignis" stattgefunden hat, sich also die Struktur des Objekts verändert hat (beispielsweise eine Kante verschwunden ist). Besonders einfach ist es, beide Winkel äquidistant zu diskretisieren (also z.B. in

1 Dieses Konzept liegt auch der gegenseitigen Anpassung von Nutzenfunktionen verschiedener Sensoren zugrunde, wie in [Basir 92] vorgestellt. Dabei nimmt ein Sensor zunächst eine lokale Entscheidung gemäß einer Nutzenfunktion vor. Danach wird die Information von allen anderen Sensoren empfangen und die Nutzenfunktion modifiziert (*team–consensus–approach*, siehe auch [Durrant–Whyte 88]).

2 Eine Untersuchung verschiedener (mit unterschiedlicher lokaler Schätzung vor Weiterleitung an einen zentralen Fusionsprozessor arbeitender) verteilter Algorithmen findet sich in [Tucci 90].

Schritten von 10^o zu rastern). Bei der Objekterkennung wird nun wie folgt vorgegangen: Das beobachtete Bild von O_A wird mit allen Ansichten im Aspektgraphen verglichen. Dabei ergibt sich für jede Ansicht ein bestimmtes Maß an Übereinstimmung. Der Vergleichsalgorithmus mißt nun dem Vorliegen einer Ansicht *oder einer bestimmten Menge* von Ansichten eine Wahrscheinlichkeit zu. Ebenso wird bei O_B verfahren. Mit Hilfe der Kombinationsregel von Dempster werden die derart für O_B und O_A erhaltenen Wahrscheinlichkeitsverteilungen miteinander verrechnet und die Ansicht ausgewählt, für die sich die höchste Wahrscheinlichkeit ergibt. Dieser Ansatz wird auch in [Hutchinson 88; Hutchinson 89] verfolgt, um verschiedene, bekannte Teile in einer Montagezelle zu identifizieren und den Ort festzustellen, von dem aus eine Sensorbeobachtung für den weiteren Fortgang des Erkennungsprozesses den größten Nutzen haben könnte. Generell kann festgehalten werden, daß der Einsatz der Dempster–Shafer–Theorie nur selten erwogen wird. Dies liegt zum einen sicher daran, daß die Vorteile gegenüber Bayesschen Verfahren mit höherem Aufwand erkauft werden. Dies kann sich aber in Zukunft durchaus ändern, wenn den Vorteilen, die sich aus der zentralen Idee der beliebigen Verteilung von Wahrscheinlichkeitsmassen auf die Elementarereignisse ergeben, bei komplexeren Systemen mehr Beachtung geschenkt wird (siehe 2.6).

Die dynamische Plazierung von Robotik–Sensoren stellt ein neues Forschungsgebiet dar, bei dem Erkenntnisse aus den Bereichen Planung, Steuerung und Sensorik zusammenfließen. In [Sakane 91] wird eine Methode entwickelt, die Position der Lichtquellen für die Objekterfassung mit photometrischem Stereo den jeweiligen Gegebenheiten anzupassen, siehe auch [Tsai 89]. Demgegenüber wird in [Cameron 89] eine auf statistischen Prinzipien gründende Methode entwickelt, um mit möglichst wenigen Tastpunkten ein dennoch umfassendes Bild eines Objekts bei Abtastung durch einen taktilen Sensor zu erhalten. Dabei gilt es, zwischen den durch das Verfahren des Sensors entstehenden Kosten (hauptsächlich dem Zeitfaktor) und dem Gewinn einer neuen Messung abzuwägen. Eine solche Kombination von statistischen Verfahren zur Schätzung des Nutzens bei gegebener Konfiguration und folgender Entscheidung über einen Sensoreinsatz findet sich auch in [Dean 90]. Dabei wird zum Zwecke der Navigation bei einem mobilen Roboter entschieden, ob einer der zur Verfügung stehenden Sensoren benutzt werden soll, um zusätzliche Information über die Umwelt zu gewinnen[1]. Parameter, die diese Entscheidungen beeinflussen, werden als Kostenfunktionen formuliert. In einer gegebenen Situation wird darüber entschieden, einen Sensor einzusetzen, indem geprüft wird

- wie gut seine spezifischen Fähigkeiten sind, die verschiedenen momentan zur Auswahl stehenden Hypothesen zu unterscheiden,
- wie hoch die Kosten für den Sensoreinsatz sind,
- wie hoch die Wahrscheinlichkeit dafür ist, daß die gegenwärtige Hypothese stimmt und damit überhaupt einer Veränderung bedarf,
- wie hoch die Kosten für eine falsche Hypothese sind.

Bei Erkundungen der einzelnen Planquadrate eines Raumes wird in bestimmten Abständen geprüft, welcher Sensor nach obigem Katalog einzusetzen ist, um die Abweichung zwischen wahrer Position und intern als wahr angenommener Position möglichst gering zu halten. Ähnlich wird auch in [Miller 91] vorgegangen, um eine globale Umweltkarte zu erstellen, dabei wird jeweils lokal entschieden, welche Information an welchem Raumpunkt auswertbar ist, um in die globale Umweltkarte über-

[1] Zu dieser Problematik siehe auch [Jacubowicz 88].

nommen zu werden (eine ähnliche Fragestellung behandelt [Kweon 92; Kweon 92b]). Ein gegenüber diesen eher willkürlichen Methoden mathematisch fundierter, wenn auch noch recht vage formulierter Ansatz[1] wird dagegen in [Hager 87] vorgestellt. Die Problemstellung ist die oben bereits formulierte: Gegeben sei ein Sensor und eine spezifische Anfrage nach der Bereitstellung von Umweltinformation. Es sei ferner a–priori–Wissen über die Umwelt vorhanden, beispielsweise in der Form, daß Kanten immer von Ecken begrenzt werden und Ecken eine interessante Information für das System darstellen, weil sie ihrerseits Objekte begrenzen. Die Aufgabe besteht nun darin, eine optimale Strategie (einen optimalen Plan) dafür zu formulieren, aus welchen Blickwinkeln das Objekt (oder Teile der Umwelt) zu betrachten ist, um möglichst schnell und möglichst umfassend die gewünschte Information bereitstellen zu können. Es versteht sich von selbst, daß dies eine äußerst abstrakte Problemstellung ist. Zunächst stellt sich die Frage, wie überhaupt die Anfrage nach Information *formuliert* werden kann, wie also dem Planungsalgorithmus mitgeteilt werden soll, *wonach* zu suchen ist. Entsprechend ergibt sich die Frage nach der Bewertung der *Güte* der gelieferten Information. Eine korrespondierende Nutzenfunktion kann zwar für die mathematische Behandlung vorausgesetzt werden, wie diese aber zweckmäßig beschaffen sein muß, ist ein vollkommen ungeklärtes Problem. Schließlich ist offen, wie Vorwissen oder bereits erworbenene Information in den Planungsprozeß einbezogen werden kann. In [Hager 87] wird die Interaktion zwischen einer nachfragenden Instanz und dem Planer für die Sensorstrategie bzw. dem Sensor mit vier Größen modelliert: einer Nachfrage nach einem bestimmten *Typ* von Information, einer geforderten *Genauigkeit* Δ für die gelieferte Information, einer *Priorität* w (maximal zulässiger Zeitaufwand für ein bestimmtes Maß an Verbesserung der Information) und eine harte obere *Zeitschranke*, bis zu der die Information zur Verfügung stehen muß. Die Strategie wird als Reihe von Sensorpositionen oder Sensoreinstellungen $U = (U_1, U_2, \ldots, U_N)$ repräsentiert, zu der eine entsprechende Menge von Beobachtungen $X = (X_1, X_2, \ldots, X_N)$ gehört. Die Gesamtkostenfunktion $\mathrm{L}(\theta, \theta', N, U)$ für eine bestimmte Sensoreinsatzstrategie hängt dann von vier Parametern ab: dem wahren Umweltzustand θ, dem geschätzten Zustand[2] $\theta' = \delta(X)$, der Anzahl vorzunehmender Messungen N und der Einstellungssequenz U. Diese Funktion ist separierbar in zwei Teile: den Kosten für die Schätzung des wahren Umweltzustandes $\mathrm{L}(\theta, \theta', N, U)$ und den Kosten für die Durchführung der Strategie, insgesamt also: $\mathrm{L}(\theta, \theta', N, U) = \mathrm{L}^{\theta}(\theta, \theta') + \mathrm{L}^{U}(N, U)$. Je nach Wahl des Kriteriums für die Verlustfunktion[3] können nun unterschiedliche, dem Problem angepaßte Kostenfunktionen erzeugt werden (analog dem in 2.3 angegebenen Vorgehen). Um eine geeignete Strategie U zu finden, verweist [Hager 87] auf die einschlägigen Ergebnisse der Theorie zum Aufbau von Experimenten [Fedorov 72]. Hier wird dasselbe Problem behandelt: Das Experiment soll bei gegebenen Kosten so gestaltet werden, daß ein Maximum von Information gewonnen werden kann. Ein aussichtsreicher Weg scheint ferner die Übernahme von Erkenntnissen aus der Spieltheorie [Berger 88] zur Wahl einer günstigen Strategie zu sein. Ferner können Ergebnissen für optimale Stoppregeln [Liptser 79] übernommen werden, die den optimalen Zeitpunkt bestimmen, wann eine Beobachtungssequenz abgebrochen werden soll, um mit höchster Wahrscheinlichkeit das Maximum an Information gewinnen zu können. Die Realisierung eines geeigneten Anwendungsbeispieles scheint auch hier dringend geboten, bevor Aussagen über die

---

1 Zu dieser Veröffentlichung gibt es nach Kenntnis des Autors keine Nachfolgearbeit. Dies mag sich aus der Komplexität des Ansatzes erklären, ist aber auf jeden Fall bedauerlich.

2 $\delta(\cdot)$ bezeichnet den Entscheidungsalgorithmus, der X in θ' überführt.

3 Für $\mathrm{L}^{\theta}(\theta, \theta')$ kann zweckmäßigerweise die Einheitsverlustfunktion $\mathrm{L}(\theta, \theta') = \begin{cases} c & \text{für } \|\theta - \theta'\| < \Delta \\ w & \text{sonst} \end{cases}$ verwendet werden. In diesem Fall entspricht die Verlustfunktion der Wahrscheinlichkeit dafür, daß der Schätzwert außerhalb des Toleranzbereiches liegt.

praktische Realisierbarkeit, die Nützlichkeit und das Potential dieses Ansatzes getroffen werden können.

Schließlich werden auch Techniken der künstlichen Intelligenz für die Sensordatenfusion vorgeschlagen. Als aussichtsreich haben sich *blackboard*–Architekturen [Nii 86] erwiesen, bei denen verschiedene Wissensquellen ihre Beobachtungen und Erkenntnisse an eine „Tafel" (gemeinsame Datenbasis) schreiben. Dieses Wissen steht dann allen anderen Wissensquellen und übergeordneten Instanzen zur Verfügung[1]. Intern findet eine Bewertung des abgeleiteten Wissens mit Hilfe von *certainty–factors* [Shortliffe 84] statt (siehe dazu die Bemerkungen in [Cheeseman 85]). Verwendet werden diese Methoden zum Beispiel bei der Kombination der Ergebnisse verschiedener Sensoren im Bereich der Flugzeugerkennung und –verfolgung [Sikka 89; Chaudhuri 89], die Übertragbarkeit zumindest auf den Bereich mobiler Roboter ist offensichtlich (siehe z.B. [Belknap 86; Raczkowsky 91; Brooks 91]). Es würde den Rahmen dieser Arbeit sprengen, wenn wir auf die dem Bereich der KI zuzurechnenden Techniken eingehen würden, die bei der Szeneninterpretation verwendet werden. Es sei aber zumindest angedeutet, daß Techniken, bei denen Hypothesen über die Szene unter Berücksichtigung von Randbedingungen aufgestellt und verifiziert werden, sich für die Fusion eignen. Dabei können die von einem Sensor erhaltenen Meßergebnisse und daraus generierten Hypothesen als mehr oder weniger genau einzuhaltende Randbedingungen (*constraints*) für die Interpretation eines von einem anderen Sensor gelieferten Bildes benutzt werden (siehe [Hanson 88; Hu 87], eine Beschreibung des in diesem Zusammenhang zuweilen verwendeten *relaxation*–Algorithmus findet sich in [Ballard 82; Chen 92]). Die Auswertung von Randbedingungen wird auch in [Hager 92] zur Grundlage der Kombination von Sensordaten gemacht. Statt einer statistischen Modellierung der Sensorfehler wird dabei von einer Menge möglicher Meßwerte für jeden Sensor ausgegangen (was voraussetzt, daß die Meßfehler beschränkt sind). Die von einem anderen Sensor gelieferten Meßwerte schränken diese Menge ein, so daß die gelieferte Information genauer wird (in [Hager 90b] wird dabei auch die Interaktion mit einem menschlichen Benutzer in bestimmten Situationen vorgeschlagen). Grundlage der Interpretation ist ein parametriertes Umweltmodell; dessen Parameter werden im Zuge eines auf die Beobachtungen abgestützen Entscheidungsprozesses immer enger spezifiziert.

1.5 Zusammenfassende Bewertung

Es hat sich bei unseren Betrachtungen gezeigt, daß sich bei der Sensordatenfusion der statistische Ansatz als leistungsfähig erwiesen hat und deshalb auch der hauptsächlich bearbeitete ist. Das gilt unabhängig von den Anwendungen, also für mobile ebenso wie für stationäre Systeme oder Sensorsysteme in Luft– und Raumfahrt[2]. Umwelt– und Sensormodelle lassen sich hier ebenso mathematisch formulieren, wie der eigentliche Fusionsalgorithmus; es gibt bewährte Techniken aus der Statistik und der Nachrichtentechnik, die sich problemlos übertragen lassen und schließlich sind solche Methoden flexibel sowie vor allem *schnell* (daraus erklärt sich auch die häufige Verwendung des als Standard–Methode anzusehenden Kalman–Filters sowie seiner vielen Abwandlungen) und *einfach zu*

1 Man beachte hier die konzeptionelle Ähnlichkeit zu den am Beginn dieses Abschnitts beschriebenen Ansatz nach [Durrant–Whyte 90].

2 Zukünftig wird sicher auch die zusätzliche Verwendung von Konzepten aus dem Bereich der neuronalen Netze eine Rolle spielen, wenn sich die Probleme mit den normalerweise zu langen „Lernzyklen" und der Stabilität/Determiniertheit des Verhaltens eines solchen Netzwerkes lösen lassen; eine Ankündigung für ein entsprechendes Forschungsvorhaben findet sich in [Schürmann 91; Eckmiller 91].

implementieren. Die wesentlichen Arbeiten zur Übertragung von Methoden aus der Statistik und Entscheidungstheorie auf die bei der Fusion anstehenden Probleme wurden von Durrant–Whyte bzw. Hager und Mintz veröffentlicht. Die eher theoretisch angelegten Arbeiten von Hager zur effizienten Implementierung des Satzes von Bayes (erforderlich für die Implemtierung von Fehlermodellen, die nicht der Normalverteilung entsprechen) haben nach Kenntnis des Autors noch keine praktische Verwendung gefunden. Demgegenüber ist sowohl das von Durrant–Whyte verwendete Umweltmodell (welches allerdings von bekannten Objekten in ungefähr bekannter Lage ausgeht), seine Methode zur Überprüfung der Konsistenz zwischen Sensordaten und Objektbeziehungen sowie die von ihm benutzte Fusionsmethode des statischen Kalman–Filters (bzw. Maximum–a–posteriori–Schätzung) die allgemein anerkannte Grundlage für die statistische Datenfusion. Was den wichtigen Punkt der Konsistenzüberprüfung angeht, erwies sich bei den durchgeführten Simulationen die Konsistenzüberprüfung bei Durrant-Whyte gegenüber der von Luo zwar als etwas weniger leistungsfähig, dafür aber sehr viel robuster. Insbesondere ist kein Schwellwert festzulegen. Die Konsistenzüberprüfung nach [Luo 90] kann zwar bei optimal gewählten Schwellwerten zu besseren Schätzwerten führen, es zeigte sich aber, daß eine geeignete Schwellwertbestimmung im allgemeinen sehr schwierig ist.

Abgesehen von der Annahme von beliebigen (allerdings durch Treppenfunktionen zu approximierenden) Fehlerverteilungsdichten wird bei allen anderen Ansätzen von Normalverteilungen für die Meßfehler ausgegangen. Diese Approximation wird durch den zentralen Grenzwertsatz und durch die Behauptung gerechtfertigt, daß die Berechnungen bei anderen Verteilungsannahmen aufwendig werden und daß eine akzeptable Dauer der Berechnung der Schätzwerte nicht erreicht werden kann. Auch verteilungsfreie Verfahren (also solche Verfahren, die keine Annahme über eine gegebene Verteilung der Stichproben machen) sind nicht einsetzbar, da durch verteilungsfreie Methoden mehr Information verlorengeht als an Fehlern durch die möglicherweise nicht korrekte Annahme einer Normalverteilung hinzukommen; so wird beispielsweise bei der Methode der kleinsten Quadrate die Sensorvarianz nicht mit in die Berechnung einbezogen. Daraus folgt aber, daß sich die normalerweise unterschiedlichen Fusionsmethoden Kalman–Filter, Maximum–a–posteriori– und Maximum–Likelihood–Schätzung für statische Systeme nur noch darin unterscheiden, ob mit oder ohne a–priori Information gearbeitet wird. Bei einer relativ geringen Anzahl von Sensoren sollte infolge der möglichen sich lange „durchziehenden“ Auswirkungen von a–priori–Informationen immer die Maximum-Likelihood-Methode gewählt werden. Bei Vorhandensein zuverlässiger a–priori–Information sollte diese allerdings verwendet werden; so könnte bei Fusion mit dem Kalman–Filter diese Information als Startwert für Schätzwert und Kovarianzmatrix verwendet werden. Hier sind in jeder Beziehung Verbesserungen möglich.

Größere Projekte, die den Themenkreis bearbeiten, sind beschrieben in [Greenway 91] und in weiterem Sinne auch [Eckmiller 91; Schürmann 91]. Die Methoden zur Verarbeitung von redundanter Information stehen in engem Zusammenhang mit den Verfahren zur deren Gewinnung. Sie gehen bei den oben erwähnten Ansätzen zur automatisierten Plazierung von Sensoren zum Zwecke des höchsten Informationsgewinns bei der Fusion sogar ineinander über.

2 Mathematische Grundlagen der verwendeten Schätzverfahren

Im folgenden werden Grundbegriffe der Schätztheorie definiert und darauf aufbauend wichtige, im Rahmen der Sensordatenfusion verwendete Methoden übersichtsartig dargestellt: Die *Maximum–Likelihood–Methode*, die *Maximum–a–posteriori–(MAP)*–Schätzung, das *Kalman–Filter* und die Datenkombination nach der Methode von *Dempster–Shafer*.

2.1 Grundkonzepte

Bei der Sensordatenfusion soll aufgrund der verschiedenen Sensordaten ein möglichst guter Schätzwert für den unbekannten wirklichen Wert eines indirekt beobachteten Merkmals (oder Merkmalsparameters) bestimmt werden. Dazu wird ein bestimmter, plausibler Wert für das Merkmal angenommen (geschätzt) und diese Annahme (Schätzung) durch Hinzunahme von immer mehr Beobachtungen verändert bzw. bestätigt. In der Entscheidungstheorie wird aufgrund des geschätzten Wertes dann eine Aktion ausgeführt, also z.B. ein bestimmtes Produkt gekauft oder aber das beobachtete Merkmal mit Hilfe einer Klassifizierungsfunktion einer bestimmten *Klasse* zugewiesen. Es wird also eine Entscheidung für einen bestimmten Wert eines Merkmals getroffen, von dem angenommen werden kann, daß er dem wahren Wert des Merkmalsparameters entspricht oder ihm zumindest so nahe wie möglich kommt. Normalerweise sind die Informationen, anhand derer eine Entscheidung getroffen werden soll, gestört. Werden sie durch eine Messung ermittelt, treten Meßfehler auf. Um Entscheidungen auch bei vorhandenen Ungenauigkeiten treffen zu können, werden statistische Annahmen (Verteilungsfunktion, a–priori–Wissen, Mittelwert, Varianz) gemacht und diese in den Entscheidungsprozeß einbezogen.

Der zu schätzende wahre *„Naturzustand“* (im folgenden einfach mit *Zustand* bezeichnet) werde durch eine Größe θ beschrieben, die ein Skalar oder ein Vektor sein kann. Soll also beispielsweise die Lage einer Geraden in der Ebene geschätzt werden, dann würde diese Lage durch einen Vektor θ beschrieben, dessen eine Komponente die Steigung und die andere der Schnittpunkt mit der Ordinate des Bezugskoordinatensystems bestimmt. Unter Umständen sind diese beiden Komponenten nicht statistisch unabhängig voneinander, d.h. bei Änderung der einen Komponente ändert sich nach einer bestimmten Gesetzmäßigkeit auch die andere. Wenn die wahre Farbe eines Gegenstandes geschätzt werden soll, sind die drei Komponenten z.B. die Intensitäten von Rot, Grün und Blau.

Vielfach werden statistische Kenngrößen geschätzt: Es sei bekannt, daß der Umfang eines Nagels einer Normalverteilung $N(\mu, \sigma^2)$ entspreche. Die die Normalverteilung vollständig beschreibenden Parameter μ (Mittelwert) und σ^2 (Varianz) sind jedoch unbekannt. Sie sollen mit Hilfe von Stichproben geschätzt werden. In diesem Fall hat θ die beiden zu schätzenden Komponenten μ und σ^2.

Es wird vorausgesetzt, daß θ den wahren Naturzustand beschreibt, also alle interessierenden Kenngrößen des Naturzustandes in θ zusammengefaßt sind. Ferner sei es möglich, daß für alle Komponenten von θ eine Angabe (in Form einer Wahrscheinlichkeitsdichte) darüber gemacht werden kann, welche Ungenauigkeiten auftreten. Die zur Schätzung von θ erforderlichen Informationen werden durch Experimente (z.B. Umfragen) oder durch direkte Messung (physikalischer Größen) gewonnen. Das Ergebnis dieser Experimente kann durch die Eigenschaften einer Zufallsvariablen, die im folgenden mit X bezeichnet wird, beschrieben werden. Meistens handelt es sich bei X um eine

Menge $X = \{X_1, ..., X_N\}$, wobei die X_i verschiedene unabhängige Beobachtungen einer gemeinsamen Verteilung sind. X kann auch ein Vektor der Dimension K sein, dann bezieht sich die vorige Aussage auf die korrespondierenden Komponenten der N Vektoren.

2.2 Die Normalverteilung

In den meisten Fällen ist es schwierig, die bedingte Dichtefunktion $p(X|\theta)$ zu ermitteln, die bei vielen Schätzverfahren benötigt wird. Falls allerdings infolge der allgemeinen Beschreibung der Problemstruktur eine feste, parametrisierte Dichtefunktion angenommen werden kann, wird die Schätzung wesentlich vereinfacht. In der Sensordatenfusion wird fast immer davon ausgegangen, daß die Dichtefunktion eine Normalverteilung (auch als Gauß–Verteilung bezeichnet) mit Mittelwert μ und Varianz σ^2 ist. Damit wird die Bestimmung der Dichtefunktion auf die Bestimmung der Parameter μ und σ^2 reduziert. Im folgenden wird nur diese Art von Schätzproblemen behandelt. Das findet seine Rechtfertigung zum einen darin, daß die Abbildungsfunktion von erfaßter Größe nach Meßwert bei praktisch allen Sensoren mit einer Normalverteilung beschreibbar ist.

Im skalaren Fall wird eine Normalverteilungsdichte durch zwei Parameter spezifiziert: Mittelwert μ und Varianz σ^2. Man schreibt deshalb oft $p(X) \sim N(\mu, \sigma^2)$. Die Wahrscheinlichkeitsdichte ergibt sich zu:

$$\mathrm{p}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

Im mehrdimensionalen Fall erhalten wir für die Verbunddichtefunktion eines Zufallsvektors X mit K Komponenten $X = (X_1, X_2, ... , X_K)^T$ und der Realisierung $x = (x_1, x_2, ..., x_K)^\mathrm{T}$ die Verbundwahrscheinlichkeitsdichte:

$$\mathrm{p}_X(x) = p_{X_1,X_2,\ldots,X_K}(x_1,x_2,\ldots,x_K) = \frac{1}{\sqrt{(2\pi)^K}\sqrt{\det(\Sigma_X)}} e^{-\frac{1}{2}\left((x-\mu)^\mathrm{T} \Sigma_X^{-1} (x-\mu)\right)}$$

wo x ein K–dimensionaler Vektor ist, μ der zugehörige K–dimensionale Vektor der Erwartungswerte von X und Σ_X die $K \times K$ – Autokovarianzmatrix von X. Schließlich ist Σ_X^{-1} die Inverse dieser Matrix, $\det(\Sigma_X)$ ihre Determinante und $(x - \mu)^\mathrm{T}$ ist der zum Spaltenvektor $(x - \mu)$ transponierte Zeilenvektor. Bei der Multiplikation beachte man, daß für einen beliebigen Spaltenvektor X das Produkt $X^\mathrm{T}X$ ein Skalar, das Produkt XX^T aber eine quadratische Matrix ist. Dementsprechend läßt sich für die Kovarianzmatrix auch schreiben: $\Sigma_X = E[(x - \mu)(x - \mu)^\mathrm{T}]$. Wenn die einzelnen Komponenten X_i des Zufalls*vektors* X voneinander unabhängige Zufallsvariable sind, so sind nur die Elemente σ_{ii}^2 = cov $(X_i, X_i) = \mathrm{var}(X_i)$ in der Hauptdiagonalen von Σ_X von Null verschieden und die Verbunddichte degeneriert in das Produkt der Verteilungen der einzelnen Zufallsvariablen:

$$\mathrm{p}_X(x) = p_{X_1,X_2,\ldots,X_K}(x_1,x_2,\ldots,x_K) =$$

$$\frac{1}{\sqrt{(2\pi)^K}\sigma_1\sigma_2\cdots\sigma_K} e^{-\frac{1}{2}\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2} e^{-\frac{1}{2}\left(\frac{x_2-\mu_2}{\sigma_2}\right)^2} \cdots e^{-\frac{1}{2}\left(\frac{x_K-\mu_K}{\sigma_K}\right)^2}$$

Die mehrdimensionale Normalverteilung wird vollständig durch Mittelwertvektor μ und Kovarianzmatrix Σ_X spezifiziert. Stellt man die Realisierungen, die einer mehrdimensionalen Normalverteilung gehorchen, graphisch dar, so erhält man eine Gruppierung um den Mittelwertvektor μ herum, wobei die Dichte mit zunehmendem Abstand von μ abnimmt. Die Ortskurven gleicher Dichten erhält man, indem man $p_X(x)$ = konstant setzt und dies für X auswertet. Die Dichte ist konstant für konstanten Exponenten, also:

$$r^2 = (x-\mu)^{\mathrm{T}} \Sigma_X^{-1} (x-\mu)$$

Im Fall eines zweidimensionalen Vektors x beschreibt diese Beziehung die „Höhenlinien" der im $\mathbf{R}^3$ aufgezeichneten Dichtefunktion. Man bezeichnet diesen Ausdruck auch als quadrierten *Mahalanobis–Abstand* zwischen den Merkmalsvektoren x und μ. Einfach veranschaulichen läßt sich die Bedeutung dieses Maßes im skalaren Fall für $r = 1$. In diesem Fall wird $r^2 = 1 = (x - \mu)^2/\sigma_X^2$ oder $(x - \mu)^2 = \sigma_X^2$. Der Mahalanobis–Abstand einer Variablen x von einem Mittelwert μ wird also genau dann 1, wenn sich x genau um die Standardabweichung vom Mittelwert entfernt hat. Dies ist ein vernünftiger Grenzwert für eine Konsistenzüberprüfung, d.h. für die Entscheidung, ob ein bestimmter Sensorwert noch einer Klasse um den Vektor m zuzurechnen ist, oder nicht. Dieses Kriterium findet in [Durrant–Whyte 87] Verwendung (siehe auch [Bolle 86]).

2.3 Risiko, Verlustfunktion, Klassifikation

Oftmals ist es bei der Entscheidung für einen bestimmten Schätzwert hilfreich, die Risiken aller möglichen Entscheidungen mit zu bewerten, mithin also zu schätzen, welches Risiko man bei der Entscheidung für einen bestimmten Wert (und implizit des Verwerfens möglicher anderer) eingeht. Dieses Risiko wird durch den Erwartungswert einer sogenannten *Verlustfunktion* berechnet. Die Verlustfunktion $L(D, \theta)$ beschreibt für jede mögliche Entscheidung D und jedes mögliche θ den Verlust (die Kosten), die unter Annahme von θ als wahrem Wert durch eine Entscheidung D entsteht. Bei der Schätzung beruht die Entscheidung immer auf der Berechnung eines bestimmten geschätzten Wertes, den wir mit $\hat{\theta}$ bezeichnen wollen. Bei Annahme von θ als wahr entspricht das Risiko $R(\hat{\theta},\theta)$ der verschiedenen Entscheidungen für $\hat{\theta}$ dem Erwartungswert der Verlustfunktion $L(\hat{\theta},\theta)$, also $R(\hat{\theta},\theta) = E[L(\hat{\theta},\theta)]$. Das Risiko ist damit berechenbar, wenn eine Verteilungsfunktion von θ bekannt ist. Üblicherweise wird eine Verlustfunktion so definiert, daß sie einen größeren Wert annimmt, wenn der Schätzwert sich weiter vom wahren Wert entfernt. Beispielsweise ist eine häufig verwendete Verlustfunktion der quadratische Abstand $L(\hat{\theta}, \theta) = c(\theta)(\hat{\theta} - \theta)^2$, wo $c(\theta)$ eine beliebige Konstante bezüglich des Schätzwertes $\hat{\theta}$ ist.

Wir betrachten ein Beispiel [Berger 88]: Eine Firma möchte ein neues Produkt einführen und möglichst vorher genau wissen, welche Produktionszahl zu realisieren ist. Der sich tatsächlich ergebende Marktanteil ist $\theta \in [0,1]$. Die Produktionszahl $\hat{\theta}$, angegeben ebenfalls als Marktanteil und hinmit $\hat{\theta} \in [0,1]$, sollte so gut wie möglich geschätzt werden, im Idealfall sollte natürlich $\theta = \hat{\theta}$ sein. Es werde hier der Einfachheit halber eine lineare Verlustfunktion definiert, nämlich

$$L(\theta,\hat{\theta}) = \begin{cases} \theta - \hat{\theta} & \text{falls} \quad \theta \geq \hat{\theta} \\ -2(\theta - \hat{\theta}) & \text{falls} \quad \theta < \hat{\theta} \end{cases}$$

Das bedeutet nichts anderes, als daß eine Überproduktion als zweimal so teuer angenommen wird wie eine Unterproduktion. Ferner sei aus früheren Produkteinführungen bekannt, daß der Marktanteil sich gleichverteilt zwischen 10% und 20% bewegt, also einer Gleichverteilung $\pi(\theta) = 10 \cdot I_{(0,1;0,2)}(\theta)$ gehorcht (a–priori–Wahrscheinlichkeitsdichte). Damit ergibt sich für das Risiko:

$$R(\theta,\hat{\theta}) = E[L(\theta,\hat{\theta})] = \int_0^1 L(\theta,\hat{\theta})\,\pi(\theta)\,d\theta = \begin{cases} 0.15 - \hat{\theta} & \text{falls } \hat{\theta} \le 0.1 \\ 15 \cdot \hat{\theta}^2 - 4 \cdot \hat{\theta} + 0.3 & \text{falls } 0.1 \le \hat{\theta} \le 0.2 \\ 2 \cdot \hat{\theta} - 0.3 & \text{falls } \hat{\theta} \ge 0.2 \end{cases}$$

Das geringste Risiko ergibt sich beim Minimum der Risikofunktion, hier also für $\hat{\theta} = 2/15 \approx 13{,}3\%$. Dies wird als zu erwartender Marktanteil geschätzt und die Produktion entsprechend eingerichtet.

Wir wenden uns nun kurz dem *Klassifikationsproblem* zu, bei dem eine Beobachtung X einer bestimmten Klasse zugeordnet werden soll. Wir nehmen zunächst an, daß der wahre Zustand θ_i eines Merkmals einer aus der Menge von J möglichen Zuständen ist, also $\theta_i \in \{\theta_1, \theta_2, ..., \theta_J\}$. Die aus einer Beobachtung resultierende Merkmalsvariable X ist (wie θ_i) eine skalare Zufallsvariable oder ein K–dimensionaler Zufallsvektor. Aufgrund der Beobachtung(en) X soll eine Entscheidung getroffen werden, welches der wahre Zustand θ_i ist. Es gebe ferner für jedes i eine initiale (a–priori–)Wahrscheinlichkeit $\pi(\theta_i)$ dafür, daß θ_i der wahre Zustand ist und eine bekannte Verteilung $p(X|\theta_i)$, die die Wahrscheinlichkeit für die Beobachtung X angibt, wenn θ_i der wahre Zustand ist.

Zunächst wird die Beobachtung X dazu benutzt, eine a–posteriori–Wahrscheinlichkeit $\pi(\theta_i|X)$ für alle möglichen Zustände zu berechnen. Es gilt mit dem Satz von Bayes, wenn X die Beobachtung ist :

$$p(\theta_i \mid X) = \frac{p(X \mid \theta_i)\pi(\theta_i)}{\sum_{j=1}^{J} p(X \mid \theta_j)\pi(\theta_j)}$$

Dies ist die Wahrscheinlichkeit dafür, daß θ_i der wahre Zustand ist, wenn die Beobachtung X mit einbezogen wird. Wenn wir die Entscheidung für θ_j mit D_j bezeichnen, gibt eine geeignete Verlustfunktion $L(D_j, \theta_i)$ den Verlust an, der entsteht, wenn θ_i der wahre Zustand ist und die Entscheidung D_j (also für θ_j) getroffen wird (häufig verwendete Verlustfunktionen sind die z.B. bereits erwähnte quadratische Verlustfunktion oder die weiter unten angegebene Einheitsverlustfunktion). Das Risiko für alle möglichen Entscheidungen, das sogenannte *Bayes–Risiko*, kann damit angegeben werden zu:

$$R(D_i, X) = E[L(D_i,\theta_i)] = \sum_{j=1}^{J} L(D_i,\theta_j)\,p(\theta_j \mid X)$$

Bei Klassifikationsproblemen soll eine Beobachtung X (die bei einem wirklich vorliegenden Zustand θ_i gemacht wird) einer Klasse i zugeordnet werden. Es wird also aufgrund einer Beobachtung X eine Entscheidung D_i gefällt, die X der mit θ_i assoziierten Klasse i zuweist, weil θ_i als wahrer Zustand vermutet wird. Wenn die Entscheidung D_i getroffen wird und θ_i der wahre Zustand ist, ist die

Entscheidung korrekt. Wenn die Entscheidung für D_j fällt, obwohl der θ_i der wahre Zustand ist, ist die Entscheidung falsch. Die hier zur Bewertung der Entscheidung normalerweise herangezogene *Einheitsverlustfunktion* ist:

$$\mathrm{L}(\mathrm{D}_i,\theta_i)=\begin{cases}1 & \text{für} \quad i=j\\ 0 & \text{für} \quad i\neq j\end{cases}$$

Damit entstehen bei einer richtigen Klassifikation keine, bei allen falschen Entscheidungen entstehen gleich hohe Kosten. Mit dieser Verlustfunktion nimmt das Bayes–Risiko die einfache Form an: $R(D_i, X) = 1 - p(\theta_i|X)$. Dies ist kein überraschendes Ergebnis: Die Entscheidung mit dem geringsten Risiko ist demnach die, für die die a–posteriori–Wahrscheinlichkeit, daß i der tatsächliche Zustand ist, maximal ist. Anders ausgedrückt: Die Entscheidung ist für die Klasse i zu treffen, für die gilt $p(\theta_i|X) > p(\theta_j|X)$ für alle $i \neq j$. Bei der Klassifikation ist also wie folgt vorzugehen: Mit jeder Klasse i wird eine Entscheidungsfunktion $g_i(X)$ assoziiert. Bei J Klassen hat man damit J solcher Funktionen $g_1(X)...g_J(X)$. Alle diese Funktionen werden auf einen beobachteten und zu klassifizierenden Merkmalsvektor X angewendet und die Entscheidung für die Klasse i getroffen, für die der Wert der Funktion größer ist als für alle anderen, also $g_i(X) > g_j(X)$ für alle $i \neq j$. Im Falle des Bayes–Risikos ist $g_i(X) = p(\theta_i|X)$ proportional zur a–posteriori–Wahrscheinlichkeit, daß i die richtige Klasse ist. Damit gilt:

$$\mathrm{g}_i(X)=\mathrm{p}(\theta_i \mid X)=\frac{\mathrm{p}(X \mid \theta_i)\pi(\theta_i)}{\sum_{j=1}^{J}\mathrm{p}(X \mid \theta_j)\pi(\theta_j)}$$

Da nur das Verhältnis der einzelnen Entscheidungsfunktionen zueinander interessiert, der Nenner auf der rechten Seite aber für alle Funktionen $g_j(X)$ gleich ist, ergeben sich mit

$$\mathrm{g}_i(X)=\mathrm{p}(X \mid \theta_i)\pi(\theta_i)$$

die gleichen Klassifikationsresultate. Das gilt auch für die logarithmierte Version

$$\mathrm{g}'_i(X)=\log(\mathrm{p}(X \mid \theta_i))+\log(\pi(\theta_i))$$

die sich insbesondere im Zusammenhang mit Gauß–Verteilungen anbietet, da die Exponentialfunktion dann verschwindet.

Diesen Fall betrachten wir nun noch näher: Wir suchen für den Fall der Normalverteilung einen Satz von Entscheidungsfunktionen. Es sei ein K–dimensionaler Vektor X zu klassifizieren und J Klassen i vorhanden. Alle Klassen i haben eine zugehörige Dichte $p(X,\theta_i)$, jeweils mit $p(X,\theta_i) \sim N(\mu_i,\Sigma_{X_i})$, d.h. einer Normalverteilung mit Mittelwertvektor μ_i und Kovarianzmatrix Σ_{X_i}. Mit der logarithmierten Version der Entscheidungsfunktionen ergibt sich:

$$\mathrm{g}'_i(X)=-\frac{1}{2}\left(x-\mu_i\right)^{\mathrm{T}}\Sigma_{X_i}^{-1}\left(x-\mu_i\right)-\frac{1}{2}\log(\det(\Sigma_{X_i}))+\log(\pi(\theta_i))$$

wobei der Term –K/2·log(2π) nicht angeschrieben wurde, da es sich bei ihm nur um eine für alle Klassen gleichermaßen zutreffende additive Konstante handelt. Wir untersuchen zwei Spezialfälle: Zum einen die statistische Unabhängigkeit der Merkmale des Merkmalsvektors X, zum anderen identische Kovarianzmatrizen Σ_{X_i} für alle Klassen i.

I) Statistische Unabhängigkeit der Merkmale. Hier wird davon ausgegangen, daß alle Merkmale in X unabhängig voneinander sind und die gleiche Varianz σ^2 aufweisen. Ein Beispiel dafür ist die Vermessung eines Würfels in allen drei Raumrichtungen und Einteilung der vermessenen Würfel in „groß“ und „klein,“ je nach Ähnlichkeit zum Mittelwertvektor μ_i (Prototypvektor) der beiden Klassen. Man kann davon ausgehen, daß bei Verwendung desselben Meßwerkzeuges dieselben Varianzen für alle drei Merkmale auftreten. In diesem Fall gilt

$$g_i(X) = -\frac{1}{2\sigma^2}\left(x-\mu_i\right)^{\mathrm{T}}\left(x-\mu_i\right)+\log(\pi(\theta_i))$$

Dementsprechend ist jeweils nur der euklidische Abstand zwischen Merkmalsvektor und Mittelwertvektor für die jeweilige Klasse zu berechnen und dann die Entscheidung zu fällen. Es ist leicht einzusehen, daß im Falle von zwei Klassen gleicher a–priori–Wahrscheinlichkeit die „Entscheidungsgrenze“ je nach Dimension des Merkmalsvektors eine Linie, Ebene oder Hyperebene senkrecht auf dem Differenzenvektor $\mu_1 - \mu_2$ ist.

II) Gleiche Kovarianzmatrix für alle Klassen: Σ_{X_i}=Σ_X. Ein Beispiel für diesen Fall ist die Klassif ängigkeit von der Farbe des Objekts). Man bekommt hier für den Satz von Entscheidungsfunktionen:

$$g_i(X) = -\frac{1}{2\sigma^2}\left(x-\mu_i\right)^{\mathrm{T}}\Sigma_X^{-1}\left(x-\mu_i\right)+\log(\pi(\theta_i))$$

Im Falle gleicher a–priori–Wahrscheinlichkeiten ist die Entscheidungsregel wieder einfach: Gemessen wird der quadrierte Mahalanobis–Abstand zum jeweiligen Mittelwertvektor und der beobachtete Merkmalsvektor wird der Klasse mit dem geringsten Abstand zugeschlagen.

2.4 Die Güte von Schätzmethoden

Beim Einsatz einer bestimmten Methodik zur Schätzung interessiert vor allem deren Schätzqualität, d.h. die Fähigkeit, unter allen Randbedingungen ein Resultat $\hat{\theta}$ zu liefern, das dem wahren Wert des zu schätzenden Parametervektors θ möglichst nahe kommt. Diese Qualität kann anhand bestimmter Kriterien näher spezifiziert werden.

2.4.1 Erwartungstreue

Zunächst sollte eine Schätzmethode in der Lage sein, für alle möglichen Stichprobenrealisierungen einen geeigneten Schätzwert für den Parametervektor zu ermitteln. Deshalb beziehen sich alle Güte-kriterien auf den Fehler $\theta - \hat{\theta}$, der im Idealfall Null sein sollte. Es wird also von einer Schätzmethode verlangt, daß der Erwartungswert des Schätzfehlers Null ist. Allgemein formuliert ergibt sich aus dieser Überlegung das *Kriterium der Erwartungstreue*: Eine Schätzmethode, die den Schätzwert $\hat{\theta}$ liefert, wird als erwartungstreu für den Parametervektor θ bezeichnet, falls bei Vorliegen von θ der Erwartungswert für die Schätzung $\hat{\theta}$ mit θ übereinstimmt, also $\mathrm{E}[\hat{\theta}|\theta] = \theta$ gilt.

Beispiel: Sollen Mittelwert μ und Varianz σ^2 einer Normalverteilung $\mathrm{N}(\mu, \sigma^2)$ anhand von N Stichproben $\{x_1, ..., x_N\}$ geschätzt werden, handelt es sich bei der Verwendung vom Stichprobenmittelwert

$$\bar{x} = \frac{1}{N}\sum_{k=1}^{N} x_k$$

um eine erwartungstreue Schätzmethode, denn

$$\mathrm{E}[\bar{x} \mid \mu;\sigma^2] = \frac{1}{N}\sum_{k=1}^{N} E[x_k \mid \mu;\sigma^2] = \mu$$

2.4.2 Konsistenz

Für die Schätzung von μ ist auch die Übernahme einer beliebigen einzelnen Beobachtung als Schätzwert eine erwartungstreue Methode. Die Erwartungstreue reicht somit als Gütekriterium allein nicht aus, denn die Schätzgenauigkeit sollte sich verbessern, wenn der Stichprobenumfang erhöht wird. Daraus läßt sich ein weiteres Kriterium ableiten: Der Schätzwert sollte mit zunehmender Stichprobenanzahl gegen den wirklichen Wert des Parameters konvergieren. Es sei N wieder die Anzahl der Stichproben und $\hat{\theta}_N$ der anhand dieser N Stichproben bestimmte Schätzwert. Eine Schätzmethode wird als *konsistent* bezeichnet, falls mit Wahrscheinlichkeit 1 bei zunehmender Stichprobenanzahl der Schätzfehler kleiner als eine beliebige Schranke $\varepsilon \in \mathbf{R}^+$ wird, also:

$$\lim_{N\to\infty} P(|\hat{\theta} - \theta| < \varepsilon) = 1$$

Der oben betrachtete Stichprobenmittelwert ist offensichtlich eine konsistente Schätzmethode für den Mittelwert der Normalverteilung.

2.4.3 Varianz des Schätzfehlers

Als drittes Gütekriterium wird die *Varianz des Schätzfehlers* einer Schätzmethode verwendet. Seien $\hat{\theta}_1$ und $\hat{\theta}_2$ Schätzwerte zweier erwartungstreuer, aber unterschiedlicher Schätzmethoden für den Parametervektor θ. Als bessere der beiden Schätzmethoden wird diejenige angesehen, bei der die Varianz $\mathrm{E}[(\hat{\theta} - \theta)^{\mathrm{T}}(\hat{\theta} - \theta)]$ aller geschätzten Parameter kleiner ist, als bei der anderen. Eine Schätzmethode, die dieses Kriterium erfüllt, heißt erwartungstreue Methode mit minimaler Varianz (*Minimum Variance Unbiased Estimator, MVUE*).

2.5 Schätzmethoden

In den folgenden Abschnitten werden zunächst zwei in der Sensordatenfusion häufig eingesetzte Schätzmethoden beschrieben: die Maximum–Likelihood–Methode und die Maximum–a–posteriori–Methode, beides klassische statistische Verfahren. Obwohl beide in vielen Fällen ähnliche Ergebnisse liefern, sind sie konzeptionell unterschiedlich:

a) Bei der Maximum–Likelihood–Methode werden die Parameter in θ als feste Größen angesehen, deren Werte unbekannt sind. Die Schätzung wird aufgrund einer Beobachtung (Stichprobe) X durchgeführt. Als beste Schätzung für θ wird dann der Wert definiert, bei dessen Vorliegen die Wahrscheinlichkeit $p(X|\theta)$ am größten ist, daß man die Beobachtung X erhält.

b) Bei der Schätzung nach der Maximum–a–posteriori–Methode wird der Parameter als Zufallsvariable angesehen, der eine Verteilungsdichte $p(X|\theta)$ und eine a–priori–Wahrscheinlichkeit $\pi(\theta)$ zugeordnet ist. Mit Hilfe der Beobachtung X und dieser a–priori–Wahrscheinlichkeit wird über den Satz von Bayes eine *a–posteriori*–Wahrscheinlichkeit $p(\theta|X)$ berechnet. Basierend auf dieser und den gegebenen Werten der Verlustfunktion für jeden möglichen Schätzwert wird dann das *Bayes–Risiko* bestimmt. Als Schätzung wird der Wert mit dem kleinsten Risiko genommen.

2.5.1 Die Maximum–Likelihood–Methode

Wir gehen davon aus, daß eine Menge von N *statistisch unabhängigen* Stichproben $\{X_1, ..., X_N\}$ entsprechend der Dichtefunktion $p(X|\theta)$ ermittelt wurden. Die X_i können dabei Skalare oder Vektoren sein, ebenso der zu schätzende Zustand θ. Das Problem besteht darin, die durch die Stichproben gegebene Information zu nutzen, einen guten Schätzwert für θ zu finden. Es wird vorausgesetzt, daß die Verteilung $p(X|\theta)$ bekannt ist und durch den Parametervektor θ vollständig bestimmt wird; so könnte für skalare X die Dichte $p(X|\theta)$ normalverteilt sein mit Mittelwert μ und Varianz σ^2. Diese Größen sind zu schätzen, also würden die Komponenten von θ in diesem Fall μ und σ^2 sein. Alle Stichproben sind dann gemäß einer Normalverteilung mit Mittelwert μ und Varianz σ^2 verteilt. Wenn die N Stichproben Vektoren der Dimension K sind, besteht der Parametervektor aus dem Mittelwertvektor μ der Dimension K und den relevanten Elementen der $K \times K$–Kovarianzmatrix Σ_X. Da die einzelnen Stichproben unabhängig voneinander sind, kann $p(X|\theta)$ als Produkt der Einzeldichten bestimmt werden:

$$p(X \mid \theta) = \prod_{i=1}^{N} p(X_i \mid \theta)$$

Wird $p(X|\theta)$ als Funktion von θ aufgefaßt, bezeichnet man $p(X|\theta)$ als *Likelihood–Funktion* der Stichproben. Als Schätzwert $\hat{\theta}$ der Maximum–Likelihood–Methode wird der Wert definiert, für den $p(X|\theta)$ am größten ist. Die Methode kann damit wie folgt definiert werden [Bamberg 80]: Zu einer festen Stichprobenrealisation $\{X_1, ..., X_N\}$ ist der Schätzwert $\hat{\theta}$ so zu bestimmen, daß gilt

$$p(X \mid \hat{\theta}) \geq p(X \mid \theta)$$

für alle als möglich erachteten Werte von θ. Offensichtlich entspricht dieser Schätzwert dem Wert, der in einem gewissen Sinne am besten mit den beobachteten Stichproben übereinstimmt.

Da in vielen Fällen die Dichtefunktion einer Normalverteilung entspricht, ist es häufig leichter, den Logarithmus der Likelihood–Funktion zu benutzen. Infolge der strengen Monotonie des Logarithmus ist jener der Likelihood–Funktion für denselben Wert von θ' maximal, für den auch die Likelihood–Funktion maximal wird. Handelt es sich bei $p(X|\theta)$ um eine stetige, differenzierbare Funktion, kann ihr Maximum bei θ' durch einfaches Differenzieren und Nullsetzen berechnet werden. Es sei θ ein M–dimensionaler Spaltenvektor $\theta = (\theta_1\ \theta_2\ ...\ \theta_M)^T$ und

$$\nabla_\theta = \left(\frac{\partial}{\partial\theta_1} \quad \frac{\partial}{\partial\theta_2} \quad \cdots \quad \frac{\partial}{\partial\theta_M}\right)^T$$

der Gradientenoperator von θ, also der Operator, der bei Anwendung auf θ alle partiellen Ableitungen bildet. Ferner bezeichne $l(\theta) = \log(p(X|\theta))$ den Logarithmus der Likelihood–Funktion ($l(\theta)$ wird als *log–likelihood*–Funktion bezeichnet). Diese Funktion wird für das oben angeschriebene Produkt der Einzeldichten der N Stichproben:

$$l(\theta) = \log(p(X \mid \theta)) = \log\left(\prod_{i=1}^{N} p(X_i \mid \theta)\right) = \sum_{i=1}^{N} \log(p(X_i \mid \theta))$$

wobei dies M gleichartige Gleichungen für jede Komponente des Parametervektors darstellt. Für jede Komponente von θ wird die Bedingung für den Schätzwert θ', daß die Ableitung der log–likelihood–Funktion Null wird, also

$$\mathrm{grad}_\theta(l(\theta))\Big|_{\theta=\theta'} = \nabla_\theta\, l(\theta)\Big|_{\theta=\theta'} = 0$$

wobei für $\mathrm{grad}_\theta(l(\theta))$ gilt:

$$\mathrm{grad}_\theta(l(\theta)) = \nabla_\theta\, l(\theta) = \sum_{i=1}^{N} \nabla_\theta \log(p(X_i \mid \theta))$$

Dies sind wieder M gleichartige Gleichungen für jede Komponente des Parametervektors, die den Maximum–Likelihood–Schätzwert für θ vollständig bestimmen.

Als Beispiel für eine Maximum–Likelihood–Schätzung wird der wichtige Fall einer normalverteilten Dichtefunktion mit unbekanntem Mittelwert μ und unbekannter Varianz σ^2 betrachtet. Der zu schätzende Parametervektor θ besteht hier aus M = 2 Komponenten, nämlich $\theta_1 = \mu$ und $\theta_2 = \sigma^2$. Für die Likelihood–Funktion ergibt sich

$$p(X_i \mid \theta) = \frac{1}{\sqrt{2\pi\theta_2}}\, e^{-\frac{(x_i-\theta_1)^2}{2\theta_2}}$$

und für die zugehörige log–likelihood–Funktion:

$$\log(\mathrm{p}(X_i \,|\, \theta)) = -\frac{1}{2}\log(2\pi\theta_2) - \frac{1}{2\theta_2}(x_i - \theta_1)^2$$

Mithin hat man für den Gradienten:

$$\mathrm{grad}_\theta(\theta) = \nabla_\theta \log(\mathrm{p}(X_i \,|\, \theta)) = \begin{bmatrix} \frac{1}{\theta_2}(x_i - \theta_1) \\ -\frac{1}{2\theta_2} + \frac{(x_i - \theta_1)^2}{2\theta_2^2} \end{bmatrix}$$

und also die Bedingungen für die Berechnung der Schätzwerte

$$\sum_{i=1}^{N} \frac{1}{\hat{\sigma}^2}(x_i - \hat{\mu}) = 0$$

und

$$\sum_{i=1}^{N} \frac{(x_i - \hat{\mu})^2}{2\hat{\sigma}^4} - \frac{1}{2\hat{\sigma}^2} = 0$$

wobei $\hat{\mu}$ und $\hat{\sigma}^2$ die Maximum–Likelihood–Schätzwerte für θ_1 und θ_2 bezeichnen. Nach einigen Umformungen berechnen sich die Maximum–Likelihood Schätzwerte für μ und σ^2 zu

$$\hat{\mu} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

und

$$\hat{\sigma}^2 = \frac{1}{N}\sum_{i=1}^{N} (x_i - \hat{\mu})^2$$

Diese Schätzwerte entsprechen dem Stichprobenmittelwert und der Stichprobenvarianz. Im mehrdimensionalen Fall ergeben sich mit den gleichen Überlegungen die folgenden Gleichungen zur Bestimmung der Maximum–Likelihood Schätzwerte für den Mittelwertvektor μ und die Kovarianzmatrix Σ_X [Duda 73]:

$$\hat{\mu} = \frac{1}{N}\sum_{i=1}^{N} x_i \quad \text{und} \quad \hat{\Sigma}_X = \frac{1}{N}\sum_{i=1}^{N} (x_i - \hat{\mu})(x_i - \hat{\mu})^{\mathrm{T}}$$

Hier ist wieder der Schätzwert für den Mittelwertvektor der Mittelwertvektor der Stichprobe und der Schätzwert für die Kovarianzmatrix (die die Abhängigkeiten der einzelnen Komponenten der Stichprobenvektoren widerspiegelt) der arithmetische Mittelwert der N Matrizen $\left(x_i - \hat{\mu}\right)\left(x_i - \hat{\mu}\right)^{\mathrm{T}}$

Als zweites Beispiel wird eine Maximum–Likelihood–Schätzung für eine in der Sensordatenfusion typische Problemstellung betrachtet. Unter der Annahme einer Normalverteilung soll durch die N Stichprobenrealisationen x_i mit stichprobenabhängigen, aber bekannten Varianzen σ_i^2 ein Schätzwert für den unbekannten Mittelwert μ bestimmt werden. Hier ist also nur ein Parameter $\theta = \mu$ zu schätzen. Der natürliche Logarithmus der Likelihood–Funktion wird durch

$$\log(\mathrm{p}(X_i \mid \theta)) = \log\left(\frac{1}{\sqrt{2\pi\sigma_i^2}} e^{-\frac{(x_i-\theta)^2}{2\sigma_i^2}}\right) = -\frac{1}{2}\log(2\pi\sigma_i^2) - \frac{1}{2\sigma_i^2}(x_i - \theta)^2$$

gegeben. Für die Ableitung ergibt sich

$$\frac{\partial}{\partial\theta} \log \mathrm{p}(X_i \mid \theta) = \frac{1}{\sigma_i^2}(x_i - \theta)$$

und als Bedingung für den Schätzwert

$$\sum_{i=1}^{N} \frac{1}{\sigma_i^2}\left(x_i - \hat{\mu}\right) = 0$$

wobei $\hat{\mu}$ der Maximum–Likelihood–Schätzwert für θ ist. Damit kann der Maximum–Likelihood–Schätzwert für μ durch die Beziehung

$$\hat{\mu} = \frac{\sum_{i=1}^{N}\left(\sigma_i^2\right)^{-1} x_i}{\sum_{i=1}^{N}\left(\sigma_i^2\right)^{-1}}$$

berechnet werden. Ein analoges Ergebnis für den Schätzwert des Mittelwertvektors μ ergibt sich im mehrdimensionalen Fall (d.h. jede der N Stichprobenwerte x_i besteht aus einem Vektor mit Kovarianzmatrix $\boldsymbol{\Sigma}_{X_i}$):

$$\hat{\mu} = \left(\sum_{i=1}^{N} \Sigma_{X_i}^{-1}\right)^{-1} \sum_{i=1}^{N} \Sigma_{X_i}^{-1} x_i$$

2.5.2 Bayessche Methoden

Auch hier wird von einer Menge von N unabhängigen Stichproben $\{X_1, ..., X_N\}$, die entsprechend der Dichtefunktion $p(X|\theta)$ ermittelt wurden, ausgegangen. Die Werte von θ werden wieder als unbekannt und fest vorausgesetzt. Die Dichte $p(X|\theta)$ wird erneut als bekannt vorausgesetzt und durch θ eindeutig bestimmt. Darüberhinaus existieren für jede Komponente von θ Vermutungen in Form einer sogenannten *a–priori–Verteilungsdichte* $\pi(\theta)$ über die Wahrscheinlichkeiten dafür, daß ein bestimmter Wert der Komponente von θ der zutreffende ist. In der Möglichkeit der Hereinnahme von Vorwissen besteht der wesentliche Vorteil gegenüber der Maximum–Likelihood–Methode (wenn kein Vorwissen vorhanden ist, ist $\pi(\theta)$ gleichverteilt über den gesamten Merkmalsraum und die Bayessche Methode entspricht der Maximum–Likelihood–Methode). Weiterhin ist eine Verlustfunktion $L(\theta', \theta)$ gegeben, die für jeden möglichen Schätzwert θ' von θ definiert ist. Die Methode wird hier für einen Parametervektor θ mit kontinuierlichen Komponenten beschrieben. Für einen diskreten Parametervektor müssen die auftretenden Integrale sinngemäß durch Summen ersetzt werden.

Das Problem besteht zunächst darin, aus den vorhandenen Informationen (also X, $p(X|\theta)$, und $\pi(\theta)$) die sogenannte *a–posteriori–Dichtefunktion* $p(\theta|X)$ zu bestimmen. Der Name rührt daher, daß die a–posteriori–Dichtefunktion das Wissen repräsentiert, *nachdem* die Daten angesehen worden sind. Sie kann demnach als „Verbesserung“ der a–priori–Dichtefunktion hinsichtlich der aufgetretenen Stichprobenrealisierungen interpretiert werden. Mit dem Satz von Bayes ergibt sich die a–posteriori Dichtefunktion $p(\theta|X)$ zu

$$p(\theta \mid X) = \frac{p(X \mid \theta)\pi(\theta)}{p(X)}$$

wo die Bezugsdichte $p(X)$ durch Integration über den gesamten Wertebereich von θ

$$p(X) = \int_\theta p(X \mid \theta)\pi(\theta)\, d\theta$$

bestimmt wird. Da die N Stichproben X_i nach Voraussetzung unabhängig sind, läßt sich $p(X|\theta)$ berechnen zu

$$p(X \mid \theta) = \prod_{i=1}^{N} p(X_i \mid \theta)$$

Es soll, wie üblich, angenommen werden, daß die gegebene Verlustfunktion größer wird, je weiter der Schätzwert θ' von θ entfernt ist. Je nach Wahl der Verlustfunktion erhält man unterschiedliche Schätzverfahren. Das sogenannte *Bayes–Risiko* B berechnet sich als Erwartungswert dieser Verlustfunktion:

$$B(\theta') = E_\theta[L(\theta',\theta)] = \int_\theta L(\theta',\theta)\, p(\theta) d\theta$$

Nun ist aber p(θ) nur als a–priori–Wahrscheinlichkeitsdichte $\pi(\theta)$ bekannt. Wünschenswert ist es deshalb, statt p(θ) die a–posteriori–Wahrscheinlichkeit p($\theta|X$) in die Rechnung einzubeziehen. Man erhält dann das auf *X bedingte Bayes–Risiko*:

$$\mathrm{B}(\hat{\theta}|\,X) = \mathrm{E}_\theta[\mathrm{L}(\hat{\theta},\theta)\,|\,X] = \int_\theta \mathrm{L}(\hat{\theta},\theta)\,\mathrm{p}(\theta|\,X)\mathrm{d}\theta$$

wobei das Integral über den gesamten Wertebereich von θ zu erstrecken ist. Diese Beziehung gibt das Risiko an für die Wahl eines Schätzwertes für θ unter Berücksichtigung der N bei der Messung erhaltenen Stichprobenrealisationen (die sich eben im Verlauf der a–posteriori–Wahrscheinlichkeitsdichte p($\theta|X$) widerspiegelt). Sie muß erst noch durch eine geeignete Verlustfunktion ergänzt werden, welche angibt, wie falsche Schätzungen zu bewerten sind. In jedem Fall wird aber als Bayessche Schätzung für θ unter der Beobachtung X der Wert $\hat{\theta}$ definiert, für den das Bayessche Risiko am kleinsten ist. Nachdem für *jede beliebige Stichprobe* immer der Schätzwert gewählt wird, der die bedingte Risikofunktion B($\hat{\theta}|X$) minimiert, ist auch das Gesamtrisiko B($\hat{\theta}$) minimal, wenn das bedingte Risiko minimal ist. Das Gesamtrisiko ist der Erwartungswert des Risikos bei Variation der Stichproben über den gesamten möglichen Bereich für die Stichproben, das entspricht $\mathrm{E}_X[\mathrm{B}(\hat{\theta}|X)]$. Um $\mathrm{B}(\hat{\theta}) = \mathrm{E}_X[\mathrm{B}(\hat{\theta}|X)]$ zu minimieren, genügt es also immer, B($\hat{\theta}$|X) zu minimieren. Dies geschieht üblicherweise durch Differentiation von B($\hat{\theta}$|X) nach $\hat{\theta}$ und Nullsetzen, also

$$\mathrm{grad}_{\hat{\theta}}(\mathrm{B}(\hat{\theta})) = \nabla_{\hat{\theta}}\mathrm{B}(\hat{\theta}) = 0$$

Das Bayes–Risiko läßt sich also immer anhand einer konkreten Stichprobe berechnen und der damit ausgewählte Schätzwert ist im Sinne einer (frei wählbaren) Verlustfunktion optimal.

Für jede Anwendung lassen sich durch die Wahl einer Verlustfunktion die vom Schätzwert geforderten Eigenschaften spezifizieren. Alle dabei erzeugten Schätzverfahren gehören der Klasse der Bayesschen Verfahren an, weil sie mit a–priori–Wissen arbeiten, mit Hilfe dessen eine a–posteriori–Wahrscheinlichkeit berechnet wird. Die Auswahl des Schätzwertes geschieht immer durch Minimierung der Risiko–Funktion. Wir betrachten im folgenden drei gebräuchliche Verfahren: Die *Maximum–a–posteriori–(MAP)*–Methode, die Schätzung mit *quadratischer Kostenfunktion* und die Methode der *minimalen Varianz*.

2.5.2.1 Die Maximum–a–posteriori–(MAP)–Methode

Die Verlustfunktion ist hier die Einheitsverlustfunktion

$$\mathrm{L}(\hat{\theta},\theta) = \begin{cases} 1 & \text{für } \theta \neq \hat{\theta} \\ 0 & \text{für } \theta = \hat{\theta} \end{cases}$$

wobei dies für jede Komponente von θ gilt. Zur Herleitung des Schätzwertes führen wir die modifizierte Verlustfunktion

$$L^*(\hat{\theta},\theta) = \begin{cases} 1 & \text{für } |\theta - \hat{\theta}| > \varepsilon \\ 0 & \text{für } |\theta - \hat{\theta}| \leq \varepsilon \end{cases}$$

ein, wobei ε sehr klein sein soll und offensichtlich $L^* \to L$ für $\varepsilon \to 0$. Unter Benutzung der modifizierten Verlustfunktion folgt für das bedingte Bayes–Risiko sofort:

$$\begin{aligned} B(\hat{\theta}|X) &= \int_\theta L(\hat{\theta},\theta)\, p(\theta|X)\, d\theta \\ &= \int_\theta p(\theta|X)\, d\theta - \int_{\hat{\theta}-\varepsilon}^{\hat{\theta}+\varepsilon} p(\theta|X)\, d\theta \end{aligned}$$

Das links stehende Integral repräsentiert dabei den Bereich, in dem die Verlustfunktion den Wert Eins aufweist. Davon abgezogen wird der Wert des Integral für die Region, in der die Verlustfunktion den Wert Null hat und keinen Beitrag zum Risiko leistet. Das linke Integral ist Eins, denn es repräsentiert die Wahrscheinlichkeit dafür, daß θ irgendeinen Wert aus seinem gesamten Wertebereich aufweist. Damit wird die Bedingung für den minimalen Wert des Risikos in diesem Fall:

$$\operatorname{grad}_{\hat{\theta}}(B(\hat{\theta}|X)) = -\operatorname{grad}_{\hat{\theta}}\left(\int_{\hat{\theta}-\varepsilon}^{\hat{\theta}+\varepsilon} p(\theta|X)\, d\theta \right) = 0$$

Nun gilt mit dem Hauptsatz der Integralrechnung für $\varepsilon \to 0$ und also für $L^* \to L$ (wir betrachten nun nur eine Komponente von θ, die Überlegungen gelten aber für alle anderen Komponenten ebenso):

$$\int_{\hat{\theta}-\varepsilon}^{\hat{\theta}+\varepsilon} p(\theta|X)\, d\theta \approx 2\varepsilon\, p(\hat{\theta}|X)$$

und dementsprechend wird die Forderung für eine Nullstelle des Gradienten der Risikofunktion:

$$\operatorname{grad}_{\hat{\theta}}(p(\hat{\theta}|X)) = 0$$

Mit anderen Worten: Wird als Verlustfunktion die Einheitsverlustfunktion angenommen, dann wird das Kriterium für die Wahl des Schätzwertes die Suche nach dem *Maximum der a–posteriori–Verteilungsdichte* für θ. Der Schätzwert $\hat{\theta}$ ist dann an die Stelle dieses Maximums zu legen. In vielen Fällen ist es sinnvoller, statt der a–posteriori–Dichtefunktion $p(\theta|X)$ die Verbunddichtefunktion $p(\theta, X) = p(\theta|X)\,p(X) = p(X|\theta)\,p(\theta)$ zu maximieren. Da die im Satz von Bayes auftretende Normierung $p(X)$ unabhängig von θ ist und nur als Normierungskonstante dient, wird die Verbunddichtefunktion $p(\theta,X)$ genau dann maximal, wenn die a–posteriori–Dichtefunktion $p(\theta|X)$ maximal wird.

Die Maximum–Likelihood–Methode und die Maximum–a–posteriori–Methode stehen, wie bereits erwähnt, in einer gewissen Beziehung zueinander. Wird nämlich angenommen, daß keine a–priori–Information zur Verfügung steht, so entspricht dies einer Gleichverteilung der a–priori–Dichtefunktion. In diesem Fall wird $p(\theta|X)$ genau dann maximal, wenn auch $p(X|\theta)$ maximal wird.

Beide Verfahren liefern also denselben Schätzwert. Wenn a–priori–Informationen über θ vorhanden sind, verursacht dies eine Verschiebung des Maximums der Likelihood–Funktion.

Als Beispiel für eine Maximum–a–posteriori–Schätzung werde eine normalverteilte Dichtefunktion mit unbekanntem Mittelwert μ und bekannter Varianz σ_i^2 für jeden Stichprobenwert X_i betrachtet. Jede Stichprobe besteht hier also aus zwei Komponenten: μ und σ_i^2; der Parametervektor θ hat ebenfalls zwei entsprechende Einträge. Zusätzlich sei bekannt, daß μ eine Zufallsvariable mit a–priori–Normalverteilung $\pi(\mu) \sim \mathrm{N}(0, \sigma_\mu^2)$, d.h. mit Mittelwert 0 und bekannter Varianz σ_μ^2, sei. Hier wird sinnvollerweise statt $\mathrm{p}(\theta|X)$ die Verbunddichtefunktion $\mathrm{p}(\theta, X) = \mathrm{p}(X \mid \theta)\pi(\theta)$ maximiert. Für die Verbunddichtefunktion ergibt sich für jeden Stichprobenwert

$$\mathrm{p}(\theta, X_i) = \mathrm{p}(X_i \mid \theta)\pi(\theta) = \frac{1}{\sigma_i \sqrt{2\pi}} e^{-\frac{(x_i-\theta)^2}{2\sigma_i^2}} \frac{1}{\sigma_\mu \sqrt{2\pi}} e^{-\frac{\theta^2}{2\sigma_\mu^2}}$$

Analog zur Berechnung bei der Maximum–Likelihood–Methode ist es auch hier einfacher, mit dem natürlichen Logarithmus weiterzurechnen. Der Schätzwert $\hat{\theta}$ wird dann ebenso durch Nullsetzen der Ableitung berechnet, die sich wie folgt ergibt:

$$\mathrm{grad}_\theta(\log(\mathrm{p}(\theta, X)) = \sum_{i=1}^{N} \mathrm{grad}_\theta(\log(\mathrm{p}(X_i \mid \theta)\pi(\theta))) = \sum_{i=1}^{N} \frac{1}{\sigma_i^2}(x_i - \theta) - \frac{\theta}{\sigma_\mu^2}$$

und folglich als Bedingung für die Berechnung des Schätzwertes:

$$\sum_{i=1}^{N} \frac{1}{\sigma_i^2}(x_i - \hat{\mu}) - \frac{\hat{\mu}}{\sigma_\mu^2} = 0$$

wobei $\hat{\mu}$ den Maximum–a–posteriori–Schätzwert bezeichnet. Nach einigen Umformungen erhält man den Schätzwert zu

$$\hat{\mu} = \frac{\sum_{i=1}^{N} \left(\sigma_i^2\right)^{-1} x_i}{\left(\sigma_\mu^2\right)^{-1} + \sum_{i=1}^{N} \left(\sigma_i^2\right)^{-1}}$$

Wie bei der Maximum–Likelihood–Methode erhält man ein analoges Ergebnis für den Schätzwert des Mittelwertvektors μ im mehrdimensionalen Fall (d.h. jede der N Stichproben X_i besteht aus einem K-dimensionalen Vektor und einer Kovarianzmatrix Σ_i, außerdem ist eine a–priori–Normalverteilung $\mathrm{N}(0, \Sigma_\mu)$ gegeben):

$$\hat{\mu} = \left(\Sigma_\mu^{-1} + \sum_{i=1}^{N} \Sigma_i^{-1} \right)^{-1} \sum_{i=1}^{N} \Sigma_i^{-1} x_i$$

was sich für den Fall eines nicht verschwindenden a–priori–Erwartungswertes μ_μ noch ergänzen läßt zu:

$$\hat{\mu} = \left(\Sigma_\mu^{-1} + \sum_{i=1}^{N} \Sigma_i^{-1} \right)^{-1} \left(\sum_{i=1}^{N} \Sigma_i^{-1} x_i + \Sigma_\mu^{-1} \mu_\mu \right)$$

An diesem Beispiel kann die Beziehung zwischen der Maximum–Likelihood Methode und der Maximum–a–posteriori Methode sehr gut verdeutlicht werden: Falls keine a–priori–Informationen vorhanden sind, d.h. σ_μ^2 gegen unendlich strebt, ergibt sich

$$\lim_{\sigma_\mu^2 \to \infty} \left(\frac{\sum_{i=1}^{N} \left(\sigma_i^2\right)^{-1} x_i}{\left(\sigma_\mu^2\right)^{-1} + \sum_{i=1}^{N} \left(\sigma_i^2\right)^{-1}} \right) = \frac{\sum_{i=1}^{N} \left(\sigma_i^2\right)^{-1} x_i}{\sum_{i=1}^{N} \left(\sigma_i^2\right)^{-1}}$$

Dies entspricht genau dem Schätzwert, der mit Hilfe der Maximum–Likelihood–Methode bestimmt wurde. Dieses Ergebnis ergibt sich auch für $N \to \infty$. Der Einfluß der a–priori–Information nimmt also mit zunehmender Stichprobenanzahl ab und der Schätzwert nähert sich dem Maximum–Likelihood–Schätzwert an. Die a–priori–Information hat somit bei kleiner Stichprobenanzahl einen größeren Einfluß.

Für die Güte der Bayesschen Methode gilt die gleiche Aussage wie für die Maximum–Likelihood–Methode: Im allgemeinen Fall ist die Erwartungstreue nicht erfüllt. Für den in der Sensordatenfusion relevanten Fall der Schätzung des Mittelwertes einer Normalverteilung handelt es sich jedoch um eine gute Schätzmethode im Rahmen der Kriterien aus Abschnitt 2.4. Der Nachteil der Bayesschen Methode gegenüber der Maximum–Likelihood–Methode liegt darin, daß neben der Dichtefunktion auch noch a–priori–Information und eine Verlustfunktion bekannt sein müssen. Besonders bei kleiner Stichprobenanzahl ist der Einfluß der a–priori–Information am größten und kann bei schlechter a–priori–Information zu schlechteren Schätzwerten führen. In praktischen Anwendungen verzichtet man deshalb häufig auf die Möglichkeit der Berücksichtigung von a–priori–Information und benutzt die Maximum–Likelihood–Methode.

2.5.2.2 Bayes–Schätzung mit quadratischer Kostenfunktion

Bei dieser Methode wird als Verlustfunktion der quadratische Abstand zwischen Schätzwert und wahrem Wert benutzt: $\mathrm{L}(\theta, \hat{\theta}) = (\theta - \hat{\theta})^2$. Wir erhalten sofort für das bedingte Bayes–Risiko:

$$\mathrm{B}(\hat{\theta} \mid X) = \int_\theta (\theta - \hat{\theta})^2 \, \mathrm{p}(\theta \mid X) \, \mathrm{d}\theta$$

und als Bedingung für den Schätzwert:

$$\mathrm{grad}_{\hat{\theta}} \mathrm{B}(\hat{\theta} \mid X) = \int_\theta \mathrm{grad}_{\hat{\theta}} (\theta - \hat{\theta})^2 \, \mathrm{p}(\theta \mid X) \, \mathrm{d}\theta = 0$$

Damit wird

$$0 = 2\hat{\theta}\int_{\theta} \mathrm{p}(\theta \mid X)\,\mathrm{d}\theta - 2\int_{\theta} \theta\,\mathrm{p}(\theta \mid X)\,\mathrm{d}\theta$$

und schließlich der Schätzwert:

$$\hat{\theta} = \int_{\theta} \theta\,\mathrm{p}(\theta \mid X)\,\mathrm{d}\theta = \mathrm{E}[\theta \mid X]$$

Diese Schätzung beruht auf der Kenntnis von $\mathrm{p}(\theta|X)$ und sollte nicht mit der Schätzung mit der Methode der kleinsten Fehlerquadrate verwechselt werden, bei der diese Kenntnis nicht benötigt wird.

2.5.2.3 Die Bayes–Schätzung mit minimaler Varianz

Die Schätzung mit minimaler Varianz kann als Verallgemeinerung der vorigen Methode aufgefaßt werden. Hier werden nicht die Komponenten des Schätzvektors einzeln betrachtet, sondern das Risiko $\mathrm{B}[\hat{\theta}] = \mathrm{E}_{\theta}[\mathrm{L}(\theta,\hat{\theta})] = \mathrm{E}_{\theta}\left[(\theta - \hat{\theta})^{\mathrm{T}}(\theta - \hat{\theta})\right]$ entspricht der Varianz des gesamten Schätzfehlers, d.h. einer Summe über alle Komponenten. Damit ist die Verlustfunktion $\mathrm{L}(\theta,\hat{\theta}) = (\theta - \hat{\theta})^{\mathrm{T}}(\theta - \hat{\theta})$. Die Komponenten können auch mit einer Wichtungsmatrix W bewertet werden, so daß $\mathrm{L}(\theta,\hat{\theta}) = (\theta - \hat{\theta})^{\mathrm{T}} W(\theta - \hat{\theta})$, was wir aber hier nicht weiter betrachten. Wir erhalten für das bedingte Bayes–Risiko:

$$\mathrm{B}(\hat{\theta} \mid X) = \int_{\theta} (\theta - \hat{\theta})^{\mathrm{T}}(\theta - \hat{\theta})\,\mathrm{p}(\theta \mid X)\,\mathrm{d}\theta$$

und als Bedingung für den Schätzwert:

$$\mathrm{grad}_{\hat{\theta}}\mathrm{B}(\hat{\theta} \mid X) = \int_{\theta} \mathrm{grad}_{\hat{\theta}}[(\theta - \hat{\theta})^{\mathrm{T}}(\theta - \hat{\theta})]\,\mathrm{p}(\theta \mid X)\,\mathrm{d}\theta = 0$$

Mit $\mathrm{grad}_{\hat{\theta}}[(\theta - \hat{\theta})^{\mathrm{T}}(\theta - \hat{\theta})] = -2(\theta - \hat{\theta})^{\mathrm{T}}$ wird daraus nach entsprechender Umformung

$$0 = 2\hat{\theta}\int_{\theta} \mathrm{p}(\theta \mid X)\,\mathrm{d}\theta - 2\int_{\theta} \theta\,\mathrm{p}(\theta \mid X)\,\mathrm{d}\theta$$

und schließlich, wie im vorigen Abschnitt, der Schätzwert für jede Komponente:

$$\hat{\theta} = \mathrm{E}[\theta \mid X]$$

Dies ist der Schätzer, der als Funktion einer Stichprobe X den im Sinne des Kriteriums minimaler Varianz (über alle Komponenten des Schätzvektors) optimalen Schätzwert liefert. Ausführlicher könnte man auch schreiben: $\hat{\theta}(x) = \mathrm{E}[\theta \mid X = x]$.

Beispiel: Es sei eine skalare Zufallsvariable S gegeben, von der bekannt sei, daß sie normalverteilt mit Mittelwert μ_S und Varianz σ_S^2 ist, also $S \sim \mathrm{N}(\mu_S, \sigma_S^2)$. Für diese Zufallsvariable sei der Schätzwert nach der Methode der minimalen Varianz zu bestimmen, wenn eine skalare Beobachtung

$Y = y$ gegeben ist und diese sich aus der additiven Überlagerung von S mit einer zweiten, ebenfalls normalverteilten Zufallsvariablen V (Rauschanteil) ergibt. Es gelte $V \sim \mathrm{N}(0, \sigma_V^2)$, V ist also mittelwertfrei. Damit ergibt sich für die a–priori–Dichte der Summenvariable Y ebenfalls eine Normalverteilung: $\pi(Y) = \mathrm{N}(\mu_S, \sigma_S^2 + \sigma_V^2)$. Der Schätzwert der Variablen S berechnet sich mit dem Kriterium der minimalen Varianz (bzw. der minimalen quadratischen Verlustfunktion):

$$\hat{S} = \mathrm{E}_S[S \mid Y]$$

Um diesen Schätzwert berechnen zu können, ist die Kenntnis der bedingten Dichtefunktion $\mathrm{p}(S|Y)$ erforderlich. Sie berechnet sich mit dem Satz von Bayes zu:

$$\mathrm{p}(S \mid Y) = \mathrm{p}(Y \mid S) \frac{\mathrm{p}(S)}{\pi(Y)}$$

Die beiden Dichten $\mathrm{p}(s)$ und $\pi(y)$ sind bekannt, die Dichte $\mathrm{p}(Y|S)$ ist ebenfalls leicht herzuleiten: Wenn S den Wert s annimmt, variiert die Realisation y um s mit der Varianz von V, also: $\mathrm{p}(Y|S) = \mathrm{N}(s, \sigma_V^2)$. Damit ergibt sich für die a–posteriori–Dichte $\mathrm{p}(S|Y)$:

$$\mathrm{p}(S \mid Y) = \frac{1}{\sqrt{2\pi}\sigma_V} e^{-\frac{1}{2}\left(\frac{y-s}{\sigma_V}\right)^2} \frac{1}{\sigma_S} e^{-\frac{1}{2}\left(\frac{s-\mu_S}{\sigma_S}\right)^2} (\sigma_S + \sigma_V) e^{\frac{1}{2}\frac{(y-\mu_S)^2}{(\sigma_S+\sigma_V)^2}}$$

Wie sich nach einigen Umformungen zeigen läßt [Candy 86], entspricht diese bedingte Dichtefunktion einer Normalverteilung

$$\mathrm{N}\left(\left[\frac{y}{\sigma_V^2} + \frac{\mu_S}{\sigma_S^2}\right]\frac{\sigma_V^2\sigma_S^2}{\sigma_V^2 + \sigma_S^2}, \frac{\sigma_V^2\sigma_S^2}{\sigma_V^2 + \sigma_S^2}\right)$$

und damit ist der gesuchte Schätzwert

$$\hat{S} = \mathrm{E}_S[S \mid Y] = \left[\frac{y}{\sigma_V^2} + \frac{\mu_S}{\sigma_S^2}\right]\frac{\sigma_V^2\sigma_S^2}{\sigma_V^2 + \sigma_S^2}$$

In diesem Fall von normalverteilter a–posteriori–Dichte hätte man im Falle von MAP–Schätzung den gleichen Wert erhalten.

2.5.5 Das Kalman–Bucy–Filter

Das Kalman–Filter [Kalman 69] löst im allgemeinsten Fall die Aufgabe, den sich ständig ändernden Zustand eines linearen, zeitvarianten Systems zu schätzen. Ein solches System liefert in Abhängigkeit von seinem inneren Zustand $\theta(t)$ ein Ausgabesignal $y(t)$. Grundlage der Schätzung des Systemzustandes sind zum einen die beobachteten Ausgangssignale des Systems und zum anderen ein *Modell* des Systems, das erforderlich ist, um vom Meßsignal auf den inneren Systemzustand schließen zu können. Das Kalman–Filter selbst stellt einen *Zustandsschätzer* dar, dessen Eingangsgrößen aus den beobachtbaren Werten am Systemausgang bestehen und an dessen Ausgang die geschätzten Zustandsgrößen zur Verfügung stehen. Es extrahiert oder „filtert" also aus den beobachteten Größen die gewünschten Schätzwerte für den wahren Systemzustand heraus und zwar unter Zuhilfenahme des Systemmodells, welches die Signalentstehung modelliert. Diese Arbeitsweise entspricht der eines normalen Filters in der Nachrichtentechnik, das ebenfalls ein Signal von Störgrößen befreit; daher wird dieser modellbasierte Schätzer als *Filter* bezeichnet. Wir betrachten zunächst das Signalmodell, in einem zweiten Schritt dann den eigentlichen Schätzer. Wir beschränken uns auf den zeitdiskreten Fall.

In Abschnitt 2.5.4 wurde im Beispiel das einfachste denkbare Signalmodell eingeführt: das eines konstanten Signals S, dessen Wert unbekannt und daher unter Zuhilfenahme von Meßwerten Y_k zu schätzen war. Diese Meßwerte waren von einer Störgröße W_k überlagert. Diese Schätzung basierte also auf dem einfachen Modell $Y_k = S + W_k$. Beim Kalman–Filter geht man hingegen vom allgemeinen Modell eines zeitvarianten, linearen Systems aus. Im allgemeinen Fall werden keine Zufallsvariablen geschätzt, sondern Zufallsprozesse. Das System wird im Zeitbereich mit einem Zustandsraummodell beschrieben. Das hat bekanntermaßen gegenüber der Systembeschreibung im Frequenzbereich den Vorteil, daß auch instationäre Vorgänge bei endlicher Beobachtungsdauer einfach beschrieben werden können. Beim hier behandelten Problem wird angenommen, daß sich die Ausgabe des dynamischen Systems aus Informationen über den Systemzustand und einer Störung zusammensetzt. Das Signal und die Störung werden durch diskrete stochastische Prozesse mit bekanntem Mittelwert und bekannter Kovarianzmatrix beschrieben.

Für das System, dessen Zustand zu schätzen ist und das ein vektorielles Signal ausgibt, aufgrund dessen die Schätzung erfolgen soll, wird angenommen, daß es sich um ein allgemeines lineares, zeitvariantes und dynamisches System handelt. Ein solches System wird bekanntlich durch die Zustandsgleichungen

$$\begin{aligned} \theta(k+1) &= A(k)\theta(k) + B(k)u(k) \\ z(k) &= C(k)\theta(k) \end{aligned} \tag{2.5.5–1}$$

beschrieben. Dabei repräsentiert der Vektor $\theta(k)$ den inneren Zustand des Systems zum Zeitpunkt k (die Schreibweise mit Klammern soll andeuten, daß es sich um mit dem Index k zeitabhängige Größen handelt). Jede Komponente von θ beschreibt eine innere Systemgröße. Der Vektor $u(k)$ ist die auf das System wirkende Eingangsgröße („Zwangskraft") und der Vektor $z(k)$ die Ausgangsgröße. Die Systemmatrizen $A(k)$, $B(k)$ und $C(k)$ und ihre Zeitabhängigkeiten sind bekannt. $C(k)$ wird auch als Beobachtungsmatrix bezeichnet, weil sie den Zusammenhang zwischen innerem Systemzustand und äußeren Beobachtungsgrößen herstellt.

Für alle Komponenten $u_i(k)$ des Eingangsvektors $u(k)$ wird ein *weißer Rauschprozeß* angenommen, d.h. $\mathrm{E}[u_i(k)] = 0$ und die (Kreuz–)Kovarianzmatrix Q zwischen dem Zufallsvektor $u(k)$ und dem Zufallsvektor $u(j)$ verschwindet für alle $k \neq j$. Die Kovarianzmatrix Q ist bekannt und zeitabhängig; es gilt $\mathrm{cov}\{u(k), u(j)\} = \mathrm{E}[u(k)u^{\mathrm{T}}(j)] = Q(k)\delta_{kj}$. Dieser weiße Eingangsprozeß spielt hier eine ähnliche Rolle wie der δ–Impuls oder der Einheitsimpuls bei der Bestimmung der Übertragungsfunktion linearer deterministischer Systeme. Die völlige Regellosigkeit des weißen Eingangsprozesses entspricht dem Vorhandensein aller Frequenzen im δ–Impuls. Die stochastischen Parameter des Prozesses $u(k)$ können durch die Gestalt der Matrix $B(k)$ verändert werden. Die Parameter Erwartungswert des Systemzustandes und die Kreuzkovarianzmatrix P des Schätzfehlers sind für den Systemzustand θ zum Startzeitpunkt k_0 bekannt: $\bar{\theta}_0 = \mathrm{E}[\theta(k_0)]$ und $P_0 = P(k_0) = E\left[[\theta(k_0) - \hat{\theta}(k_0)][\theta(k_0) - \hat{\theta}(k_0)]^{\mathrm{T}}\right]$.. Außerdem ist der zukünftige Eingangsvektor nicht mit dem gegenwärtigen Systemzustand korreliert, also: $\mathrm{E}[\theta(j)u^{\mathrm{T}}(k)] = 0$ für $k > j$. Aufgabe des Kalman–Filters ist die Schätzung des Zustandsvektors θ aus dem beobachteten Signal

$$y(k) = F(k)\theta(k) + w(k)$$

mit $w(k)$ als weißem Rauschprozeß. Dessen Kovarianzmatrix wird mit $R(k)$ bezeichnet, ansonsten hat er die gleichen Eigenschaften wie $u(k)$. Die Kovarianzmatrix $\mathrm{E}[\theta(j)w^{\mathrm{T}}(k)] = 0$ für alle k und j, d.h. der Störvektor $w(k)$, ist nicht mit dem Zustandsvektor korreliert. $F(k)$ ist eine weitere Beobachtungsmatrix. Als Beobachtungssignal könnte man auch den Vektor $z(k) = C(k)\theta(k)$ verwenden, die Matrix $C(k)$ soll jedoch für die Beobachtung des geschätzten Zustandsvektor $\hat{\theta}$ verwendet werden (siehe unten, Bild 2). Zusammengefaßt erhält man also das in Bild 1 graphisch dargestellte Signalmodell.

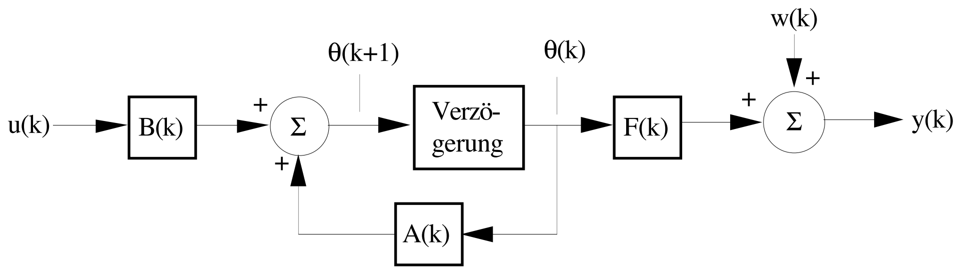

Bild 1: Signalmodell beim Kalman–Filter.
u(k): weißer Eingangsprozeß, B(k): Einflußmatrix, θ(k) Systemzustandsvektor, A(k): Systemmatrix
F(k): Beobachtungsmatrix, w(k): weißer Störprozeß, in Bildmitte: Verzögerungsglied um einen Abtasttakt

Man sollte sich bewußt sein, daß $u(k)$ der Prozeß ist, der das System „in Bewegung" hält, also den inneren Systemzustand $\theta(k)$ verändert, während $w(k)$ ein Störprozeß ist, der die Beobachtung des Vektors $\theta(k)$ verfälscht.

Gesucht wird nun ein Schätzer, der ebenfalls ein lineares, zeitvariantes System (wie das Signalmodell) darstellt[1] und bei dem sich der nächste Zustand aus dem gegenwärtigen Zustand und dem Hinzufügen der beobachteten Information berechnet, also:

$$\hat{\theta}(k+1) = G\hat{\theta}(k) + Ky(k)$$

Dabei stellen G und K Matrizen (oder bei skalarem θ auch Skalare) dar, die beim Entwurf des Schätzers unter Berücksichtigung des Signalmodells (also der Systemmatrizen A und F) bestimmt werden müssen. Beim Entwurf wird zunächst die *Übereinstimmung der Erwartungswerte von Zustandsvektor θ und Schätzvektor* $\hat{\theta}$ gefordert, also $\mathrm{E}[\theta] = \mathrm{E}[\hat{\theta}]$. Daraus folgt mit vorstehender Gleichung

[1] Ein solcher Ansatz liegt nahe, weil er die Form des Signalmodells hat, aber er ist keineswegs zwingend. Es ist nicht von vornherein klar, daß der optimale Schätzer eine solche endliche Form haben kann. Es könnte sich im Verlauf der Rechnung durchaus herausstellen, daß mit einer solchen Gleichung keine Schätzung vorgenommen werden kann.

$$
\begin{aligned}
\mathrm{E}[\hat{\theta}(k+1)] &= \mathrm{E}[G(k)\hat{\theta}(k) + K(k)y(k)] \\
&= \mathrm{E}[G(k)\hat{\theta}(k)] + \mathrm{E}[K(k)\{F(k)\theta(k) + w(k)\}] \\
&= G(k)\,\mathrm{E}[\hat{\theta}(k)] + K(k)F(k)\,\mathrm{E}[\theta(k)] \\
&= [G(k) + K(k)F(k)]\,\mathrm{E}[\theta(k)]
\end{aligned}
$$

Bildet man den Erwartungswert der Gleichung für das Signalmodell, erhält man:

$$
\begin{aligned}
\mathrm{E}[\theta(k+1)] &= \mathrm{E}[A(k)\theta(k) + B(k)u(k)] \\
&= \mathrm{E}[A(k)\theta(k)] + \mathrm{E}[B(k)u(k)] \\
&= A(k)\,\mathrm{E}[\theta(k)] + B(k)\,\mathrm{E}[u(k)] \\
&= A(k)\,\mathrm{E}[\theta(k)]
\end{aligned}
$$

Durch Vergleich der beiden letzten Gleichungen hat man sofort die Bedingung für die Matrix $G(k)$:

$$G(k) = A(k) - K(k)F(k)$$

Damit läßt sich die Matrix G(k) in der Differenzengleichung eliminieren und man erhält:

$$
\begin{aligned}
\hat{\theta}(k+1) &= A(k)\hat{\theta}(k) + K(k)[y(k) - F(k)\hat{\theta}(k)] \\
&= A(k)\hat{\theta}(k) + K(k)[F(k)\theta(k) + w(k) - F(k)\hat{\theta}(k)]
\end{aligned}
\qquad 2.5.5\text{–}2
$$

mit einer Anfangsbedingung, die üblicherweise definiert wird zu: $\hat{\theta}(k_0) = \mathrm{E}[\theta(k_0)]$. Diese beiden Gleichungen beschreiben den Zustandsschätzer vollständig. Man bedenke, daß zur Gewinnung dieser beiden Gleichungen neben dem linearen Systemmodell nur die vernünftige und sehr allgemeine Annahme der Identität zwischen den Erwartungswerten von Zustands– und dessen Schätzvektor notwendig war. Es geht nun noch darum, die Matrix K zu bestimmen, die sogenannte Zielmatrix.

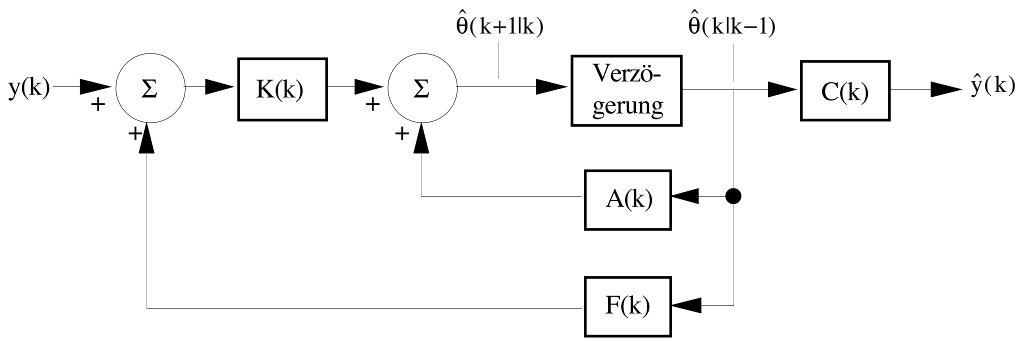

Bild 2: Blockbild des Kalman–Schätzers,
K(k): Zielmatrix, alle anderen Bezeichnungen wie in Bild 1.

Sie vermittelt den Zusammenhang zwischen aktuellem und nächstem Schätzwert und der Beobachtung zum Zeitpunkt k. Wir führen an dieser Stelle zunächst eine leicht modifizierte Notation ein. Um auszudrücken, daß die letzte Gleichung den Schätzwert für den Zeitpunkt $k + 1$ unter der Berücksichtigung von Information bis einschließlich zum Zeitpunkt k bestimmt (also mit dem Wissen um das Systemverhalten vorausberechnet), schreiben wir für diesen Wert: $\hat{\theta}(k+1 \mid k)$. Es handelt sich hier also um ein Vorhersageproblem. Beim eigentlichen Filtern wünscht man einen Wert für den Zeitpunkt k unter Berücksichtigung aller Beobachtungen bis k. In dieser Notation würde man dann schreiben: $\hat{\theta}(k \mid k)$. Man erhält für das Kalman–Filter das in Bild 2 dargestellte Signalflußbild.

Zur Bestimmung der Matrix K, des einen Freiheitsgrades in der Differenzengleichung für den Schätzer, wird neben der oben benutzten Bedingung der Gleichheit der Erwartungswerte für Zustand und Zustandsschätzung eine weitere Bedingung an den Schätzwert gestellt: *Der Erwartungswert der Summe der quadrierten Differenzen zwischen allen Komponenten des Schätzwertes und des wahren Zustandes soll minimal werden.* Mit dem Schätzfehler $\tilde{\theta}(k) = \theta(k) - \hat{\theta}(k)$ ist also

$$
\begin{aligned}
e(k) &= \mathrm{E}[(\theta(k) - \hat{\theta}(k))^{\mathrm{T}}(\theta(k) - \hat{\theta}(k))] \\
&= \mathrm{E}[\tilde{\theta}^{\mathrm{T}}(k)(\tilde{\theta}(k)]
\end{aligned}
$$

zu minimieren. Dies läßt sich ebenso über die Kovarianzmatrix $P(k) = \mathrm{E}[\tilde{\theta}(k)\tilde{\theta}^{T}(k)]$ des Schätzfehlers ausdrücken, und zwar mit: $e(k) = \mathrm{sp}\, P(k)$. Dabei ist sp $P(k)$ die Spur der Kovarianzmatrix $P(k)$, also die Summe aller ihrer Hauptdiagonalelemente, und diese ist zu minimieren.

Zunächst muß der Schätzfehler berechnet werden. Für ihn läßt sich durch Subtraktion der beiden Gleichungen (2.5.5–1) und (2.5.5–2) die folgende Differenzengleichung ableiten:

$$
\begin{aligned}
\tilde{\theta}(k+1 \mid k) &= \theta(k+1 \mid k) - \hat{\theta}(k+1 \mid k) \\
&= A(k)\theta(k \mid k-1) + B(k)u(k) \\
&\quad - A(k)\hat{\theta}(k \mid k-1) + K(k)F(k)\tilde{\theta}(k \mid k-1) + K(k)w(k) \\
&= A(k)\tilde{\theta}(k \mid k-1) - K(k)F(k)\tilde{\theta}(k \mid k-1) + s(k) \\
&= [A(k) - K(k)F(k)]\tilde{\theta}(k \mid k-1) + s(k)
\end{aligned}
$$

mit $s(k) = B(k)u(k) - K(k)w(k)$. Die gleich noch benötigte Kovarianzmatrix dieses Prozesses ist:

$$
\begin{aligned}
\mathrm{E}[s(k)s^{\mathrm{T}}(k)] &= \mathrm{E}[\{B(k)u(k) - K(k)w(k)\}\{B(k)u(k) - K(k)w(k)\}^{\mathrm{T}}] \\
&= \mathrm{E}[\{B(k)u(k) - K(k)w(k)\}\{u^{\mathrm{T}}(k)B^{\mathrm{T}}(k) - w^{\mathrm{T}}(k)K^{\mathrm{T}}(k)\}] \\
&= \mathrm{E}[B(k)u(k)u^{\mathrm{T}}(k)B^{\mathrm{T}}(k) - B(k)u(k)w^{\mathrm{T}}(k)K^{\mathrm{T}}(k) \\
&\quad - K(k)w(k)u^{\mathrm{T}}(k)B^{\mathrm{T}}(k) + K(k)w(k)w^{\mathrm{T}}(k)K^{\mathrm{T}}(k)]
\end{aligned}
$$

und mit der vorausgesetzten Unkorreliertheit zwischen $u(k)$ und $w(k)$ sowie deren Kovarianzmatrizen $Q(k)$ und $R(k)$ wird das zu

$$
\mathrm{E}[s(k)s^{\mathrm{T}}(k)] = B(k)Q(k)B^{\mathrm{T}}(k) + K(k)R(k)K^{\mathrm{T}}(k)
$$

Es handelt sich bei diesem Prozeß also um einen weißen Prozeß, wobei sich außerdem noch zeigen läßt [Unbehauen 89], daß er mit dem Schätzfehler unkorreliert ist. Wenn man diese letztere Eigenschaft berücksichtigt, läßt sich die gesuchte Kovarianzmatrix des Schätzfehlers bestimmen zu:

$$\begin{aligned} P(k+1) &= \mathrm{E}[\tilde{\theta}(k+1 \mid k)\tilde{\theta}(k+1 \mid k)^T] \\ &= \mathrm{E}\Big[[(A-KF)\tilde{\theta}+s][(A-KF)\tilde{\theta}+s]^T\Big] \\ &= \mathrm{E}\Big[(A-KF)\tilde{\theta}\tilde{\theta}^T(A^T-F^TK^T)+ss^T\Big] \end{aligned}$$

wobei aus Gründen der Übersichtlichkeit die Indizes für den Zeitpunkt k bei allen Größen weggelassen und die erwähnte Tatsache berücksichtigt wurde, daß der Schätzfehler und $s(k)$ nicht korreliert sind, daß also die Kovarianzmatrizen $\mathrm{E}[\tilde{\theta}s^{\mathrm{T}}]=0$ und $\mathrm{E}[s\tilde{\theta}^{\mathrm{T}}]=0$. Mit der eben berechneten Matrix $\mathrm{E}[ss^{\mathrm{T}}]$ und $P=\mathrm{E}[\tilde{\theta}\tilde{\theta}^{\mathrm{T}}]$ wird die Kovarianzmatrix des Schätzfehlers zu:

$$\begin{aligned} P(k+1) = {} & (A(k)-K(k)F(k))P(k)(A^{\mathrm{T}}(k)-F^{\mathrm{T}}(k)K^{\mathrm{T}}(k))+ \\ & B(k)Q(k)B^{\mathrm{T}}(k)+K(k)R(k)K^{\mathrm{T}}(k) \end{aligned}$$

Jetzt kann die optimale Zielmatrix K bestimmt werden. Dazu wird die Spur von $P(k+1)$ nach allen Elementen von K abgeleitet und das Ergebnis zu Null gesetzt. Mit den Ableitungsregeln für die Spur $\partial\mathrm{sp}(KM)/\partial K = M^{\mathrm{T}}$, $\partial\mathrm{sp}(MK)/\partial K = M$ und $\partial\mathrm{sp}(KMK^{\mathrm{T}})/\partial K = 2KM$ wird sofort die Bedingung für die optimale Matrix K:

$$\hat{K}(k) = A(k)P(k)F^{\mathrm{T}}(k)\big[F(k)P(k)F^{\mathrm{T}}(k)+R(k)\big]^{-1}$$

Diese Matrix minimiert den mittleren quadratischen Fehler. Setzt man diese in die vorige Gleichung ein und faßt geeignet zusammen, so erhält man für die Berechnung von $P(k+1)$ noch eine etwas kürzere, aber ansonsten zur oberen identische Form (für die Umformung siehe [Mohanty 86]):

$$P(k+1) = A(k)P(k)A^{\mathrm{T}}(k) - \hat{K}(k)F^{\mathrm{T}}(k)P(k)A^{\mathrm{T}}(k) + B(k)Q(k)B^{\mathrm{T}}(k)$$

Damit läßt sich der diskrete *Kalman–Schätzer in Prädiktor–Form* angeben:

1. Berechne den Startwert für den Schätzvektor $\hat{\theta}(k_0)=\mathrm{E}[\theta(k_0)]$ und den Startwert für die Kovarianzmatrix $P(k_0)=\mathrm{E}[\hat{\theta}(k_0)\hat{\theta}^{\mathrm{T}}(k_0)]$. Setze $k = k_0$.

2. Berechne die Zielmatrix (*Kalman–Gain–Matrix*) für den Zeitpunkt k:
$\hat{K}(k) = A(k)P(k)F^{\mathrm{T}}(k)\big[F(k)P(k)F^{\mathrm{T}}(k)+R(k)\big]^{-1}$

3. Berechne den Voraussage–Schätzwert für den Zeitpunkt $k+1$ unter Benutzung des bekannten Meßvektors $y(k)$ und des alten Voraussage-Schätzwertes:
$\hat{\theta}(k+1 \mid k) = A(k)\hat{\theta}(k \mid k-1) + \hat{K}(k)[y(k) - F(k)\hat{\theta}(k \mid k-1)]$

4. Berechne die korrigierte Kovarianzmatrix des Schätzfehlers $P(k+1)$:
$P(k+1) = A(k)P(k)A^{\mathrm{T}}(k) - \hat{K}(k)F^{\mathrm{T}}(k)P(k)A^{\mathrm{T}}(k) + B(k)Q(k)B^{\mathrm{T}}(k)$

5. Setze $k := k+1$. Gehe zu 2.

Der Schätzwert für den nächsten Zeitpunkt wird aus der Information zum Zeitpunkt k berechnet, nämlich dem vorigen Vorhersagewert und dem aktuellen Meßwert. Wird ein eigentliches Filter gewünscht, d.h. die Berechnung von $\hat{\theta}(k \mid k)$, muß der Schätzwert für den Zeitpunkt $k + 1$ auf den Zeitpunkt k zurückgerechnet werden. Dies geschieht mit Hilfe der Inversen der Systemübergangsmatrix, die sich mit Hilfe der Matrix A bestimmen läßt (siehe dazu [Mohanty 86]).

Es sei abschließend bemerkt, daß für statische Probleme mit normalverteiltem Fehler w das Kalman–Bucy–Filter dem Verfahren der Bayesschen Schätzung mit dem Optimalitätskriterium der minimalen Varianz und der Methode des kleinsten quadratischen Erwartungswertes für den Schätzfehler entspricht. Das statische Kalman–Filter hat durch seine rekursive Struktur gegenüber der Bayesschen Methode den Vorteil, daß auch schon mit einem Teil der Meßdaten die Berechnung begonnen werden kann. Wenn keine Voraussetzungen über die Dichten der Fehler (u und w) gemacht werden, ist das Kalman–Filter das *lineare* Schätzverfahren, welches den Schätzwert mit geringster Varianz bestimmt. Im Falle von Normalverteilungen ist das Kalman–Filter ebenfalls der Schätzer, der den Schätzwert mit geringster Varianz liefert, wobei hier auch ein nichtlinearer Schätzer kein besseres Ergebnis im Sinne minimaler Varianz liefern kann.

Wir betrachten nun noch drei Beispiele. Im ersten soll mit einfachen Überlegungen das Kalman–Filter für ein skalares Problem angewendet werden. Die vorstehenden Beziehungen werden nicht benutzt, sondern es wird anhand der Problemstellung gezeigt, daß man die Gleichungen für das Filter auch mit Hilfe einfacher heuristischer Überlegungen erhält. Die Anwendung der Beziehungen auf ein System mit zwei Zustandsgrößen zeigen dann die Beispiele 2 und 3.

Beispiel 1. Mit Hilfe eines Entfernungsmeßgerätes soll die Entfernung zu einem Gegenstand bestimmt werden [Loffeld 90]. Da jede einzelne Messung mit einem Meßfehler behaftet ist, soll durch die Kombination der Meßwerte die Genauigkeit der Messung gegenüber jeder für sich genommenen einzelnen Messung verbessert werden. Die Messung sei gestört und der zur Verfügung stehende Meßwert werde durch eine *normalverteilte* skalare Zufallsvariable $y(t)$ mit Varianz $\sigma_y^2(k)$ modelliert. Die tatsächliche Entfernung werde in der Zufallsvariablen $\theta(k)$ mit der Varianz $\sigma_\theta^2(k)$ repräsentiert. Zum Zeitpunkt 1 liefere das Meßgerät den Entfernungsmeßwert $y(k_1)$ und zusätzlich noch eine Angabe über die Vetrauenswürdigkeit des gelieferten Meßwertes in Form einer Varianz $\sigma_y^2(k_1)$. Damit läßt sich der Schätzwert für die Entfernung θ zum Zeitpunkt 1, also $\hat{\theta}(k_1)$, sinnvoll nur zu

$$\hat{\theta}(k_1) = \mathrm{E}[\theta(k_1) \mid y(k_1)] = y(k_1)$$

mit Varianz

$$\sigma_{\hat{\theta}}^2(k_1) = \sigma_y^2(k_1)$$

bestimmen. Zum Zeitpunkt 2 wird die zweite Messung $y(k_2)$ mit Varianz $\sigma_y^2(k_2)$ vorgenommen. Sie soll genauer als die erste Messung sein, d.h. $\sigma_y^2(k_2) < \sigma_y^2(k_1)$. Gesucht wird eine optimale Kombination dieser beiden Messungen mit dem Ziel, daß die Kombination genauer sein soll als jede einzelne Messung. Da die zweite Messung die genauere ist, sollte diese bei der Kombination entsprechend stärker gewichtet werden. Das Gesamtergebnis soll sich aus der linearen gewichteten Überlagerung der beiden Einzelergebnisse ergeben. Also: $\hat{\theta}(k_2) = ay(k_1) + by(k_2)$. Die Gewichte sollen umgekehrt verteilt werden wie die Varianzen, also $a / b = \sigma_y^2(k_2)/\sigma_y^2(k_1)$. Außerdem muß natürlich $a + b = 1$ gelten. Damit ergibt sich für den Schätzwert zum Zeitpunkt k_2:

$$\hat{\theta}(k_2) = \frac{\sigma_y^2(k_2)}{\sigma_y^2(k_1) + \sigma_y^2(k_2)} y(k_1) + \frac{\sigma_y^2(k_1)}{\sigma_y^2(k_1) + \sigma_y^2(k_2)} y(k_2) \tag{2.5.5–3}$$

Für die Varianzen gilt dann (bei normalverteilten Dichten, siehe [Trivedi 82]):

$$\frac{1}{\sigma_{\hat{\theta}}^2(k_2)} = \frac{1}{\sigma_y^2(k_1)} + \frac{1}{\sigma_y^2(k_2)} \tag{2.5.5–4}$$

Die Varianz der Linearkombination beider Messungen ist kleiner als die Varianz jeder einzelnen Messung. Diese beiden Gleichungen geben den gefilterten Schätzwert und die neue Varianz.

Für einen rekursiven Berechnungsalgorithmus müssen die obigen Gleichungen modifiziert werden. Unter Verwendung der Tatsache, daß als Voraussage für die Entfernung $y(k_1) = \hat{\theta}(k_1)$ gilt, erhält man

$$\begin{aligned} \hat{\theta}(k_2) &= y(k_1) + K(k_2)[y(k_2) - y(k_1)] \\ &= \hat{\theta}(k_1) + K(k_2)[y(k_2) - \hat{\theta}(k_1)] \end{aligned}$$

mit K als (skalarer) Zielmatrix. Deren Funktion wird hier offensichtlich: Sie vermittelt den Zusammenhang zwischen neuem Meßwert und altem Schätzwert. Ihr Wert ergibt sich durch Vergleich mit Gl. (2.5.5–3) zu:

$$K(k_2) = \frac{\sigma_y^2(k_1)}{\sigma_y^2(k_1) + \sigma_y^2(k_2)} = \frac{\sigma_{\hat{\theta}}^2(k_1)}{\sigma_{\hat{\theta}}^2(k_1) + \sigma_y^2(k_2)}$$

und entsprechend rekursiv für jeden weiteren Schätzwert. Für die neue Varianz ergibt sich in dieser Notation:

$$\sigma_{\hat{\theta}}^2(k_2) = \sigma_{\hat{\theta}}^2(k_1) - K(k_2)\sigma_{\hat{\theta}}^2(k_1)$$

und entsprechend für jeden weiteren Meßwert. Die letzten drei Beziehungen sind die Gleichungen des Kalman–Filters für zeitinvariante Probleme (vgl. die Beziehungen am Ende von 2.5.2.1). Sie sind lediglich die rekursive Formulierung der Gl. 2.5.5–3 und 2.5.5–4. Da das Meßgerät zu jeder Messung eine Varianz ausgegeben hat, ist es hier möglich, die Filtergleichungen anzugeben, d.h. die Berechnung von $\theta(k|k)$ vorzunehmen. Diese Gleichungen werden beispielsweise direkt in dieser Form in [Smith 86] zur Fusion geometrischer Daten und in [Crowley 89] für die Fusion von Ultraschalldaten zur Steuerung eines mobilen Roboters eingesetzt.

Wir betrachten die Gleichungen (2.5.5–3) und (2.5.5–4) nun noch für den Fall eines mittels mit normalverteilten Beobachtungen X_i (mit Kovarianzmatrizen Σ_i) zu schätzenden, festen Parametervektors θ. Zunächst wird Gl. 2.5.5–3 auf den Fall von N Beobachtungen erweitert. Dazu wird sie zunächst umgeformt zu:

$$\hat{\theta}(k_2) = \frac{\sigma_y^{-2}(k_1) y(k_1)}{\sigma_y^{-2}(k_1) + \sigma_y^{-2}(k_2)} + \frac{\sigma_y^{-2}(k_2) y(k_2)}{\sigma_y^{-2}(k_1) + \sigma_y^{-2}(k_2)}$$

Die Erweiterung auf N Beobachtungen ist damit offensichtlich und wird zu:

$$\hat{\theta}(k_2) = \left[\sum_{i=1}^{N} \sigma_y^{-2}(k_i) \right]^{-1} \sum_{i=1}^{N} \sigma_y^{-2}(k_i) y(k_i)$$

Dies ist für den vektoriellen Fall entsprechend:

$$\hat{\theta} = \hat{\theta}(k_N) = \left[\sum_{i=1}^{N} \Sigma_i^{-1} \right]^{-1} \sum_{i=1}^{N} \Sigma_i^{-1} X_i$$

für den Zustandsvektor. Für die zu diesem Schätzvektor gehörige Kovarianzmatrix hat man Gl. (2.5.5–4) entsprechend:

$$\hat{\Sigma} = \hat{\Sigma}(k_N) = \left[\sum_{i=1}^{N} \Sigma_i^{-1} \right]^{-1}$$

Diese Schätzung wird beispielsweise in [Durrant–Whyte 87] als „Kalman–Filter“ verwendet. Sie ist mit der Maximum–a–posteriori–Schätzung (Abschnitt 2.5.2.1) vollkommen identisch.

Um auch das Vorgehen bei zeitvarianten Problemen illustrieren zu können, nehmen wir nun an, daß sich der wahre Entfernungswert zwischen der Messung zum Zeitpunkt k_2 und einer dritten Messung zum Zeitpunkt k_3 kontinuierlich ändert. Es muß zuerst bekannt sein, nach welchen Gesetzmäßigkeiten sich die Entfernung ändert. Es sei bekannt, daß die Entfernung kontinuierlich mit der Zeit zunehme. Als Systemgleichung ergibt sich also:

$$\theta(k + 1) = \theta(k) + v$$

wobei v eine konstante, bekannte Entfernungszunahme (Geschwindigkeit) darstellt. Für die Messung $y(k)$ gelte:

$$y(k) = \theta(k) + w$$

Die normalverteilte Zufallsvariable w mit Mittelwert Null und Varianz σ_w^2 gibt die Abweichung von der tatsächlichen Entfernung an. Die Ungenauigkeit in bezug auf die Entfernung wird mit fortschreitendem Zeitabstand von der letzten Messung zum Zeitpunkt k_2 immer größer. Dies führt dazu, daß durch die angesammelten Ungenauigkeiten seit der letzten Messung nur noch vage Vorkenntnisse über die aktuelle Entfernung vorliegen. Die bestmögliche *Voraussage* der Entfernung ergibt sich aus

$$\hat{\theta}(k_3 \mid k_2) = \hat{\theta}(k_2) + v(k_3 - k_2)$$

Die Varianz dieses Schätzwertes ist infolge der vorhandenen Unsicherheiten *größer* als die Varianz des Schätzwertes zum Zeitpunkt t_2:

$$\sigma^2_{\hat{\theta}}(k_3 \mid k_2) = \sigma^2_{\hat{\theta}}(k_2) + \sigma^2_w(k_3 - k_2)$$

Die dritte *Messung* liefert $y(k_3)$ mit Varianz $\sigma^2_y(k_3)$. Analog zur Verarbeitung des zweiten Meßwertes folgt:

$$\hat{\theta}(k_3) = \hat{\theta}(k_3 \mid k_2) + K(k_3)[y(k_3) - \hat{\theta}(k_3 \mid k_2)]$$

mit

$$K(k_3) = \frac{\sigma^2_{\hat{\theta}}(k_3 \mid k_2)}{\sigma^2_{\hat{\theta}}(k_3 \mid k_2) + \sigma^2_y(k_3)}$$

und der Varianz

$$\sigma^2_{\hat{\theta}}(k_3) = \sigma^2_{\hat{\theta}}(k_3 \mid k_2) - K(k_3)\sigma^2_{\hat{\theta}}(k_3 \mid k_2)$$

Dies sind die skalaren Kalman–Filtergleichungen. Mit diesen Ergebnissen läßt sich der Vorhersagewert für den Zeitpunkt k_4 berechnen und dann schließlich der gefilterte Wert für k_4, nachdem $y(k_4)$ bekannt ist.

Beispiel 2. Um die Wirksamkeit des Kalman–Bucy–Filter für dynamische Probleme zu veranschaulichen, wurde ein zweidimensionales Filter implementiert [Wegner 91]. Vereinfacht wurde davon ausgegangen, daß keine Störung bei der Berechnung des neuen Systemzustandes $\theta(k+1)$ auftritt, d.h. die Kovarianzmatrix des Kanalrauschens $Q(k)$ ist die Nullmatrix. Als erstes Ausgangssignal wurde eine Gerade mit Steigung 1.1 betrachtet. Die konstante Systemmatrix A ist also:

$$A = \begin{bmatrix} 1.1 & 0 \\ 0 & 1.1 \end{bmatrix}$$

Es wurden zwanzig Meßwerte erzeugt, indem auf den wahren Meßwert (der einfach über die Matrix A zu bestimmen ist) Zufallsvektoren $w(k)$, $k = 1\ldots20$, addiert wurden. Die Mittelwerte der Vektoren lagen entlang der Geraden und die Varianzen in den Kovarianzmatrizen $R(k)$ nahmen kontinuierlich zu. Es wurde gesetzt:

$$R(k) = k \begin{bmatrix} 0.02 & 0 \\ 0 & 0.02 \end{bmatrix}$$

und der erzeugende Zufallszahlengenerator entsprechend eingestellt. Das Ausgangssignal und diese miteinander verbundenen gestörten Meßwerte zeigt Bild 3. Das Ergebnis der Filterung im Vergleich zum Ausgangssignal stellt Bild 4 dar. Trotz der immer größer werdenden Störung liegen die erhaltenen Schätzwerte sehr nahe am Ausgangssignal.

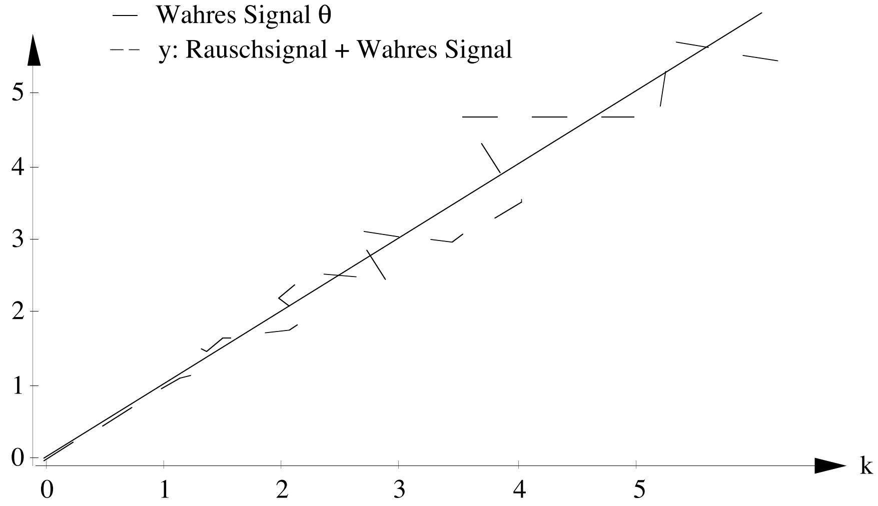

Bild 3: Wahres Signal θ und gestörtes Signal y

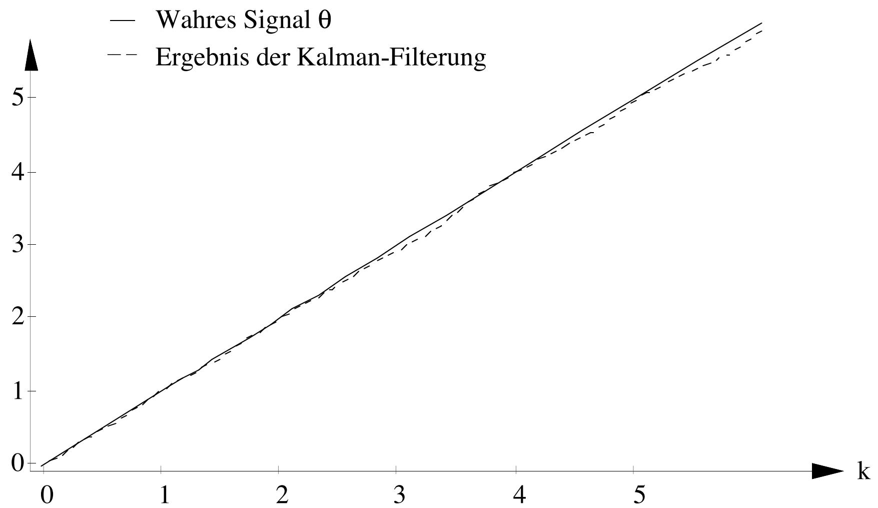

Bild 4: Wahres Signal θ und Ergebnis der Kalman–Filterung θ'

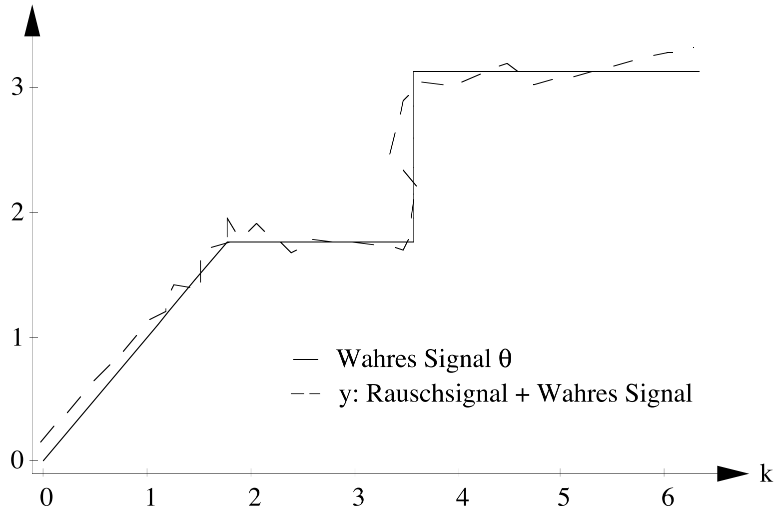

Bild 5: Wahres Signal θ und gestörtes Signal y

Beispiel 3. Nunmehr wird das Ausgangssignal nach Bild 5 gewählt. Dieses Ausgangssignal wird durch eine zeitabhängige Systemmatrix erzeugt. Für die Matrix $A(k)$ ergibt sich analog zu A im ersten Beispiel

$$A(k) = \begin{bmatrix} 1.1 & 0 \\ 0 & 1.1 \end{bmatrix} \text{ für } k = 1\ldots7$$

und danach zu

$$A(k) = \begin{bmatrix} 1.1 & 0 \\ 0 & 1 \end{bmatrix} \text{ für } k = 8\ldots14 \text{ und } k = 22\ldots28$$

bzw.

$$A(k) = \begin{bmatrix} 1 & 0 \\ 0 & 1.1 \end{bmatrix} \text{ für } k = 15\ldots21$$

Es wurden 28 Meßwerte zufällig durch Normalverteilungen mit Mittelwerten entlang des wahren Signalverlaufs und konstanten Kovarianzmatrizen mit der Varianz 0.1 auf der Hauptdiagonalen ermittelt. Bild 5 zeigt das Signal und die erhaltenen Meßwerte. Das Ergebnis des Kalman–Bucy–Filters im Vergleich zum Ausgangssignal stellt Bild 6 dar. Auch in diesem Beispiel führt der Einsatz des Kalman–Bucy–Filters zu einer wesentlichen Verbesserung. Ein Kalman–Bucy–Filter kann als gute Schätzmethode charakterisiert werden, denn es ist erwartungstreu, konsistent, effizient und der Schätzwert wird durch alle vorhandenen Stichprobenwerte bestimmt. Für ein statisches System mit $A(k) = I$ (Identität) und normalverteiltem Fehler entspricht das Kalman–Bucy–Filter Verfahren einer Bayesschen Schätzmethode mit dem Optimalitätskriterium der minimalen Varianz und der Methode des kleinsten quadratischen Mittelwertes.

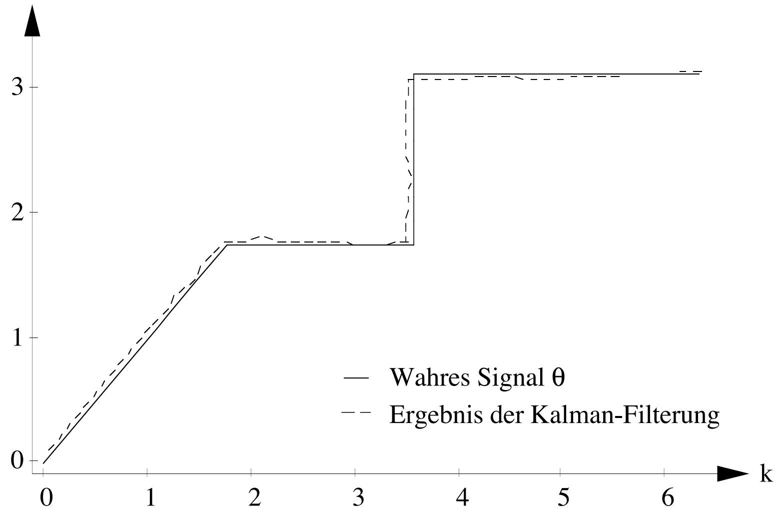

Bild 6: Wahres Signal θ und Ergebnis der Kalman–Filterung θ'

Das statische Kalman–Filter hat gegenüber der Bayesschen Methode (bei der erst alle Meßdaten vorliegen müssen bevor der Schätzwert berechnet werden kann) den Vorteil, daß auch schon mit einem Teil der Meßdaten die Berechnung begonnen werden kann.

Wird das Kalman–Filter bei der Sensordatenfusion eingesetzt, dann wird häufig nur der statische Fall zugrunde gelegt. Die Beobachtungen x_i werden als gestörte Eingangsgröße (oben mit $y(k)$ bezeichnet) nacheinander dem Filteralgorithmus übergeben und von diesem jeweils ein neuer Schätzwert für den (konstanten) Zustand θ (basierend auf allen bislang übergebenen Meßwerten) berechnet. Im hier vorliegenden Fall einer statischen Umweltbeobachtung ist (mit der oben benutzten Notation) die Systemmatrix A die Einheitsmatrix und die Matrix B verschwindet, was nichts anderes bedeutet, als daß eine eventuell vorhandene Zwangskraft keinen Einfluß auf die Umwelt nimmt. Der zu schätzende Vektor θ bleibt mithin konstant, es ist $\theta(k+1) = \theta(k)$. Im Falle dynamischer Umwelten würden sich die Bewegungsgesetze in den Matrizen A und B spiegeln.

Die Beobachtungen resultieren im statischen Fall lediglich aus dem einfachen Meßmodell $x_i = F_i\theta + w_i$, wo F_i die für den jeweiligen Sensor gültige Sensortransformation darstellt und w_i dem jeweiligen Sensorfehler entspricht. Damit sind in jedem Schritt für die Parameter von w_i die Parameter des jeweiligen Fehlers einzusetzen (Kovarianzmatrix R_i für den Sensor i). Mit jedem übergebenen Meßwert nähert sich der Schätzwert dann dem wahren Parameter θ an. Wenn alle Beobachtungen eingegeben wurden, ist der resultierende Schätzwert der beste im Sinne des Kriteriums der minimalen Varianz. Wir betrachten dazu ein abschließendes Beispiel.

2.6 Grundprinzipien der Dempster–Shafer–Theorie

Beim praktischen Einsatz der auf dem Satz von Bayes beruhenden Verfahren zur Sensordatenfusion ergeben sich zuweilen Probleme in zwei Bereichen:

a) **Sensorgerechte Abstraktion.** Insbesondere bei Sensorhierarchien oder bei der Fusion von Sensoren, die über unterschiedliche Arbeitsweisen verfügen, tritt das Problem der sensorgerechten Datendarstellung auf adäquatem Abstraktionsniveau auf. Dies kann am besten anhand eines Beispiels verdeutlicht werden: Angenommen, ein Sensorverbund soll sechs verschiedene Objekttypen unterscheiden. Diese sechs Objekte haben drei verschiedene Formen F_1, F_2 und F_3 und zwei verschiedene Farben C_1 und C_2. Es gibt zwei Sensoren, einen für die Formerkennung und einen für die Farberkennung. Als Meßergebnis nach einer Beobachtung soll für jeden Objekttyp die Wahrscheinlichkeit dafür angegeben werden, daß er vorliegt. Bei den Verfahren nach Bayes arbeiten beide Sensoren auf demselben Ereignisraum, der der Produktmenge der den einzelnen Sensoren zugeordneten Ereignismengen entspricht, in diesem Fall also der Menge $\{F_1C_1, F_1C_2, F_2C_1, F_2C_2, F_3C_1, F_3C_2,\}$. Sowohl der Farb–, als auch der Formsensor müssen allen sechs Ereignissen eine Wahrscheinlichkeit zuweisen, so daß sich eine Gesamtwahrscheinlichkeit von Eins ergibt (obwohl beide Sensoren nur über jeweils ein Merkmal der Ereignisse eine Aussage treffen können). Wenn diese Unsicherheit über alle Elemente gleich verteilt wird, ergibt sich zwar dennoch das richtige Ergebnis, aber dieses Vorgehen ist unübersichtlich. Bei der Betrachtung des Entscheidungsraumes des Sensors ist es nämlich nicht mehr möglich, zu ersehen, welche Entscheidungen der Sensor treffen kann (mithin für welchen Bereich der Sensor eine *source of evidence* ist). Wünschenswert ist es also, für jeden Sensor einen Entscheidungsraum definieren zu können, der genau die Entscheidungen repräsentiert, die der Sensor treffen kann, und irrelevante Kombinationen nicht mitzuschleppen.

b) **Unsichere Information.** Bei den Bayesschen Methoden muß jedem Ereignis eine Wahrscheinlichkeit zugeordnet werden. Wenn jedoch über eine Teilmenge B eines Ereignisraumes A keine Information zur Verfügung steht, kann diese Ungewißheit nur so ausgedrückt werden, daß allen Elementen von B die gleiche Wahrscheinlichkeit zugeordnet wird. Zweckmäßiger wäre es jedoch, ausdrücken zu können, daß über die Auftrittswahrscheinlichkeiten der Elemente von B keine Aussage gemacht werden kann. Dies geschieht bei der Theorie nach Dempster–Shafer durch die Angabe eines unteren und oberen Zahlenwerts für die Wahrscheinlichkeit des Auftritts eines Ereignisses.

Die Theorie nach Dempster und Shafer [Shafer 75; Gordon 84; Bogler 87] ermöglicht es, diese Unsicherheiten auf dem dem jeweiligen Sensor zugeschnittenen Ereignisraum zu behandeln. Zunächst wird ein *frame of discernment* Θ (Unterscheidungsrahmen) eingeführt, der eine Menge von L disjunkten Elementarereignissen umfaßt. Jedem der 2^L Elemente der Potenzmenge von Θ kann eine Wahrscheinlichkeit zugeordnet werden (zu diesen Elementen gehören natürlich auch die Elementarereignisse). Die Potenzmenge wird in diesem Kontext mit 2^Θ bezeichnet. Wir betrachten ein Beispiel: Angenommen, ein Sensor kann $L = 3$ verschiedene Formen unterscheiden, es sei $\Theta = \{F1, F2, F3\}$. Dann ist die Potenzmenge $2^\Theta = \{\emptyset,\{F1\},\{F2\},\{F3\},\{F1, F2\},\{F2, F3\},\{F1, F3\}, \Theta\}$. Jedem Element dieser Menge kann eine Wahrscheinlichkeit zugewiesen werden (*basic probability assignment*). Bei allen mehrelementigen Teilmengen von Θ repräsentiert die zugewiesene „Wahrscheinlichkeitsmasse" die Wahrscheinlichkeit dafür, daß eines der in ihr enthaltenen Elementarreignisse eingetreten ist. Diese Masse kann sich beliebig auf die Elemente der Teilmenge verteilen. Die Summe der den einzelnen Teilmengen von Θ zugewiesenen Massen muß Eins betragen. Beim Bayes–Ansatz müssen demgegenüber den drei Elementen F1, F2, F3 „Punktmassen" zugewiesen werden, deren Summe Eins beträgt. Die der üblichen Wahrscheinlichkeitsfunktion P vergleichbare Funktion m bildet jede Teilmenge A von 2^{Θ}in das Intervall [0,1] der reellen Zahlen ab, wobei m(Ø) = 0 und

$$\sum_{A \subseteq \Theta} \mathrm{m}(A) = 1$$

Diese Abbildung wird auch m: $2^{\Theta} \rightarrow$ [0,1] geschrieben. Die Wahrscheinlichkeit für das Eintreten der durch eine Teilmenge A von Θ spezifizierten Ereignisse wird durch Summation der Wahrscheinlichkeiten aller Teilmengen von A berechnet und in der sogenannten *support*–Funktion spt (oder auch *belief*–Funktion bel) ausgedrückt, wobei

$$\mathrm{spt}(A) = \sum_{T \subseteq A} \mathrm{m}(T)$$

und spt(Ø) = 0 sowie spt(Θ) = 1. Im obigen Beispiel wäre spt({F1, F2}) = m({F1}) + m({F2}) + m({F1, F2}). Neben der Funktion spt, die angibt, mit welcher Wahrscheinlichkeit ein Ereignis eintritt, wird nun ein zweites Maß „pls" eingeführt, welches angibt, mit welcher Plausibilität (Wahrscheinlichkeit) der Eintritt eines Ereignisses *nicht widerlegt* werden kann. Es ist

$$\mathrm{pls}(A) = \sum_{A \cap T \neq È} \mathrm{m}(T) = 1 - \mathrm{spt}(\overline{A})$$

Im oberen Beispiel ist pls({F1, F2}) = 1 – m(F3) = m({F1}) + m({F2}) + m({F1, F2}) + m({F1, F3}) + m({F2, F3}) + m({F1, F2, F3}). Man sieht sofort, daß pls(A) immer größer oder gleich spt(A) ist.

Die Funktionen spt und pls sind die Grundlage der Dempster–Shafer–Theorie. Sie repräsentieren eine untere und eine obere Grenze für das Maß an Sicherheit, daß eine gewisse Hypothese A gilt (daß ein gewisses Ereignis A eintritt). Je größer die Differenz zwischen oberer und unterer Grenze ist, desto größer ist die Unsicherheit über die tatsächliche Wahrscheinlichkeit für den Eintritt von A. Diese Unsicherheit spiegelt sich in den Teilmengen von Θ, die Elemente mit A gemeinsam haben, aber keine Teilmengen von A sind. Je mehr solcher Mengen mit gemeinsamen Elementen es gibt, desto größer wird die Unsicherheit und umso größer wird das Unsicherheitsintervall pls(A) – spt(A). Im Unterschied zum Bayes–Ansatz ist hier spt(A) + spt($\overline{A}$) ≤ 1 (bei Bayes gilt immer spt(A) + spt($\overline{A}$) = 1).

Die für die Sensordatenfusion wichtige Beziehung ist die *Dempstersche Kombinationsregel*. Sie entspricht dem Satz von Bayes und erlaubt die Berechnung einer neuen Wahrscheinlichkeitsmasse. Wenn m_1 und m_2 zwei Funktionen über demselben Raum Θ sind, die durch zwei Sensoren erzeugt werden, dann gilt für die kombinierte Summenfunktion:

$$\mathrm{m}(A) = \mathrm{m}_1 \oplus \mathrm{m}_2 = \frac{1}{K} \sum_{X \cap Y = A} \mathrm{m}_1(X)\,\mathrm{m}_2(Y)$$

mit

$$K = 1 - \sum_{X \cap Y = \varnothing} \mathrm{m}_1(X)\,\mathrm{m}_2(Y)$$

Damit lassen sich für jede Teilmenge aus Θ neue Wahrscheinlichkeiten bestimmen. Die Konstante K dient dabei lediglich der Normierung. Im Beispiel gilt für {F1, F2}, wenn m_1({F1, F2}) die vom ersten Sensor der Teilmenge {F1, F2} zugeordneten Masse und m_2({F1, F2}) entsprechend die vom zweiten Sensor generierte Masse ist:

$$\mathrm{m}(\{F1, F2\}) = \mathrm{m}_1 \oplus \mathrm{m}_2 = \frac{1}{K}\left[\mathrm{m}_1(\Theta)\,\mathrm{m}_2(\{F1, F2\}) + \mathrm{m}_2(\Theta)\,\mathrm{m}_1(\{F1, F2\})\right]$$

mit

$$\begin{aligned} K = 1 \; & - \mathrm{m}_1(F1)\,\mathrm{m}_2(F2) - \mathrm{m}_2(F1)\,\mathrm{m}_1(F2) \\ & - \mathrm{m}_1(F1)\,\mathrm{m}_2(F3) - \mathrm{m}_2(F1)\,\mathrm{m}_1(F3) \\ & - \mathrm{m}_1(F2)\,\mathrm{m}_2(F3) - \mathrm{m}_2(F2)\,\mathrm{m}_1(F3) \end{aligned}$$

Die Berechnung des Zählers ist für alle Teilmengen von Θ zu wiederholen, um die Funktion m = $\mathrm{m}_1 \oplus \mathrm{m}_2$ vollständig zu erhalten. Damit können dann wieder die beiden Kenngrößen spt und pls berechnet werden.

Die Dempstersche Kombinationsregel erfreut sich zunehmender Beliebtheit im Bereich der Sensordatenfusion, weil sie numerisch leicht auszuwerten ist. Daß sie jedoch nützlichere Ergebnisse als reine Bayes–Methoden bringt, wird gelegentlich bezweifelt (siehe insbesondere [Cheeseman 85]).

Literatur

Abdulghafour 90	Abdulghafour, M., Goddard, J., Abidi, M. *Non–deterministic Approaches in Data Fusion – A Review* SPIE Vol. 1383, Sensor Fusion III: 3–D Perception and Recognition
Abidi 89	Abidi, M. *Sensor Fusion: A New Approach and its Applications* Proc. SPIE, Vol. 1198, Sensor Fusion II: Human and Machine Strategies, 1989
Acharya 88	Acharya, P., Henderson, T. *Parameter Estimation and Error Analysis of Range Data* Proc. 1988 IEEE Conf. on Robotics and Autom.
Aggarwal 90	Aggarwal, J., Nadhakumar, N. *Multisensor Fusion for Automatic Scene Interpretation* in: Jain, R., Jain., A. (eds.): Analysis and Interpretation of Range Images, New York: Springer-Verlag, 1990
Allen 85	Allen, P., Bajcsy, R. *Object recognition using vision and touch* Proceedings of the Int. Joint Conf. on Art. Intell., 1985
Allen 87	Allen, P. *Object recognition using vision and touch* Boston: Kluwer Academic, 1987
Allen 90	Allen, P., Michelman, P., Roberts, K. *A System for Programming and Controlling a Multisensor Robotic Hand* IEEE Trans. on Systems, Man, and Cybernetics, Vol. 20, No.6, Nov. 1990

Allen 90b	Allen, P., Michelman, P. *Acquisition and Interpretation of 3–D Sensor Data from Touch* IEEE Trans. on Rob. and Automation, Vol. 6, No. 4, Aug. 1990
Allen 91	Allen, P., Yoshimi, B., Timcenko, A. *Real-time visual servoing* Proc. 1991 IEEE Conf. on Robotics and Autom.
Almand 85	Almand, B. *Sensor information fusion and robot decisionmaking methodology* Proc. SPIE, Vol. 579: Intelligent Robots and Computer Vision, 1985
Anderson 79	Anderson, B., Moore, J. *Optimal Filtering* Englewood Cliffs: Prentice Hall, 1979
Andrews 77	Andrews, H., Hunt, B. *Digital Image Restoration* Engelwood Cliffs: Prentice–Hall, 1977
Ayache 87b	Ayache, N., Faugeras, O. *Maintaining representations of the environment of a mobile robot* Proc. Robotics Research: The Fourth Int. Symp., 1987
Ayache 87	Ayache, N., Faugeras, O. *Building a consistent 3D representation of a mobile robot environment by combining multiple stereo views* Proceedings of the Int. Joint Conf. on Art. Intell., 1987
Ayache 88	Ayache, N., Faugeras, O. *Building, Registrating and Fusing Noisy Visual Maps* Int. J. of Rob. Research, Vol. 7, No. 6, 1988. Cambridge: MIT Press
Ayache 88b	Ayache, N. *Construction et fusion de représentations visuelles tridimensionelles – applications à la robotique mobile* Dissertation, INRIA, 1988
Bajcsy 85	Bajcsy, R., Allen, P. *Converging Disparate Sensory Data* Proc. 3rd Int. Symp. on Robotics Research, Cambridge: MIT Press, 1985
Balchen 90	Balchen, J., Dessen, F., Skofteland, G. *Sensor Integration Using State Estimators* ASI Series F, Vol. 63: Traditional and Non–Traditional Robotic Sensors, Berlin: Springer-Verlag, 1990
Bamberg 80	Bamberg, G., Baur, F. *Statistik* München: Oldenbourg Verlag, 1980
Bar–Shalom 88	Bar–Shalom, Y., Fortmann, T. *Tracking and Data Association* Boston: Academic Press, 1988
Barnhill 75	Barnhill, R., Riesenfeld R. *Computer Aided Geometric Design* Academic Press, 1975
Bazzi 89	Bazzi, Y., Mitchell, B. *Evidential integration in a multi–sensor system with dependency information* Proc. SPIE, Vol. 1198, Sensor Fusion II: Human and Machine Strategies, 1989
Bazzi 90	Bazzi, Y., Mitchell, B. *Combination of Evidences with Dependency Information in Distributed Sensor System* Proc. SPIE, Vol. 1306, Sensor Fusion III, 1990
Beckerman 92	Beckerman, M. *A Bayes Maximum Entropy Method for Multi–Sensor Data Fusion* Proc. 1992 IEEE Conf. on Robotics and Autom.
Belknap 86	Belknap, R., Riseman, R., Hanson, A. *The information fusion problem and rule-based hypotheses applied to complex aggregations of image events* Proc. 1986 IEEE Conf. on Robotics and Autom.
Bennekers 86	Bennekers, B., Ramirez, C. *Robotics Obstacle Avoidance Using a Camera and a 3-D Laser Scanner* Proc. Vision 86, Machine Vision Association, 1986
Berger 88	Berger, J. *Statistical Decision Theory and Bayesian Analysis* New York: Springer-Verlag, 1988
Bogler 87	Bogler, P. *Shafer–Dempster Reasoning with Applications to Multisensor Target Identification Systems*

IEEE Trans. Systems, Man, and Cybernetics Vol. SMC–17, Nov. 1987

Bolle 86 Bolle, R., Cooper, D.
On Optimally Combining Pieces of Information, with Application to Estimating 3-D Complex–Object Position from Range Data
IEEE Trans. on Pat. Anal. and Mach. Intell., Vol. PAMI-8, No. 5, 1986

Bonasso 92 Bonasso, R., Antonisse, H., Slock, M.
A Reactive Robot System for Find and Fetch Tasks in an Outdoor Environment
Proc. 10. Nat. Conf. on Art. Intell. (AAAI), 1992

Braid 75 Braid, I.
The Synthesis of Solids Bounded by Many Faces,
Comm. of the ACM 18 (4), 1975

Briot 81 Briot M., Talou, J., Bauzil, G.
The multi–sensors which help a mobile robot find its place
Sensor Rev., Vol. 1, No. 1, 1981

Brooks 85 Brooks, R.
Visual Map Making for a Mobile Robot
Proc. IEEE Int. Conf. on Robotics and Automation, 1985

Brooks 91 Brooks, R.
Intelligence without Representation
Artificial Intelligence 47, 1991

Cameron 88 Cameron, A., Daniel, R., Durrant-Whyte, H.
Touch and Motion
Proc. 1988 IEEE Conf. on Robotics and Autom.

Cameron 89 Cameron, A.
Optimal Tactile Sensor Placement
Proc. 1989 IEEE Conf. on Robotics and Autom.

Caroll 91 Caroll, M., Meng, M., Cadwellendar, W.
Fusion of Ultrasonic and Infrared Signatures for Personnel Detection by a Mobile Robot
Proc. SPIE, Vol. 1611, Sensor Fusion IV, 1991

Caglioti 88 Caglioti, V., Simino, R., Somalvico, M.
Multimedia sensory systems for the solution of the error recovery problem
ASI-Series F, Vol. 43: Sensors and Sensory Systems for Advanced Robots. Berlin: Springer Verlag, 1988

Chair 86 Chair, Z., Varshney, P.
Optimum Data Fusion in Multiple Sensor Detection Systems
IEEE Trans. on Aerospace and Electron. Syst., Vol. 22, No.1, 1986

Chang 86 Chang, K., Chong, C., Bar–Shalom, Y.
Joint Probabilistic Data Association in Distributed Sensor Networks
IEEE Trans. on Automatic Control, Vol. AC–31, No. 10, 1986

Chatila 85 Chatila, R., Laumont, J.
Position Referencing and Consistent World Modeling for Mobile Robots
Proc. 1985 IEEE Conf. on Robotics and Autom.

Chaudhuri 89 Chaudhuri, S., Agrawal, R.
Artificial Intelligence Applications to Command, Control, and Communications Systems/Subsystems
Proc. SPIE, Vol. 1100, Sensor Fusion II, 1989

Cheeseman 85 Cheeseman, P.
In Defense of Probability
Proceedings of the Int. Joint Conf. on Art. Intell., 1985

Chen 87b Chen, S.
A geometric approach to multisensor fusion and spatial reasoning
Proc. 1987 AAAI-Workshop on Spatial Reasoning and Multi-Sensor Fusion. Los Altos: Morgan Kaufman, 1987

Chen 87c Chen, J., Reed, I.
A Detection Algorithm for Optical Targets in Clutter
IEEE Trans. on Aerospace and Electronic Systems, AES-23, No. 1, 1987

Chen 92 Chen, Q., Luh, J.
Multi–Sensor Integration in Intelligent Robotic Workstation
Proc. IEEE 1992 Conf. on Robotics and Automation

Chiu 86 Chiu, S., Morley, D., Martin, J.
Sensor Data Fusion on a parallel Processor
Proc. 1986 IEEE Conf. on Robotics and Autom.

Clark 90 Clark, J., Yuille, A.
Data Fusion for Sensory Information Processing Systems
Boston: Kluwer Academic, 1990

Coons 67 Coons, S.
Surfaces for Computer-Aided Design of Space Forms
MIT Project Mac, Technical Report TR-41, 1967

Cowan 88	Cowan, C. *Model-Based Synthesis of Sensor Location* Proc. 1988 IEEE Conf. on Robotics and Autom.
Crowley 85	Crowley, J. *Dynamic World Modeling for an Intelligent Mobile Robot Using a Rotating Ultra-Sonic Ranging Device* Proc. IEEE Intern. Conf. on Robotics and Automation, 1985
Crowley 86	Crowley, J. *Representation and maintenance of a composite surface model* Proc. IEEE Int. Conf. on Robotics and Automation, 1986
Crowley 87	Crowley, J., Ramparany, F. *Mathematical Tools for Representing Uncertainty in Perception* Proc. 1987 AAAI-Workshop on Spatial Reasoning and Multi-Sensor Fusion. Los Altos: Morgan Kaufman, 1987
Dean 90	Dean, T., Basye, K., Chekaluk, R. *Coping with Uncertainty in a Control System for Navigation and Exploration* Proc. 8th Nat. Conf. on Art. Intell. (AAAI–90), 1990
Dickmanns 91	Dickmanns, E., Christians, T. *Relative 3D–state estimation for autonomous visual guidance of road vehicles* Robotics and Autonomous Systems (7), 1991
Dillmann 82	Dillmann, R. *A Sensor Controlled Gripper with Tactile and Non-Tactile Sensor Environment.* Proc. Int. Conf. on Robot Vision and Sensory Controls, 1982
Donald 92	Donald, B., Jennings, J. *Constructive Recognizability for Task–Directed Robot Programming* Proc. IEEE Intern. Conf. on Robotics and Automation, 1992
Duda 73	Duda, R., Hart, P. *Pattern Classification and Scene Analysis* New York: Wiley, 1973
Durrant-Whyte 85	Durrant-Whyte, H. *Integrating Distributed Sensor Information, an Application to a Robot System Coordinator* Proc. IEEE Int. Conf. on Systems, Mans, and Cybernetics, 1985
Durrant-Whyte 86	Durrant-Whyte, H. *Consistent Integration and Propagation of Disparate Sensor Observations* Proc. 1986 IEEE Conf. on Robotics and Autom.
Durrant-Whyte 87	Durrant-Whyte, H. *Sensor Models and Multisensor Integration* Int. J. of Rob. Research, Vol. 6, No. 3, Fall 1987. Cambridge: MIT Press
Durrant-Whyte 87b	Durrant-Whyte, H. *Integration, Coordination, and Control of Multi–Sensor Robot Systems* Boston: Kluwer Academic, 1990
Durrant-Whyte 87c	Durrant-Whyte, H. *Uncertain Geometry in Robotics* Proc. 1987 IEEE Conf. on Robotics and Autom.
Durrant-Whyte 88	Durrant-Whyte, H. *Consistent Integration and Propagation of Disparate Sensor Observations* Int. J. of Rob. Research, Vol. 7, No. 6, Dec. 1988. Cambridge: MIT Press
Durrant-Whyte 90	Durrant-Whyte, H. *Toward a Fully Decentralized Architecture for Multi–Sensor Data Fusion* Proc. 1990 IEEE Conf. on Robotics and Autom.
Eckmiller 91	Eckmiller, R. *Das BMFT–Verbundvorhaben SENROB: Forschungsintegration von Neuroinformatik, Künstlicher Intelligenz, Mikroelektronik und Industrieforschung zur Steuerung sensorisch geführter Industrieroboter* Proc. 4. Int. GI–Kongreß Wissenbasierte Systeme, Berlin: Springer–Verlag, 1991
Faugeras 86	Faugeras, O., Hebert, M. *The Representation, Recognition, and Locating of 3-D Objects* Int. J. of Rob. Research, Vol. 5, No. 3, 1986. Cambridge: MIT Press
Faugeras 86c	Faugeras, O., Ayache, N., Faverjon, B. *Building Visual Maps by Combining Noisy Stereo Measurements* Proc. IEEE Conf. on Rob. and Automation, 1986
Fedorov 72	Fedorov, V. *Theory of Optimal Experiments* Academic Press, New York, 1972
Fernández 90	Fernández, M., Arifgides, A., Evans, J. *The Procrustes Algorithm for Multisensor Track Fusion* Proc. SPIE, Vol. 1306, Sensor Fusion III, 1990

Flachs 88 Flachs, G., Jordan, J., Carlson, J.
Information fusion methodology
Proc. SPIE, Vol. 931, Sensor Fusion, 1988

Flynn 88 Flynn, A.
Combining Sonar and Infrared Sensors for Mobile Robot Navigation
Int. J. of Rob. Research, Vol. 7, No. 6, Dec. 1988. Cambridge: MIT Press

Freyberger 85 Freyberger, F., Kampmann, P., Karl, G., Schmidt, G.
MICROBE – Ein autonomes, mobiles Robotersystem
VDI–Z (127) No. 7, 1985

Fröhlich 91 Fröhlich, C., Freyberger, F., Karl, G., Schmidt, G.
Multisensor System for an Autonomous Robot Vehicle
in: G. Schmidt, Proc. Int. Workshop Inform. Proc. in Autonomous Mobile Robots, Berlin: Springer–Verlag, 1991

Freericks 92 Freericks, M., Knoll, A.
The Applicative Real–Time Programming Language ALDiSP – Informal Introduction and Formal Semantics
Technischer Bericht 92/26 des Fachbereichs Informatik der Technischen Universität Berlin, 1992

Garvey 81 Garvey, T., Lowrance, J., Fischler, M.
An inference technique for integrating knowledge from disparate sources
Proceedings of the 7. Int. Joint Conf. on Art. Intell., 1981

Grandjean 89 Grandjean, P., Robert de Saint Vincent, A.
3-D modeling of indoor scenes by fusion of noisy range and stereo data
Proc. 1989 IEEE Conf. on Robotics and Autom.

Greenway 91 Greenway, P.
The SKIDS Data Fusion Project
Proc. SPIE, Vol. 1611, Sensor Fusion IV, 1991

Hackett 90 Hackett, J., Shah, M.
Multi-Sensor Fusion: A Perspective
Proc. 1990 IEEE Conf. on Robotics and Autom.

Hager 87 Hager, G., Mintz, M.
Searching for Information
Proc. 1987 AAAI-Workshop on Spatial Reasoning and Multi-Sensor Fusion. Los Altos: Morgan Kaufman, 1987

Hager 89 Hager, G., Mintz, M.
Sensor Modeling and Robust Sensor Fusion
Proc. Robotics Research: The 5. Int. Symp., Cambridge: MIT Press, 1989

Hager 89b Hager, G., Mintz, M.
Taks-Directed Multi-Sensor Fusion
Proc. 1989 IEEE Conf. on Robotics and Autom.

Hager 89c Hager, G.
A Comparison of Information Gathering Approaches
Proc. SPIE, Vol. 1198, Sensor Fusion II: Human and Machine Strategies, 1989

Hager 90 Hager, G.
Computational Methods for Sensor Data Fusion and Sensor Planning
Kluwer Academic, Boston, 1990

Hager 90b Hager, G.
Deciding Not to Decide Using Resource–Bounded Sensing
SPIE Vol. 1383, Sensor Fusion III: 3–D Perception and Recognition

Hager 91 Hager, G., Mintz, M.
Task-directed Sensor Data Fusion and Sensor Planning
Int. J. of Rob. Research, Vol. 10, No. 4, 1991. Cambridge: MIT Press

Hager 91b Hager, G.
Towards Geometric Decision Making in Unstructured Environments
Proc. SPIE, Vol. 1611, Sensor Fusion IV, 1991

Hager 92 Hager, G.
Constraint Solving Methods and Sensor–Based Decision Making
Proc. IEEE 1992 Conf. on Robotics and Automation

Han 90 Han, J., Varshney, P.
Distributed detection of moving optical objects
Proc. SPIE Vol. 1306, Sensor Fusion III, 1990

Hanson 88 Hanson, A., Riseman, E., Williams, T.
Sensor and information fusion from knowledge–based constraints
Proc. SPIE, Vol. 931, Sensor Fusion, 1988

Harmon 86 Harmon, S., Bianchini, G., Pinz, G.
Sensor Data Fusion Through A Distributed Blackboard
Proc. IEEE Int. Conf. on Robotics and Automation, 1986

Hashemipour 88 Hashemipour, H., Roy, S., Laub, A.

Decentralized Structures for Parallel Kalman–Filtering
IEEE Trans. on Automatic Control, Vol. 33, No. 1, 1988

Henderson 83 — Henderson, T., Fai, W.
A Multi-Sensor Integration and Data Acquisition System
IEEE Conf. on Comp. Vision and Pat. Recog., 1983

Henderson 84 — Henderson, T., Shilcrat, E.
Logical Sensor Systems
Journal of Robotics Systems, Vol. 1, No.2, 1984

Henderson 84b — Henderson, T., Fai, W. , Hansen, C.
MKS: A Mulisensor Kernel System
IEEE Trans. on System, Man, and Cybernetics, Vol. SMC-14, No. 5, 1984

Henderson 85 — Henderson, T., Hansen, C., Bhanu, B.
The Synthesis of Logical Sensor Specifications
Proc. SPIE Vol. 579: Intelligent Robots and Computer Vision, 1985

Henderson 85b — Henderson, T., Hansen, C., Bhanu, B.
A Framework for Distributed Sensing and Control
Proceedings of the Int. Joint Conf. on Art. Intell., 1985

Hirota 92 — Hirota, K., Pedrycz, W., Yuda, M.
Fuzzy Set–Based Models of Sensor Fusion
in: Terano, T., Sugeno, M., Mukaidono, M., Shigemasu, U. (eds): Fuzzy Engineering Toward Human Friendly Systems, Amsterdam: IOS Press, 1992

Hirzinger 86 — Hirzinger, G., Dietrich, J.
Multisensory Robots and Sensor-Based Path Generation
Proc. IEEE 1986 Conf on Robotics and Automation

Hirzinger 89 — Hirzinger, G., Heindl, J., Landzettel, K.
Predictive and Knowledge-Based Telerobotic Control Concepts
Proc. 1989 IEEE Conf. on Robotics and Autom.

Hirzinger 91 — Hirzinger, G., Dietrich, J., Heindl, J.
Zum Stand der Robotik-Entwicklung bei der DLR Oberpfaffenhofen
Robotersysteme Heft 4/1991. Berlin: Springer-Verlag

Hirzinger 92 — Hirzinger, G., Heindl, J., Landzettel, K., Brunner, B.
Multisensor Shared Autonomy – A Key Issue in the Space Robot Technology Experiment ROTEX
Proc. IEEE/RSJ Int. Conf. on Intell. Robots and Systems (IROS), 1992

Hoballah 89 — Hoballah, I., Varshney, P.
Neyman-Pearson Detection with Distributed Sensors
Proc. of the 25. IEEE Conf. on Decision and Control, 1989

Holmbom 92 — Holmbom, P., Holmberg, P., Nilsson, A., Odeberg, H.
Multi-Sensor Integration – The Sensor Guided Wine Server
Proc. IEEE/RSJ Int. Conf. on Intell. Robots and Systems (IROS), 1992

Hoppen 90 — Hoppen, P., Knieriemen, T., Puttkamer, E.
Laser-Radar based Mapping and Navigation for an Autonomous Mobile Robot
Proc. 1990 IEEE Conf. on Robotics and Autom.

Hu 87 — Hu, G., Stockman, G.
3-D Scene analysis via fusion of light striped image and intensity image
Proc. 1987 AAAI-Workshop on Spatial Reasoning and Multi-Sensor Fusion. Los Altos: Morgan Kaufman, 1987

Huber 81 — Huber, P.
Robust statistics
New York: Wiley, 1981

Huntsberger 87 — Huntsberger, T., Jayaranmamurthy, S
A framework for multi-sensor fusion in the presence of uncertainty
Proc. 1987 AAAI-Workshop on Spatial Reasoning and Multi-Sensor Fusion. Los Altos: Morgan Kaufman, 1987

Huntsberger 90 — Huntsberger, T.
Comparison of techniques for disparate sensor fusion
SPIE Vol. 1383, Sensor Fusion III: 3–D Perception and Recognition

Hutchinson 88 — Hutchinson, S., Cromwell, R., Kak, A.
Planning sensing strategies in a robot work cell with multi–sensor capabilities
Proc. 1988 IEEE Conf. on Robotics and Autom.

Hutchinson 89 — Hutchinson, S., Kak, A.
Planning sensing strategies in a robot work cell with multi–sensor capabilities
IEEE Trans. on Robotics and Automation, Vol 5., No. 6, 1989

Irving 91 — Irving, W., Tsitsiklis, J.
Some Properties of Optimal Thresholds in Decentralized Detection
Proc. SPIE, Vol. 1611, Sensor Fusion IV, 1991

Jackins 80 — Jackins, C., Tanimoto, S.
Oct-Trees and their Use in Representing Three-Dimensional Objects,
Computer Graphics and Image Processing, Vol. 14, 1980

Jänich 81 — Jänich, K.
Lineare Algebra
Berlin: Springer-Verlag, 1981

Kamberova 89 — Kamberova, G., Mintz, M.
Robust Multi–Sensor Fusion: A Decision Theoretic Approach
Proc. SPIE, Vol. 1198, Sensor Fusion II: Human and Machine Strategies, 1989

Kamberova 91 — Kamberova, G., McKendall, R., Mintz, M.
Multivariate Data Fusion Based on Fixed Geometry Sets
Proc. SPIE, Vol. 1611, Sensor Fusion IV, 1991

Kay 91 — Kay, Y., Lee, S.
A Robust 3–D Motion Estimation with Stereo Cameras on a Robot Manipulator
Proc. 1991 IEEE Conf. on Robotics and Autom.

Kent 86 — Kent, E., Sheiner, M.
Building representations from fusion of multiple views
Proc. IEEE Int. Conf. on Robotics and Automation, 1986

Kowalsky 74 — Kowalsky, H.
Einführung in die lineare Algebra
Berlin: Walter de Gruyter, 1974

Krotkov 88 — Krotkov, E., Kories, R.
Adaptive Control of Cooperating Sensors: Focus and Stereo Ranging with an Agile Camera System
Proc. 1988 IEEE Conf. on Robotics and Autom.

Krzysztofowicz 90 — Krzysztofowicz, R., Long, D.
Fusion of Detection Probabilities and Comparison of Multisensor Systems
IEEE Trans. on Systems, Man, and Cybernetics, May 1990

Kweon 92 — Kweon, I., Kuno, Y., Watanabe, M., Onoguchi, K.
Behaviour–Based Mobile Robot Using Active Sensor Fusion
Proc. IEEE 1992 Conf. on Robotics and Automation

Kweon 92b — Kweon, I., Kanade, T.
High–Resolution Terrain Map from Multiple Sensor Data
IEEE Trans. on Pat. Analysis and Mach. Intell., Vol. 14, No. 2, Feb. 1992

Lim 92 — Lim, J., Cho, D.
Physically Based Sensor Modeling for a Sonar Map in a Specular Environment
Proc. 1992 IEEE Conf. on Robotics and Autom.

Liptser 79 — Liptser, R.
Optimal Stopping Rules
New York: Springer-Verlag, 1979

Liu 91 — Liu, L., Yang, J.
Model-Based Object Classification Using Fused Data
Proc. SPIE, Vol. 1611, Sensor Fusion IV, 1991

Loughlin 82 — Loughlin, C., Hudson, E.
Eye–in–Hand Robot Vision
Proc. Conf. Robot Vision and Sensory Controls, 1982

Lowrance 86 — Lowrance, J., Garvey, T., Strat, T.
A Framework for Evidential Reasoning Systems
Proc. Nat. Conf. on Art. Intell. (AAAI), 1986

Luo 87 — Luo, R., Lin, M.
Multi-sensor integrated intelligent robot for automated assembly
Proc. 1987 AAAI-Workshop on Spatial Reasoning and Multi-Sensor Fusion. Los Altos: Morgan Kaufman, 1987

Luo 87b — Luo, R., Lin, M., Scherp, R.
The issues and approaches of a robot multi-sensor integration
Proc. 1987 IEEE Conf. on Robotics and Autom.

Luo 88 — Luo, R., Lin, M.
Robot Multi-Sensor Fusion and Integration: Optimum Estimation of Fused Sensor Data
Proc. 1988 IEEE Conf. on Robotics and Autom.

Luo 88b — Luo, R., Lin, M., Ralph, R.
Dynamic Multi-Sensor Data Fusion System for Intelligent Robots
IEEE Journal of Robotics and Automation, Vol. 4, No. 4, Aug. 1988

Luo 90 — Luo, R., Lin, M.
Hierarchical Robot Multi-Sensor Data Fusion System
ASI Series F, Vol. 58: Highly Redundant Sensing in Robotic Systems. Berlin: Springer-Verlag, 1990

Malik 87 — Malik, J.

	Recovering three dimensional shape from a single image of curved objects Proceedings of the Int. Joint Conf. on Art. Intell., 1987
Matthies 87	Matthies, L., Shafer, S. *Error Modeling in Stereo Navigation* IEEE Journal of Robotics and Automation, Vol RA–3, No. 3, 1987
Matthies 88	Matthies, L., Elfes, A. *Integration of Sonar and Stereo Range Data Using a Grid-Based Representation* Proc. 1988 IEEE Conf. on Robotics and Autom.
McKendall 90	McKendall, R., Mintz, M. *Using robust statistics for sensor fusion* SPIE Vol. 1383, Sensor Fusion III: 3–D Perception and Recognition
Meinköhn 92	Meinköhn, J., Knoll, A. *Die Modellierung nichthierarchischer Sensorverbunde als Vertragsnetze mit autonomen Sensoragenten* Technischer Bericht 92/28 des Fachbereichs Informatik der Technischen Universittät Berlin, 1992
Miller 91	Miller, D., Slack, M. *Global Symbolic Maps from Local Navigation* Proc. 9. Nat. Conf. on Art. Intell, AAAI–91, 1991
Mintz 90	Mintz, M. *Comments on „Dynamic Multi-Sensor Data Fusion System for Intelligent Robots“* IEEE Trans. on Rob. and Automation, Vol. 6, No. 1, Feb. 1990
Moravec 89	Moravec, H. *Sensor Fusion in Certainty Grids for Mobile Robots* ASI Series F, Vol. 52, Berlin: Springer–Verlag, 1989
Nakamura 89	Nakamura, Y., Xu, Y. *Geometrical fusion method for multi-sensor robotic systems* Proc. 1989 IEEE Conf. on Robotics and Autom.
Nii 86	Nii, H. *Blackboard systems: The blackboard model of problem solving and the evolution of blackboard architectures* AI magazine, Vol 7., No. 2, 1986
Nishida 87	Nishida, T., Yamada, A., Doshita, S. *Figuring out most plausible interpretation from spatial constraints* Proc. 1987 AAAI-Workshop on Spatial Reasoning and Multi-Sensor Fusion. Los Altos: Morgan Kaufman, 1987
Papanikolopoulos 91	Papanikolopoulos, N., Khosla, P., Kanade, T. *Vision and control techniques for robotic visual tracking* Proc. 1991 IEEE Conf. on Robotics and Autom.
Pollard 87	Pollard, S., Porrill, J., Mayhew, E., Frisby, J. *Matching geometrical descriptions in three space* Image and Vision Computing 5(2), 1987
Porrill 87	Porrill, J., Pollard, S., Mayhew, E. *Optimal combination of multiple sensors including stereo vision* Image and Vision Computing 5 (2), 1987
Porrill 87b	Porrill, J., Pollard, S., Pridmore, T., Bowen, J., Mayhew, E., Frisby, J. *TINA: The Sheffield AIVRU vision system* Proceedings of the Int. Joint Conf. on Art. Intell., 1987
Porrill 88	Porrill, J. *Optimal Combination and Constraints for Geometrical Sensor Data* Int. J. of Rob. Research, Vol. 7, No. 6, Dec. 1988. Cambridge: MIT Press
Raczkowsky 91	Raczkowsky, J. *Multisensordatenverarbeitung in der Robotik* Informatik Fachberichte 268, Berlin: Springer–Verlag, 1991
Rao 87	Rao, K., Nevatia, R., Medioni, G. *Issues in shape description and an approach for working with sparse data* Proc. 1987 AAAI-Workshop on Spatial Reasoning and Multi-Sensor Fusion. Los Altos: Morgan Kaufman, 1987
Requicha 80	Requicha, A. *Representations for Rigid Solids: Theory, Methods and Systems* Computing Surveys 12(4), 1980
Richardson 88	Richardson, J., Marsh, K. *Fusion of Multisensor Data* Int. J. of Rob. Research, Vol. 7, No. 6, Dec. 1988. Cambridge: MIT Press
Robert 86	Robert de Saint Vincent, A. *A 3D Perception System for the Mobile Robot Hilare* Proc. 1986 IEEE Conf. on Robotics and Automation
Ruokangas 86	Ruokangas, C., Black, M., Martin, J., Schoenwald, J.

	Integration of multiple sensors to provide flexible control strategies Proc. 1986 IEEE Conf. on Robotics and Autom.
Rozmann 92	Rozmann, M., Detlefsen, J. *Environmental Exploration based on a Three-Dimensional Imaging Radar Sensor* Proc. IEEE/RSJ Int. Conf. on Intell. Robots and Systems (IROS), 1992
Sakane 91	Sakane, S., Sato, T. *Automatic Planning of Light Source and Camera Placement for an Active Photometric System* Proc. 1991 IEEE Conf. on Robotics and Autom.
Sasiadek 87	Sasiadek, J., Wojcik, P. *Tactile Sensor Sioganl Processing Using an Adaptive Kalman–Filter* Proc. 1987 IEEE Conf. on Robotics and Autom.
Sato 92	Sato, K., Ikeuchi, K., Kanade, T. *Model–Based Recognition of Specular Objects Using Sensor Models* CVGIP: Image Understanding, Vol. 55, No. 2, March 1992
Schoess 88	Schoess, J., Castore, G. *A Distributed Sensor Architecture for Advanced Aerospace Systems* Proc. SPIE, Vol. 931, Sensor Fusion, 1988
Scholz 88	Scholz, P. *Erkennung einfacher Objekte mit einem Kamerasystem unter Zuhilfenahme eines entfernungsmessenden Sensors* Diplomarbeit, TU Berlin, Institut für Techn. Inform., 1988
Schürmann 91	Schürmann, B., Hirzinger, G., Hernández, D., Simon, H., Hackbarth, H. *Neural Control within the BMFT–Project Neres* Proc. 4. Int. GI–Kongreß Wissenbasierte Systeme, Berlin: Springer–Verlag, 1991
Shaw 88	Shaw, S., deFigueiredo, R., Krishen, K. *Fusion of radar and optical sensors for space robotic vision* Proc. 1988 IEEE Conf. on Robotics and Autom.
Shekhar 86	Shekhar , S., Khatib, O., Shimojo, M. *Sensor Fusion and Object Localization* Proc. 1986 IEEE Conf. on Robotics and Autom.
Shekhar 88	Shekhar, S., Khatib, O., Shimojo, M. *Object Localization with Multiple Sensors* Int. J. of Rob. Research, Vol. 7, No. 6, Dec. 1988. Cambridge: MIT Press
Shen 90	Shen, H., Basir, O. *A Scheme for Sensory Data Integration* SPIE Vol. 1383, Sensor Fusion III: 3–D Perception and Recognition, 1990
Shortliffe 84	Shortliffe, E., Buchanan, B. *A Model of Inexact Reasoning in Medicine* in: Buchanan, B., Shortliffe, E.(Eds.): Rule-Based Expert Systems, Addison-Wesley, 1984
Sikka 89	Sikka, D., Varshney, P., Vannicola, V. *A distributed Artificial Intelligence approach to object identification and classification* Proc. SPIE, Vol. 1198, Sensor Fusion II, 1989
Silverman 87	Silverman, G. *Locating polyhedral objects from edge point data* Proceedings of the Int. Joint Conf. on Art. Intell., 1987
Skofteland 91	Skofteland, G., Hirzinger, G. *Computing Position and Orientation of a Freeflying Polyhedron from 3D Data* Proc. 1991 IEEE Conf. on Robotics and Autom.
Smith 86	Smith, R., Cheeseman, P. *On the Representation and Estimation of Spatial Uncertainty* Int. J. on Robotics Research, Vol. 5, No. 4, 1986
Srinivasan 86	Srinivasan, R. *Distributed Radar Detection Theory* Proc. IEE, Vol. 133, No. 1, Feb. 1986
Stansfield 88	Stansfield, S. *A Robotic Perceptual System Utilizing Passive Vision and Active Touch* Int. J. of Rob. Research, Vol. 7, No. 6, Dec. 1988. Cambridge: MIT Press
Stelmaszyk 91	Stelmaszyk, P., Ishiguro, H., Tsuji, S. *Mobile Robot Navigation by an Active Control of the Vision System* Proc. 12. Int. Joint Conf. on Art. Intell., 1991
Tenney 81	Tenney, R., Sandell, N. *Detection with Distributed Sensors* IEEE Trans. on Aerospace and Electron. Syst., Vol. 17, No. 4, 1981
Tenney 81b	Tenney, R., Sandell, N. *Structures for Distributed Decisionmaking* IEEE Trans. Systems, Man, and Cybernetics, Vol. SMC–11, No. 1, Jan 1981

Tenney 81c	Tenney, R., Sandell, N. *Strategies for Distributed Decisionmaking* IEEE Trans. Systems, Man, and Cybernetics, Vol. SMC–11, No. 1, Jan 1981
Thomopoulos 91	Thomopoulos, S., Papadakis, I., Sahinoglou, H., Otello, N. *Centralized and Distributed Hypothesis Testing with Structured Adaptive Networks and Perceptron-Type Neural Networks* Proc. SPIE, Vol. 1611, Sensor Fusion IV, 1991
Tirumalai 90	Tirumalai., A., Schunck, B., Jain, R. *Robust Self–Calibration and Evidential Reasoning for Building Environment Maps* SPIE Vol. 1383, Sensor Fusion III: 3–D Perception and Recognition, 1990
Tou 90	Tou, J., Balchen, J. *Highly Redundant Sensing in Robotic Systems* ASI Series F, Vol. 58, Berlin: Springer-Verlag, 1990
Trivedi 90	Trivedi, M., Abidi, M., Eason, R., Gonzalez, R. *Developing Robotic Systems with Multiple Sensors* IEEE Trans. on Systems, Man, and Cybernetics, Vol. 20, No. 6, Nov. 1990
Tsai 87	Tsai, R., Silverman, G., Lavin, M. *Locating polyhedral objects from edge point data* Proc. 1987 AAAI-Workshop on Spatial Reasoning and Multi-Sensor Fusion. Los Altos: Morgan Kaufman, 1987
Tsai 89	Tsai, R., Tarabanis, K. *Model–Based Planning of Sensor Placement and Optical Settings* Proc. SPIE, Vol. 1198, Sensor Fusion II: Human and Machine Strategies, 1989
Vemuri 90	Vemuri, B., Skofteland, G. *Computing Motion Parameters from Sparse Multi–Sensor Range Data for Tele–Robotics* SPIE Vol. 1383, Sensor Fusion III: 3–D Perception and Recognition, 1990
Weiß 87	Weiß, S. *Untersuchung der Anwendbarkeit von Fuzzy-Logik zur Interpretation von Sensordaten* Diplomarbeit, Lehrstuhl für Prozeßrechentechnik der Universität Karlsruhe, 1987
Wen 92	Wen, W., Durrant–Whyte, H. *Model–based Multi–sensor Data Fusion* Proc. IEEE Int. Conf. on Robotics and Automation, 1992
Wise 86	Wise, B. *An Experimental Comparison of Uncertain Inference Systems* Ph. D. Thesis, Carnegie–Mellon Univ., June 1986
Wolfram 92	Wolfram, A. *Untersuchung der Anwendbarkeit unscharfen Schließens bei der Fusion von Daten in Multisensorsystemen* Studienarbeit am Institut für Technische Informatik der TU Berlin, 1992
Yoshida 92	Yoshida, K., Hirose, S. *Real–Time Stereo Vision with Multiple Arrayed Camera* Proc. 1992 IEEE Conf. on Robotics and Autom.
Zytkow 89	Zytkow, J., Pachowicz, P. *Fusion of vision and touch for spatio–temporal reasoning in learning manipulation tasks* Proc. SPIE, Vol. 1198, Sensor Fusion II: Human and Machine Strategies, 1989