

Fahrzeugklassifikation und Schätzung streckenbezogener Verkehrsdaten mit Methoden der Mustererkennung

Vom Promotionsausschuß der
Technischen Universität Hamburg-Harburg
zur Erlangung des akademischen Grades
Doktor-Ingenieur
genehmigte Dissertation

von

Jens Wohlers

aus Rendsburg

1997

1. Gutachter

Prof. Dr.-Ing. N. Fliege

2. Gutachter

Prof. Dr.-Ing. M. Cremer

3. Gutachter

Dr.-Ing. habil. A. Mertins

Tag der mündlichen Prüfung:

29. August 1997

Vorwort

Die vorliegende Arbeit entstand während meiner Tätigkeit als wissenschaftlicher Mitarbeiter im Arbeitsbereich Nachrichtentechnik der Technischen Universität Hamburg-Harburg.

Mein besonderer Dank gilt Herrn Prof. Dr.-Ing. N. Fliege für die Möglichkeit der Durchführung und die Betreuung dieser Arbeit.

Herrn Prof. Dr.-Ing. M. Cremer danke ich für das Interesse an der Arbeit und für die Übernahme des Gutachtens.

Herrn Dr.-Ing. habil. A. Mertins danke ich ebenfalls für die Übernahme des Gutachtens sowie für viele wertvolle Sachdiskussionen und Anregungen und für die besonders gute Zusammenarbeit.

Schließlich möchte ich mich bei allen Mitarbeitern und Mitarbeiterinnen des Arbeitsbereiches, insbesondere bei Herrn Dipl.-Ing. A. Manea, für die freundschaftliche und sehr angenehme Arbeitsatmosphäre bedanken.

Hamburg, im August 1997

Jens Wohlers

Inhaltsverzeichnis

1	Einführung	1
2	Signalerfassung und -vorverarbeitung	4
2.1	Signalerfassung	4
2.2	Signalvorverarbeitung	5
3	Diskrete Signaltransformationen zur Merkmalsextraktion	11
3.1	Signalapproximation mit orthonormalen Reihenentwicklungen	12
3.2	Diskrete Karhunen-Loève-Transformation	13
3.3	Whitening-Transformation	15
3.4	Merkmalsextraktion mit der Karhunen-Loève-Transformation	16
3.4.1	Klassenspezifische Transformation	17
3.4.2	Gemeinsame Transformation	18
3.4.3	Anwendung der Karhunen-Loève-Transformation auf Fahrzeugsignale	19
3.5	Diskriminanzanalyse	21
4	Fahrzeugklassifikation	25
4.1	Mustererkennungssystem	26
4.2	Signalvorverarbeitung und Merkmalsextraktion	27
4.3	Klassifikation	28
4.3.1	Bayes-Klassifikation	29
4.3.2	Maximum-a-posteriori- und Maximum-Likelihood-Klassifikation	31
4.4	Parametrische Klassifikationsverfahren	33
4.5	Nichtparametrische Klassifikationsverfahren	35
4.5.1	Nächster-Nachbar-Klassifikator	35

4.5.2	Polynomklassifikator	37
4.5.3	Histogrammklassifikator	38
4.5.4	Klassifikation mit neuronalen Netzwerken	40
4.6	Mehrstufige Klassifikationsverfahren	43
4.7	Verfahrensvergleich	46
4.8	Schätzung des Bayes-Fehlers	54
5	Streckenbezogene Verkehrsdatenerfassung	57
5.1	Mustererkennungssystem	58
5.1.1	Signalvorverarbeitung und Merkmalsextraktion	60
5.1.2	Entscheidungsfunktion	62
5.1.3	Parameterfestlegung	63
5.1.4	Voruntersuchungen	66
5.2	Schätzverfahren mit Fahrzeugfolgenkorrelation	68
5.2.1	Der Schätzalgorithmus	68
5.2.2	Simulationsergebnisse	76
5.3	Schätzverfahren mit Kostenfunktionen	79
5.3.1	Der Schätzalgorithmus	80
5.3.2	Optimierung des Verfahrens	84
5.3.3	Simulationsergebnisse	89
5.4	Schätzverfahren mit Hidden-Markov-Modellen	92
5.4.1	Hidden-Markov-Modelle	93
5.4.2	Der Schätzalgorithmus	103
5.4.3	Optimierung des Verfahrens	106
5.4.4	Beispiel zur Fahrzeugwiedererkennung	112
5.4.5	Simulationsergebnisse	117
5.5	Verfahrensvergleich	118
6	Verkehrszustandsschätzung mit streckenbezogenen Verkehrsdaten	120
6.1	Das makroskopische Verkehrsmodell	122

6.1.1	Variablen des Verkehrsflußmodells	122
6.1.2	Modellvalidierung	125
6.1.3	Modellerweiterung zur Störfallerkennung	126
6.1.4	Modellerweiterung mit streckenbezogenen Verkehrs- daten	127
6.2	Modellgestützte Verkehrszustandsschätzung	129
6.2.1	Das stochastische Verkehrsflußmodell	129
6.2.2	Zustandsraumdarstellung	131
6.2.3	Zustandsschätzung mit dem erweiterten Kalman- Filter	132
6.2.4	Verkehrszustandsschätzung mit Reisezeitinformation	134
6.3	Beobachtbarkeitsanalyse	135
6.3.1	Theoretische Grundlagen	136
6.3.2	Analyse des erweiterten Verkehrsflußmodells	138
6.4	Simulationsergebnisse	141
6.4.1	Parameterfestlegung	141
6.4.2	Ergebnisse der Verkehrszustandsschätzung	143
6.4.3	Ergebnisse der Störungserkennung	145
7	Zusammenfassung	147
	Abkürzungs- und Symbolverzeichnis	149
	Literaturverzeichnis	153

1 Einführung

Zur Verbesserung der Straßenverkehrssicherheit und -effizienz sowie der durch den Straßenverkehr verursachten Umweltschäden wurden von der Europäischen Union Mitte der achtziger Jahre Forschungsprogramme wie DRIVE (**D**edicated **R**oad **I**nfrastructure for **V**ehicle Safety in **E**urope) und PROMETHEUS (**P**rogram for an **E**uropean Traffic with **H**ighest **E**fficiency and **U**nprecedented **S**afety) ins Leben gerufen. Parallel wurden in den Vereinigten Staaten Forschungsarbeiten zur Automatisierung des Straßenverkehrs durchgeführt, die in die Programme IVHS (**I**ntelligent **V**ehicle-**H**ighway **S**ystems) und ATT (**A**dvanced **T**ransport **T**elematics) eingebettet sind.

Fazit der Programme ist, daß im Zuge der stetig wachsenden Zahl von Kraftfahrzeugen und einer weiterhin zunehmenden Mobilität der Bevölkerung eine Optimierung des Verkehrsablaufs durch Straßenverkehrskontrollsysteme unumgänglich erscheint. Die Systeme müssen jederzeit einwandfreie Informationen über den aktuellen Verkehrszustand liefern und zusätzlich in einem Streckenabschnitt auftretende Störungen detektieren. Durch Stau- und Unfallmeldungen können die Verkehrsteilnehmer speziell auf Autobahnen direkt auf der Strecke über kritische Verkehrssituationen informiert und gewarnt sowie notfalls über andere Autobahnabschnitte umgeleitet werden.

Zur Bereitstellung der notwendigen Informationen wurden in den letzten zwanzig Jahren Schätzverfahren entwickelt, die eine Prognose des Verkehrszustandes liefern und Störungen im Verkehr detektieren. Die zur Schätzung notwendigen Daten werden heutzutage vorwiegend durch Induktionsschleifen zur Verfügung gestellt. Mit diesen Sensoren können zum einen querschnittsbezogene Verkehrsdaten wie Fahrzeuggeschwindigkeiten und Verkehrsstärken erfaßt werden, zum anderen ist es auch möglich, anhand der vom Fahrzeug abhängigen Induktivitätsänderung der Schleife Fahrzeugmuster aufzunehmen und mit diesen Mustern Fahrzeuge zu

klassifizieren und streckenbezogene Verkehrsdaten wie die Reisezeit eines Fahrzeugs zu ermitteln.

Die Klassifikation von Fahrzeugen zur statistischen Auswertung des Fahrzeugverkehrs wurde erstmals in den siebziger Jahren mit relativ einfachen Methoden durchgeführt [57], [28], [68] und Mitte der achtziger Jahre mit Verfahren der Merkmalsextraktion verbessert [11]. Der nächste Schritt in der Auswertung von Fahrzeugmustern liegt in der genauen Klassifizierung von Einzelfahrzeugen, wodurch Fahrzeuge oder Fahrzeugkollektive an einem zweiten Meßquerschnitt wiedererkannt und somit streckenbezogene Verkehrsdaten geschätzt werden können. Dieses Forschungsgebiet wird seit den achtziger Jahren bearbeitet und erste Ergebnisse werden heute in verschiedenen Verkehrsleitsystemen eingesetzt [11], [62], [47]. Die Untersuchung der optimalen Entscheidungsregel sowohl für die Klassifikation als auch für die Schätzung streckenbezogener Verkehrsdaten ist hierbei in den bisherigen Arbeiten nahezu unberücksichtigt geblieben.

Die Zielsetzung der vorliegenden Arbeit liegt in der Untersuchung und im Vergleich verschiedener Verfahren zur Klassifikation und zur individuellen Wiedererkennung von Fahrzeugen. Beide Probleme lassen sich mit Methoden der Mustererkennung lösen. Diese Methoden führen auf die Bayes-Klassifikation [31], bei der die Entscheidungsregeln dahingehend optimiert werden, daß die Fehlerrate eines Mustererkennungssystems minimal wird. Daraus ergeben sich eine Vielzahl von Klassifikationsverfahren. Durch das Auffinden effizienter Lernverfahren und einer Vielzahl an Forschungsarbeiten im Rahmen der Spracherkennung haben in den letzten Jahren *neuronale Netzwerke* und *Hidden-Markov-Modelle* Einzug in zahlreiche technische Applikationen erhalten. Diese Verfahren werden in dieser Arbeit ebenfalls untersucht und zur Klassifikation bzw. streckenbezogenen Verkehrsdatenerfassung eingesetzt.

In Kapitel 2 wird die Erfassung von Fahrzeugen mit Induktionsschleifendetektoren beschrieben. Die erfaßten Fahrzeugsignale können nach einer Normierung in Zeit und Amplitude, wodurch die Signale von nicht-fahrzeugspezifischen Eigenschaften befreit werden, zur Klassifikation und zur streckenbezogenen Verkehrsdatenerfassung eingesetzt werden.

In Kapitel 3 sind diskrete Signaltransformationen zusammengestellt, die im vorliegenden Fall ihren Einsatz in der Gewinnung geeigneter Merkmale für die in dieser Arbeit untersuchten Schätzverfahren finden.

Kapitel 4 bildet den ersten Schwerpunkt dieser Arbeit und stellt die Klassifikation von Fahrzeugen dar. Hierbei wird nach der Präsentation eines Mustererkennungssystems auf die Bayes-Klassifikation und ihre Durchführung eingegangen. Dabei wird zwischen nichtparametrischen und parametrischen Verfahren unterschieden, wobei unter letztere auch die neuronalen Netzwerke fallen. Es wird gezeigt, daß mit den hier aufgeführten Klassifikatoren eine Fehlerrate erzielt werden kann, die nahe beim optimalen Bayes-Fehler liegt.

Im zentralen Kapitel 5 wird die Schätzung streckenbezogener Verkehrsdaten behandelt. Ausgehend von einem mit einer Korrelationsanalyse arbeitenden Referenzverfahren [61], das heutzutage eine weite Verbreitung gefunden hat, werden zwei neue, unterschiedliche Schätzmethoden präsentiert, die mit einer Wiedererkennung von Individualfahrzeugen im Fahrzeugverbund arbeiten. Das erste Verfahren arbeitet mit Kostenfunktionen auf der Basis von Fahrzeugmerkmalen und lokalen Streckenprognosen, das zweite Verfahren berücksichtigt die statistischen Bindungen der Fahrzeuge im Streckenabschnitt, die durch Hidden-Markov-Modelle modelliert werden. Die Verfahren werden anhand von Simulations- und Meßdaten bewertet und abschließend einander gegenübergestellt.

Aufbauend auf die Möglichkeit der streckenbezogenen Verkehrsdatenerfassung befaßt sich Kapitel 6 mit der Verkehrszustandsschätzung mit Reisezeitinformation. Das vorgestellte Modell zur Zustandsschätzung geht auf CREMER zurück [22] und arbeitet mit querschnitts- und streckenbezogenen Verkehrsdaten [25]. Es wird gezeigt, daß eine Erweiterung des Modells um die Reisezeit und die anschließende Schätzung des Verkehrszustandes mit einem erweiterten Kalman-Filter eine Verbesserung der Schätzergebnisse bringt und eine zuverlässige Störungsdetektion ermöglicht.

In Kapitel 7 erfolgt eine Zusammenfassung der Ergebnisse dieser Arbeit sowie eine abschließende Diskussion.

2 Signalerfassung und -vorverarbeitung

In diesem Kapitel wird zunächst die Erfassung von Fahrzeugsignalen mit Induktionsschleifendetektoren erläutert. Die erfaßten Fahrzeugsignale können zur Klassifikation und zur Wiedererkennung und damit zur streckenbezogenen Verkehrsdatenerfassung eingesetzt werden. Die Signale müssen allerdings durch geeignete Vorverarbeitungsverfahren von nicht fahrzeugspezifischen Eigenschaften befreit werden. Dieses gelingt durch eine Normierung der Signale in Zeit und Amplitude. Aus den normierten Fahrzeugsignalen können die Merkmalsvektoren für das entsprechende Mustererkennungssystem ermittelt werden.

2.1 Signalerfassung

Neben Infrarot- und Radardetektoren sind Induktionsschleifendetektoren die derzeit geeignetesten Sensoren zur Detektion von Fahrzeugen im Straßenverkehr. Dieses gilt vor allem unter dem Gesichtspunkt der Verfügbarkeit, der Zuverlässigkeit und der Kosten. Die Induktionsschleife ist eine in die Fahrbahn eingelassene, von einem Wechselstrom durchflossene Leiterschleife mit mehreren Windungen. Die Schleifenbreite wird durch die Fahrbahnbreite bestimmt, die Kantenlängen der Schleife variieren je nach Anwendung und betragen meistens 1 m oder 2,5 m. Der Wechselstrom erzeugt ein elektromagnetisches Feld, in dem metallische, leitfähige Oberflächen über den Effekt der Feldverdrängung eine kleine Änderung der Induktivität der Schleife verursachen. In dem Detektor wird diese Induktivitätsänderung $L + \Delta L(t)$ in ein Signal umgesetzt, das zur Detektion von Fahrzeugen eingesetzt werden kann (Bild 2.1). Dieses Signal wird hauptsächlich von der Metallmassenverteilung der Fahrzeuge im Unterbodenbereich bestimmt, und somit ist die Induktivitätsänderung neben

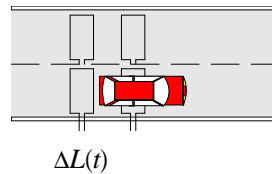


Bild 2.1: Erfassung eines Fahrzeugsignals

der Schleifengeometrie auch von der Fahrzeugart abhängig.

Verschiedene Meßverfahren finden sich in [10]. Ein neuartiges Verfahren zur Erfassung von Fahrzeugen mit Induktionsschleifen ist in [34] beschrieben. Bild 2.2 zeigt einige Beispiele von Fahrzeugsignalen $\Delta L(t)$ für verschiedene Fahrzeugklassen. Die Signale wurden mit 1m-Schleifen erfaßt. Die Geschwindigkeiten der Fahrzeuge liegen im Bereich von 50 bis 80 km/h. Deutlich erkennbar sind die von der Metallmassenverteilung hervorgerufenen unterschiedlichen Signalverläufe und Signalamplituden der verschiedenen Fahrzeugklassen.

2.2 Signalvorverarbeitung

Die am Ausgang des Induktionsschleifendetektors gewonnenen Meßdaten $\Delta L(t)$ beinhalten irrelevante Eigenschaften und Einflüsse des Meßaufnehmers und des Fahrzeugs. Hierunter fallen die geometrischen Abmessungen der Induktionsschleife, die Empfindlichkeit und Verstärkung des Detektors, Fahrbahneigenschaften, die Fahrzeuggeschwindigkeit und das Fahrverhalten. Zur Beseitigung dieser Einflüsse muß eine Normierung der Fahrzeugsignale in Zeit und Amplitude vorgenommen werden, so daß anschließend aussagekräftige und gut reproduzierbare Merkmalsvektoren aus den Schleifensignalen ermittelt werden können. Bei der Normierung können die nichtlinearen Verzerrungen, die z.B. bei einem versetzten Fahren über die Schleife entstehen, im Rahmen der Meßgenauigkeit unberücksichtigt bleiben [61].

Die Signalvorverarbeitung beinhaltet folgende Schritte:

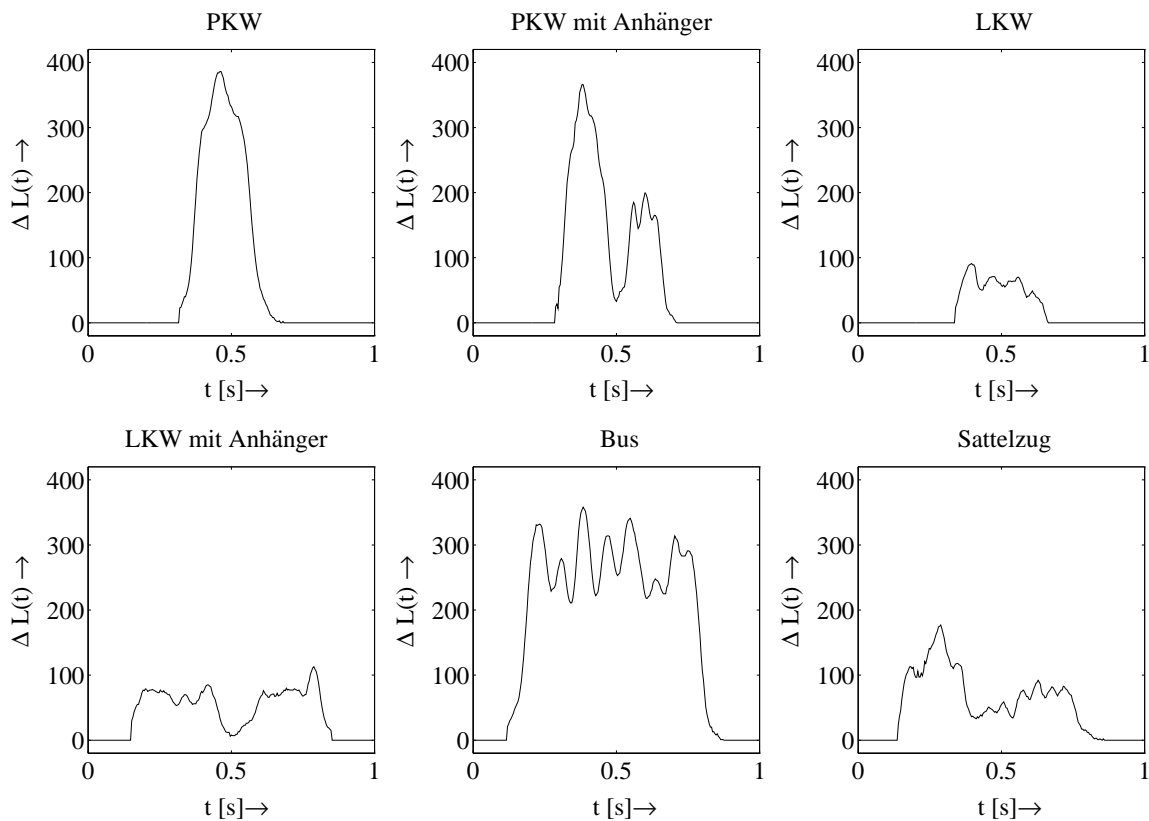


Bild 2.2: Fahrzeugsignale

Analog-Digital-Umsetzung

Nach einer Anti-Aliasing-Filterung werden aus dem zeitkontinuierlichen Signal durch eine Analog-Digital-Umsetzung Abtastwerte der Fahrzeugsignale gewonnen [33].

Signaldetektion

Die Signaldetektion erfolgt in zwei Schritten. Nach einer ersten Schwellwertabfrage erfolgt eine amplitudenunabhängige Erkennung des Signalanfangs und -endes. Die feste Schwelle muß so festgelegt werden, daß nahezu alle Fahrzeuge detektiert werden und die durch Rauschvorgänge und Meßfehler verursachten Störungen unberücksichtigt bleiben. Eine zweite, amplitudenabhängige Schwelle sorgt dafür, daß nur noch repräsentative Meßsignale weiterverarbeitet werden und die nachfolgende Normierung sowie die Geschwindigkeits- und Längenbestimmung vom Fahrverhalten unabhängig ist.

Zeitnormierung

Die Abtastwerte des Fahrzeugsignals können in einem Intervall der Länge ℓT zu einem zeitabhängigen Vektor

$$\begin{aligned} \mathbf{r}(kT) &= [\Delta L(kT), \Delta L([k+1]T), \dots, \Delta L([k+\ell-1]T)]^T \\ &= [r(kT), r([k+1]T), \dots, r([k+\ell-1]T)]^T \end{aligned} \quad (2.1)$$

zusammengefaßt werden. Hierbei ist ℓ eine natürliche Zahl und noch abhängig von der Fahrzeuggeschwindigkeit, der Fahrzeuglänge und der Schleifengeometrie. Zur Elimination der Geschwindigkeit und der Schleifenabmessung erfolgt eine Abtastratenumwandlung durch Interpolation bzw. durch Dezimation auf eine feste Signallänge n . Aufwendige Interpolationsverfahren, wie sie z.B. in [87] aufgeführt sind, sind bei dieser Anwendung nicht notwendig, da hochfrequente Signalanteile für die Klassifikation und die Wiedererkennung eine untergeordnete Rolle spielen. Vielmehr können die neuen Abtastwerte durch eine lineare Interpolation bzw. Dezimation bestimmt werden und dem nächsten Vorverarbeitungsschritt zugeführt werden.

Tiefpaßfilterung

Durch eine Tiefpaßfilterung im normierten Signalraum gelingt es, mit geringem Aufwand eine Unabhängigkeit der normierten Signale von der Fahrzeuggeschwindigkeit zu erreichen und gleichzeitig hochfrequente Störanteile zu unterdrücken. Gleichzeitig mit der Tiefpaßfilterung kann die Abtastrate noch um einen ganzzahligen Faktor reduziert werden. Diese Reduktion sorgt für eine Verringerung des Rechenaufwands bei der anschließenden Merkmalsextraktion (siehe Kapitel 3).

Amplitudennormierung

Da die Signalamplitude von der Empfindlichkeit des Meßaufnehmers abhängig ist, erfolgt eine Normierung auf eine einheitliche Signalamplitude.

Energienormierung

Eine Energienormierung ermöglicht eine Unabhängigkeit des Signalvektors von der Signalenergie.

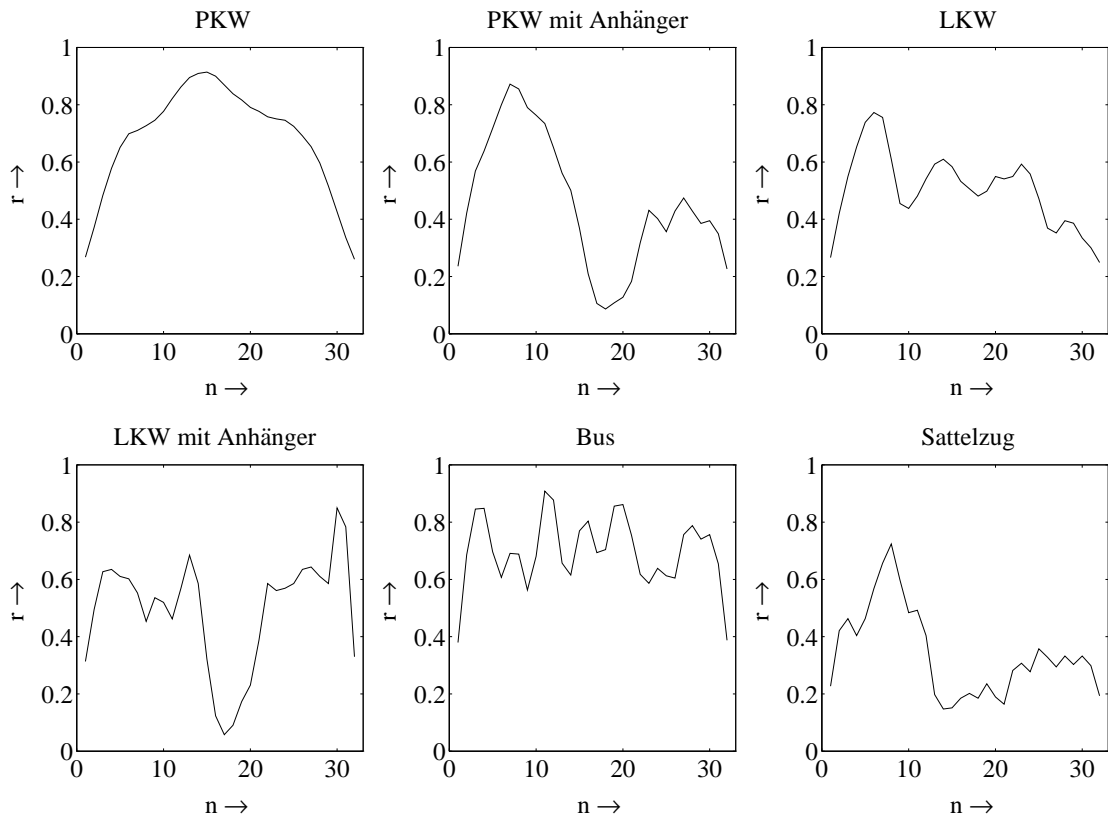


Bild 2.3: Normierte Fahrzeugsignale

Aus der Signalvorverarbeitung ergeben sich zeitabhängige Vektoren

$$\mathbf{r}_n(kT_n) = [r_n(kT_n), r_n([k+1]T_n), \dots, r_n([k+n-1]T_n)]^T. \quad (2.2)$$

Die Dimension der Vektoren, n , ist abhängig von der gewünschten Anwendung und muß sowohl bei der Fahrzeugklassifikation als auch bei der Fahrzeugwiedererkennung experimentell bestimmt werden.

In Bild 2.3 sind die zu Bild 2.2 gehörenden energienormierten Fahrzeugsignale für $n = 32$ dargestellt.

Geschwindigkeits- und Längenbestimmung

Während der Signalnormierung können Fahrzeuggeschwindigkeit und -länge bestimmt werden. Dieses geschieht mit Hilfe von Doppelschleifen. Als Doppelschleife wird ein Induktionsschleifenpaar bezeichnet, dessen Kopfabstand sehr gering ist, so daß eine eindeutige Zuordnung der Detektorausgangssignale zu einem Fahrzeug möglich ist. Die Fahrzeugge-

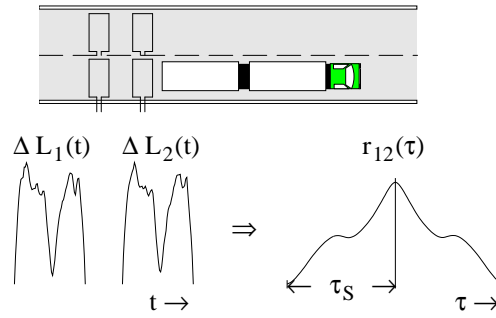


Bild 2.4: Bestimmung der Zeitverzögerung τ_S

schwindigkeit ergibt sich aus der Zeitdifferenz τ_S der Fahrzeugdetektion an den Schleifen und dem Kopfabstand d zu

$$v_{Fz} = \frac{d}{\tau_S}. \quad (2.3)$$

Die Zeitdifferenz kann aus den gemittelten Detektionszeiten des Signalanfangs und -endes an den Schleifen oder genauer, aber auch rechenintensiver, aus dem Maximum der Kreuzkorrelationsfolge der beiden Fahrzeugsignale berechnet werden (siehe Bild 2.4).

Die Fahrzeuglänge ergibt sich aus der Fahrzeuggeschwindigkeit v_{Fz} , der Signallänge τ_{Fz} , die aus der ersten Schwellwertabfrage ermittelt werden kann, sowie der Länge der Induktionsschleife in Fahrtrichtung l_S zu

$$l_{Fz} = \tau_{Fz} v_{Fz} - l_S. \quad (2.4)$$

Die Fahrzeuglänge ist ein wesentliches Merkmal eines Fahrzeugs und wird zusätzlich zum Fahrzeugsignal zur Klassifikation und Wiedererkennung eingesetzt.

Falls die Meßstelle lediglich mit einer Schleife ausgestattet ist, kann die Fahrzeuggeschwindigkeit aus einem Signal $\mathbf{r}(kT)$ mit dem Mittelwert

$$\mu = \frac{1}{\ell} \sum_{i=0}^{\ell-1} \Delta L([k+i]T) \quad (2.5)$$

durch

$$\hat{v}_{Fz} = k \frac{a}{\mu} \quad (2.6)$$

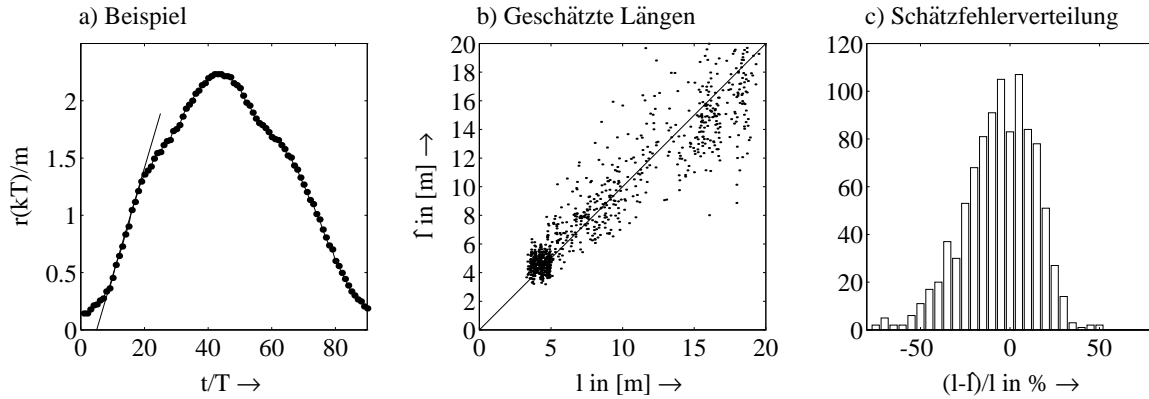


Bild 2.5: Schätzung der Fahrzeuglänge aus einem Fahrzeugsignal

geschätzt werden. k ist hierbei eine von der Abtastrate T abhängige Konstante, a gibt die maximale Steigung des Fahrzeugsignals an. Zur Bestimmung der maximalen Steigung können verschiedene Verfahren herangezogen werden. Die beste Approximation gelingt allerdings mit einer linearen Regression über eine feste Anzahl von Abtastwerten. Nach [81] kann die Steigung eines Signals mit $2p + 1$ Abtastwerten zum Zeitpunkt iT mit

$$a_i = (\mathbf{t}^T \mathbf{t})^{-1} \mathbf{t}^T \mathbf{w}_i \quad (2.7)$$

für einen Zeitvektor

$$\mathbf{t} = [-p, \dots, p]^T \quad (2.8)$$

und einen Signalvektor

$$\mathbf{w}_i = [\Delta L([k + i - p]T), \dots, \Delta L([k + i + p]T)]^T \quad (2.9)$$

bestimmt werden.

Die maximale Tangente wird dann durch

$$a = \max_{i=p}^{\ell-1-p} a_i \quad (2.10)$$

festgelegt. Die Fahrzeuglänge kann wiederum nach Gleichung (2.4) ermittelt werden. Bild 2.5a zeigt ein Beispiel eines Fahrzeugsignals und der dazugehörigen maximalen Tangente. In Bild 2.5b sind Schätzwerte für Fahrzeuglängen gegenüber den aus der Doppelschleifenmessung ermittelten Werten aufgetragen. Bild 2.5c zeigt die aus dieser Stichprobe resultierende Verteilung des Schätzfehlers.

3 Diskrete Signaltransformationen zur Merkmalsextraktion

Diskrete Signaltransformationen oder Reihenentwicklungen besitzen eine zentrale Bedeutung in der Verarbeitung von Signalen, die durch diskrete Werte repräsentiert werden. Im vorliegenden Fall finden sie ihren Einsatz in der Gewinnung geeigneter Merkmale (Merkmalsextraktion) für die in dieser Arbeit untersuchten Schätzverfahren.

In der Mustererkennung ist es wichtig, die zu klassifizierenden Signale durch einige wenige, besonders aussagekräftige Merkmale zu beschreiben. Besonders geeignet sind Verfahren, die Merkmale extrahieren, die die Trennbarkeit unterschiedlicher Klassen optimieren. Da bei der Optimierung der Merkmalsextraktion im allgemeinen nicht das gesamte Erkennungssystem berücksichtigt werden kann, werden geeignete Gütekriterien festgelegt, die eine Beurteilung der ausgewählten Merkmale ermöglichen. Die Merkmalsextraktion wird dann unabhängig von der Signalvorverarbeitung und der Signalerkennung nur unter Betrachtung des Gütekriteriums optimiert.

Bei den Signaltransformationen unterscheidet man zwischen sogenannten signalabhängigen und signalunabhängigen Transformationen. Signalabhängige Transformationen sind an den zugrundeliegenden stochastischen Prozeß angepaßt. Eine besondere Bedeutung kommt hierbei der Karhunen-Loève-Transformation zu, die optimale Eigenschaften bezüglich der Approximation von stochastischen Signalen besitzt. Die diskrete Karhunen-Loève-Transformation und ihr Einsatz zur Merkmalsextraktion wird nach einer Übersicht über diskrete Signalapproximationen mit orthonormalen Reihenentwicklungen zu Beginn dieses Kapitels erläutert. Eine Abwandlung der Karhunen-Loève-Transformation ist die Whitening-Transformation, die anschließend kurz präsentiert wird und zur optimalen Trennung unterschiedlicher Signalklassen zusammen mit der Karhunen-Loève-Transformation in der Diskriminanzanalyse eingesetzt wird.

3.1 Signalapproximation mit orthonormalen Reihenentwicklungen

Mit diskreten Reihenentwicklungen erfolgt eine Transformation eines Signals in einen geeigneten Signalraum. Die Merkmalsextraktion kann als anschließende Projektion des Signals in einen Unterraum verstanden werden.

Betrachtet werde ein n -dimensionaler Hilbert-Raum, in dem das Skalarprodukt zweier beliebiger Vektoren $\mathbf{x} = [x_1, \dots, x_n]^T$ und $\mathbf{y} = [y_1, \dots, y_n]^T$ mit

$$\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i=1}^n x_i y_i = \mathbf{x}^T \mathbf{y} \quad (3.1)$$

und die hierdurch implizierte Norm mit

$$\|\mathbf{x}\| = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle} \quad (3.2)$$

definiert sind. Der Abstand zweier Vektoren berechnet sich durch die aus der Norm abgeleitete Metrik

$$d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\| = \sqrt{\langle \mathbf{x} - \mathbf{y}, \mathbf{x} - \mathbf{y} \rangle}. \quad (3.3)$$

Die lineare Transformation eines Signals läßt sich mit Hilfe einer orthonormalen Basis $\{\mathbf{u}_1, \dots, \mathbf{u}_n\}$ und eines Repräsentanten $\mathbf{y} = [y_1, \dots, y_n]^T$ als

$$\mathbf{x} = \sum_{i=1}^n y_i \mathbf{u}_i \quad (3.4)$$

schreiben, wobei für die Koeffizienten

$$y_i = \mathbf{u}_i^T \mathbf{x} \quad (3.5)$$

gilt. Mit der orthogonalen Matrix $\mathbf{U} = [\mathbf{u}_1, \dots, \mathbf{u}_n]$ können die Gleichungen (3.4) und (3.5) kompakt als

$$\mathbf{x} = \mathbf{U} \mathbf{y} \quad \text{und} \quad \mathbf{y} = \mathbf{U}^T \mathbf{x} \quad (3.6)$$

geschrieben werden.

Für die orthonormalen Basisvektoren \mathbf{u}_i gilt:

$$\mathbf{u}_i^T \mathbf{u}_j = \delta_{ij}, \quad i, j = 1, \dots, n. \quad (3.7)$$

Hierbei ist δ_{ij} das *Kronecker-Symbol*

$$\delta_{ij} = \begin{cases} 1 & \text{für } i = j, \\ 0 & \text{sonst.} \end{cases} \quad (3.8)$$

Zur Merkmalsreduktion wird eine Annäherung des Signals \mathbf{x} mit einer begrenzten Anzahl von $m < n$ Basisvektoren gewünscht. Die Annäherung ergibt einen Schätzvektor

$$\hat{\mathbf{x}} = \sum_{i=1}^m y_i \mathbf{u}_i, \quad (3.9)$$

der resultierende Schätzfehler ergibt sich bei einem orthonormalen Basissystem zu

$$\mathbf{e} = \mathbf{x} - \hat{\mathbf{x}} = \sum_{i=m+1}^n y_i \mathbf{u}_i. \quad (3.10)$$

Gesucht wird die Matrix \mathbf{U} , die eine optimale Projektion in den m -dimensionalen Unterraum durch Weglassen der $n - m$ letzten Basisvektoren ermöglicht. Die Optimalität muß hierzu durch ein geeignetes Gütemaß festgelegt werden. Sinnvoll sind hierbei Gütemaße, die die für die Signalerkennung relevante Information aus einem Signal extrahieren können. Bei der im folgenden dargestellten diskreten Karhunen-Loève-Transformation wird die Approximation eines Signals bewertet, während bei der später diskutierten Diskriminanzanalyse eine Bewertung der Konzentration der Merkmale einer Klasse gegenüber Merkmalen anderer Klassen stattfindet.

3.2 Diskrete Karhunen-Loève-Transformation

Bei der diskreten Karhunen-Loève-Transformation erfolgt die Wahl der Basisvektoren $\{\mathbf{u}_i\}$, $i = 1, \dots, m$, derart, daß der mittlere quadratische

Fehler

$$\varepsilon = E\{\|\mathbf{e}\|^2\} = E\{\|\mathbf{x} - \hat{\mathbf{x}}\|^2\} \quad (3.11)$$

minimiert wird. Nach [31] ergibt sich für mittelwertfreie Zufallsprozesse \mathbf{x} als Lösung dieses Minimierungsproblems die Eigenwertgleichung

$$\mathbf{R}_{xx}\mathbf{u}_i = \lambda_i\mathbf{u}_i, \quad i = m + 1, \dots, n. \quad (3.12)$$

Gleichung (3.12) besagt, daß die gesuchten Basisvektoren \mathbf{u}_i die *Eigenvektoren* der Kovarianzmatrix $\mathbf{R}_{xx} = E\{\mathbf{x}\mathbf{x}^T\}$ sind. Unter dieser Bedingung kann (3.11) als

$$\varepsilon = \sum_{i=m+1}^n \lambda_i \quad (3.13)$$

geschrieben werden [31], und die auszuwählenden Basisvektoren müssen diejenigen Eigenvektoren der Kovarianzmatrix \mathbf{R}_{xx} sein, die zu den m größten Eigenwerten $\lambda_1, \dots, \lambda_m$ mit $\lambda_1 \geq \dots \geq \lambda_m$ gehören.

Die diskrete Karhunen-Loève-Transformation besitzt eine Reihe weiterer wichtiger Eigenschaften, die ebensogut zum Ansatzpunkt für die Optimierung eingesetzt werden können. So kann die Karhunen-Loève-Transformation auch als dekorrelierende Lineartransformation aufgefaßt werden. Die Repräsentanten und damit die transformierten Merkmale y_i sind unkorreliert. Die Kovarianzmatrix $\mathbf{R}_{yy} = E\{\mathbf{y}\mathbf{y}^T\}$ ist eine Diagonalmatrix und lautet

$$\mathbf{R}_{yy} = \mathbf{U}^T \mathbf{R}_{xx} \mathbf{U} = \mathbf{\Lambda} = \begin{bmatrix} \lambda_1 & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \lambda_n \end{bmatrix}. \quad (3.14)$$

Nach der Transformation sind die transformierten Merkmale somit nicht mehr durch Linearkombination auseinander vorhersagbar. Die vor der Transformation in dieser gegenseitigen Vorhersagbarkeit durch Linearkombination liegende Redundanz wird durch die Karhunen-Loève-Transformation entfernt. Die Redundanzbefreiung äußert sich in einer Umverteilung unter den Werten, die die Varianzen der einzelnen Merkmale vor und nach der Transformation annehmen. Es ergibt sich, daß

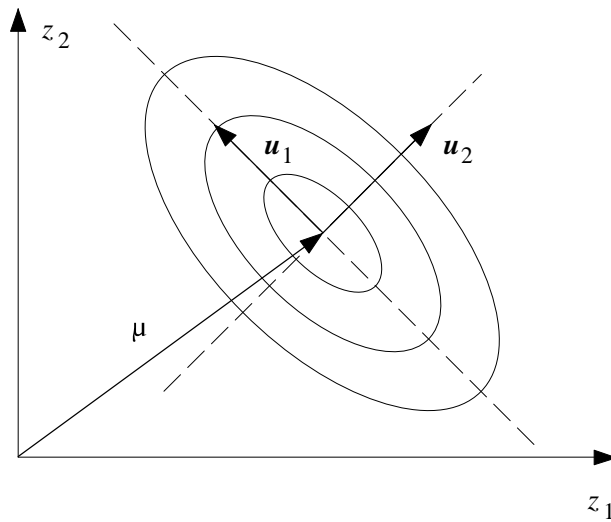


Bild 3.1: Geometrische Deutung der Karhunen-Loève-Transformation

die Karhunen-Loève-Transformation unter allen möglichen Lineartransformationen die am stärksten ungleichförmige Werteverteilung der Varianzen im Sinne minimaler Entropie erzeugt. Zudem hat die Karhunen-Loève-Transformation die Eigenschaft, daß die Gesamtvarianz des durch die Transformation erzeugten Prozesses maximal wird [72].

Bild 3.1 verdeutlicht die Eigenschaften der Karhunen-Loève-Transformation anhand einer Höhenliniendarstellung der Verteilungsdichte eines gaußschen Prozesses $\mathbf{x} = \mathbf{z} - \boldsymbol{\mu}$. Der Eigenvektor \mathbf{u}_1 , der zum größten Eigenwert gehört, zeigt in die Richtung der größten Abweichung vom Schwerpunkt $\boldsymbol{\mu}$.

3.3 Whitening-Transformation

Eine Abwandlung der Karhunen-Loève-Transformation ist die Whitening-Transformation. Die Forderung der Karhunen-Loève-Transformation nach unkorrelierten Koeffizienten der Kovarianzmatrix des Repräsentanten wird bei der Whitening-Transformation dahingehend erweitert, daß alle Koeffizienten die gleiche Varianz besitzen. Die Whitening-Transformation

überführt damit einen farbigen Prozeß mit der Kovarianzmatrix $\mathbf{R}_{xx} \neq \mathbf{I}$ in einen weißen Prozeß mit der Kovarianzmatrix

$$\mathbf{R}_{yy} = E\{\mathbf{y}\mathbf{y}^T\} = E\{\mathbf{W}^T \mathbf{x}\mathbf{x}^T \mathbf{W}\} = \mathbf{W}^T \mathbf{R}_{xx} \mathbf{W} = \mathbf{I}. \quad (3.15)$$

Durch die Karhunen-Loève-Transformation läßt sich die Kovarianzmatrix eines Prozesses \mathbf{x} als

$$\mathbf{R}_{xx} = \mathbf{U} \mathbf{\Lambda} \mathbf{U}^T \quad (3.16)$$

schreiben. Die Matrix \mathbf{U} setzt sich aus den Eigenvektoren und die Matrix $\mathbf{\Lambda}$ aus den Eigenwerten der Kovarianzmatrix \mathbf{R}_{xx} zusammen (3.14).

Mit der Matrix

$$\mathbf{\Sigma} = \mathbf{\Lambda}^{-1/2} = \begin{bmatrix} \sqrt{1/\lambda_1} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \sqrt{1/\lambda_n} \end{bmatrix}. \quad (3.17)$$

läßt sich die Whitening-Transformation als

$$\mathbf{W} = \mathbf{U} \mathbf{\Lambda}^{-1/2} = \mathbf{U} \mathbf{\Sigma} \quad (3.18)$$

angeben, was sich durch Einsetzen in Gleichung (3.15) zeigen läßt.

Alternativ ist es auch möglich, die Transformationsmatrix \mathbf{W} mit Hilfe der *Cholesky-Zerlegung* der Kovarianzmatrix \mathbf{R}_{xx} zu bestimmen [54].

3.4 Merkmalsextraktion mit der Karhunen-Loève-Transformation

Die diskrete Karhunen-Loève-Transformation ist die Entwicklung von Zufallsprozessen nach einem orthonormalen Vektorsystem mit der bestmöglichen Approximation für das Gütekriterium nach Gleichung (3.11). Allerdings bedeutet die bestmögliche Approximation eines Signals nicht unbedingt bestmögliche Klassifikation. Der Nachteil der diskreten Karhunen-Loève-Transformation liegt darin, daß sie eine Klassentrennung im

Merkmalsraum unberücksichtigt läßt. Dieses berücksichtigt die generalisierte Karhunen-Loève-Transformation, in der klassenspezifische Erwartungswerte miteinbezogen werden.

Betrachtet werden stochastische Prozesse \mathbf{z}_k , $k = 1, \dots, K$, die verschiedenen Klassen eines Signalerkennungssystems entstammen. Die Realisationen dieser Prozesse werden mit $\mathbf{z}_{i,k}$ bezeichnet, wobei $i = 1, \dots, N_k$ die Nummer der Realisation und N_k die Anzahl der Realisationen der Klasse k angibt.

Der erste Schritt zur Merkmalsextraktion ist die Erzeugung mittelwertfreier Prozesse

$$\mathbf{x}_k = \mathbf{z}_k - \boldsymbol{\mu}_k. \quad (3.19)$$

Die klassenspezifischen Mittelwerte können mit

$$\boldsymbol{\mu}_k = E\{\mathbf{z}_k\} \approx \frac{1}{N_k} \sum_{i=1}^{N_k} \mathbf{z}_{i,k} \quad (3.20)$$

abgeschätzt werden.

Zur Merkmalsextraktion können zwei verschiedene Transformationsarten verwendet werden.

3.4.1 Klassenspezifische Transformation

Eine Möglichkeit in der Merkmalsextraktion liegt darin, klassenspezifische Transformationen durchzuführen. In diesem Fall wird für jede Klasse eine Kovarianzmatrix

$$\mathbf{R}_{xx_k} = E\{\mathbf{x}_k \mathbf{x}_k^T\} \approx \frac{1}{N_k + 1} \sum_{i=1}^{N_k} \mathbf{x}_{i,k} \mathbf{x}_{i,k}^T \quad (3.21)$$

ermittelt. Die Karhunen-Loève-Transformation wird auf jede Kovarianzmatrix angewandt und es entstehen K verschiedene Basissysteme \mathbf{U}_k . Nach der Transformation liegen die Muster verschiedener Klassen somit in verschiedenen Merkmalsräumen.

Bei der Signalerkennung erfolgt eine Abbildung des unbekanntes Signals in jeden der K Merkmalsräume und die Zuordnung zu einer der

möglichen Klassen erfolgt z.B. durch Minimierung des Abstands des unbekanntes Musters zu den entsprechenden Klassenmittelpunkten (Kapitel 4). Eine ungleichmäßige Verteilung der Muster in den verschiedenen Klassen muß in diesem Fall allerdings berücksichtigt werden. Hierzu erfolgt nach der Transformation eine Skalierung der Achsen auf gleiche Varianz durch die Whitening-Transformation. Damit ergibt sich der transformierte Vektor mit (3.18) zu

$$\mathbf{y}_k = \mathbf{\Sigma} \mathbf{U}_k^T \mathbf{x} = \mathbf{W}_k^T \mathbf{x}. \quad (3.22)$$

Aufgrund der Skalierung ist die Transformationsmatrix \mathbf{W}_k nicht mehr orthonormal.

3.4.2 Gemeinsame Transformation

Für die praktische Anwendung ist es oftmals günstiger, unter Berücksichtigung der A-priori-Wahrscheinlichkeiten der Klassen p_k einen Prozeß

$$\mathbf{x} = \sum_{k=1}^K p_k \mathbf{x}_k \quad (3.23)$$

zu betrachten und eine gemeinsame Transformation und somit auch einen gemeinsamen Merkmalsraum zu verwenden.

Die Kovarianzmatrix wird aus den klassenspezifischen Kovarianzmatrizen gebildet,

$$\mathbf{R}_{xx} = E\{\mathbf{x} \mathbf{x}^T\} = \sum_{k=1}^K p_k \mathbf{R}_{xx_k}. \quad (3.24)$$

Die Karhunen-Loève-Transformation wird in diesem Fall nur einmal durchgeführt, das entstehende Basissystem $\mathbf{U} = [\mathbf{u}_1, \dots, \mathbf{u}_n]$ beinhaltet die geordneten Eigenvektoren der Kovarianzmatrix \mathbf{R}_{xx} und für den Vektor im transformierten Raum gilt

$$\mathbf{y} = \mathbf{U}^T \mathbf{x}. \quad (3.25)$$

3.4.3 Anwendung der Karhunen-Loève-Transformation auf Fahrzeugsignale

Die Approximationseigenschaften der Karhunen-Loève-Transformation sollen im folgenden mit normierten Fahrzeugsignalen dargestellt werden. Jede Fahrzeugklasse liefert einen stochastischen Prozeß, dessen Realisationen einer Lernstichprobe entnommen sind. Die klassenspezifischen Mittelwerte und Kovarianzmatrizen können mit (3.20) und (3.21) abgeschätzt werden. Die A-priori-Wahrscheinlichkeiten der Klassen sind allerdings nicht bekannt. Sie können entweder als gleich angenommen werden oder aus der Anzahl der klassenspezifischen Realisationen ebenfalls abgeschätzt werden:

$$p_k \approx \frac{N_k}{N}. \quad (3.26)$$

Hierbei ist N die Summe aller Realisationen,

$$N = \sum_{k=1}^K N_k. \quad (3.27)$$

Die Approximation der Fahrzeugsignale erfolgt dann durch

$$\hat{\mathbf{z}}_{i,k} = \mathbf{U} \mathbf{U}^T (\mathbf{z}_{i,k} - \boldsymbol{\mu}_k) + \boldsymbol{\mu}_k. \quad (3.28)$$

Die Matrix \mathbf{U} beinhaltet die m Eigenvektoren, die zu den m größten Eigenwerten der geschätzten Kovarianzmatrix gehören. Bild 3.2 zeigt die aus einer Lernstichprobe ermittelten Elemente der Kovarianzmatrizen \mathbf{R}_{xx} und $\mathbf{\Lambda}$. Deutlich erkennbar ist die dekorrelierende Wirkung der Karhunen-Loève-Transformation. Gleichzeitig zeigt sich, daß der größte Teil der Information in sehr wenigen Koeffizienten von $\mathbf{\Lambda}$ komprimiert ist, so daß eine Approximation mit sehr wenigen Koeffizienten sehr gute Ergebnisse liefert. Die zu den vier größten Eigenwerten gehörenden Eigenvektoren wurden zur Approximation von Fahrzeugsignalen der Dimension $n = 32$ verwendet. Bild 3.3 zeigt beispielhafte Signalverläufe verschiedener Fahrzeugklassen und deren Approximation für normierte Signale, die mit 2,5m-Schleifen erfaßt wurden.

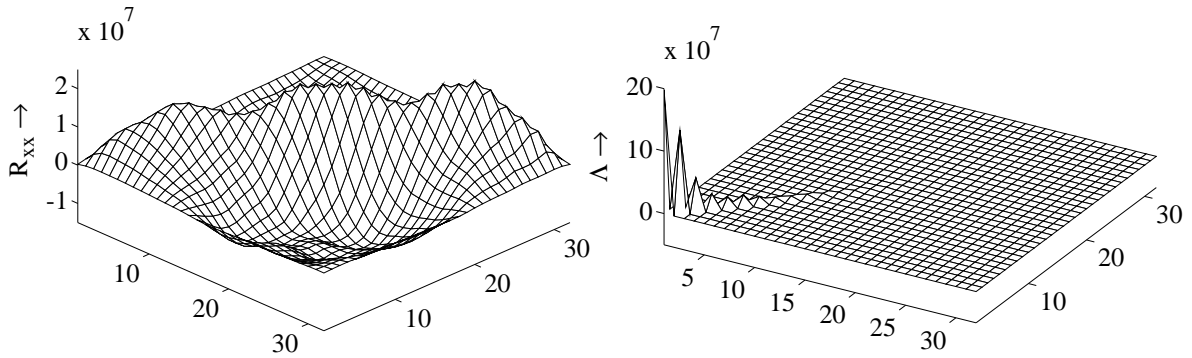


Bild 3.2: Kovarianzmatrix R_{xx} des Prozesses x und Kovarianzmatrix Λ des Repräsentanten y

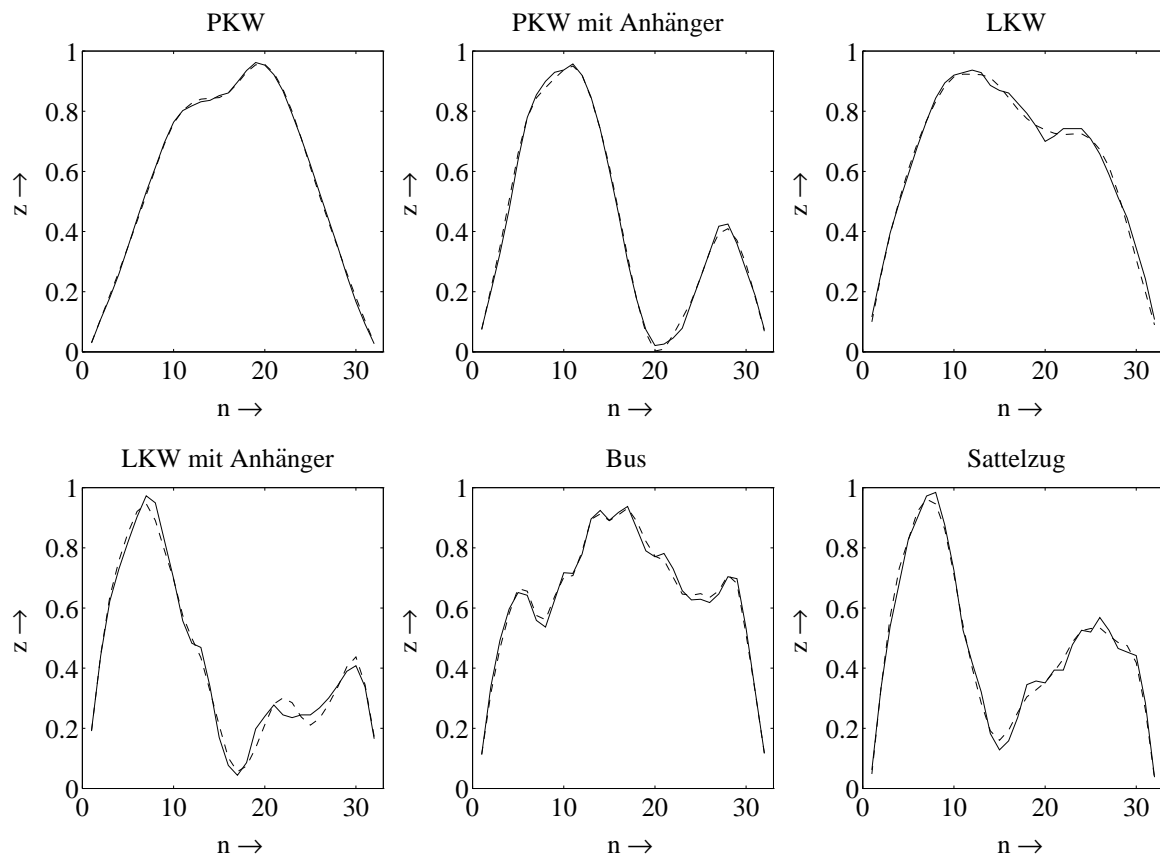


Bild 3.3: Beispielhafte Signalverläufe verschiedener Fahrzeugklassen (—) und deren Approximation (- -) (2,5m-Schleife)

3.5 Diskriminanzanalyse

Neben dem Kriterium der bestmöglichen Approximation von Zufallsprozessen nach einem orthogonalen Vektorsystem finden sich in der Praxis unter dem Begriff Diskriminanzanalyse noch weitere Kriterien zur Merkmalsextraktion [30], [31], [37]. In der Diskriminanzanalyse wird eine optimale Klassentrennung von Merkmalsvektoren dadurch angestrebt, daß der Abstand der Muster einer Klasse (*Intraklassenabstand*) minimiert wird und der Abstand der Muster verschiedener Klassen (*Interklassenabstand*) gleichzeitig maximiert wird.

Hierzu werden zur Formulierung der Kriterien sogenannte Scattermatrizen eingeführt. Die Intraklassen-Scattermatrix (*within-class scatter matrix*) gibt die Streuung der Mustervektoren um ihren Mittelwertsvektor an:

$$\begin{aligned} \mathbf{S}_W &= \sum_{k=1}^K p_k E\{(\mathbf{z}_k - \boldsymbol{\mu}_k)(\mathbf{z}_k - \boldsymbol{\mu}_k)^T\} \\ &= \sum_{k=1}^K p_k E\{\mathbf{x}_k \mathbf{x}_k^T\} = \sum_{k=1}^K p_k \mathbf{R}_{\mathbf{x}\mathbf{x}_k}. \end{aligned} \quad (3.29)$$

Diese Matrix ist mit der Kovarianzmatrix aus Gleichung (3.24) identisch.

Die Interklassen-Scattermatrix (*between-class scatter matrix*) berücksichtigt die Streuung der klassenspezifischen Mittelwertsvektoren um den gemeinsamen Schwerpunkt aller Klassen:

$$\mathbf{S}_B = \sum_{k=1}^K p_k (\boldsymbol{\mu}_k - \boldsymbol{\mu})(\boldsymbol{\mu}_k - \boldsymbol{\mu})^T, \quad (3.30)$$

wobei der gemeinsame Mittelwert mit

$$\boldsymbol{\mu} = E\{\mathbf{z}\} = \sum_{k=1}^K p_k \boldsymbol{\mu}_k \quad (3.31)$$

gegeben ist. Hieraus ergibt sich, daß der Rang der Matrix \mathbf{S}_B höchstens $K - 1$ sein kann.

Die gemeinsame Scattermatrix (*mixture scatter matrix*) ist die Kovarianzmatrix aller Mustervektoren unabhängig von ihrer Klasse:

$$\mathbf{S}_M = E\{(\mathbf{z} - \boldsymbol{\mu})(\mathbf{z} - \boldsymbol{\mu})^T\} = \mathbf{S}_W + \mathbf{S}_B. \quad (3.32)$$

Die lineare Transformation eines n -dimensionalen Prozesses \mathbf{x} in einen m -dimensionalen Prozeß \mathbf{y} ($m < n$) erfolgt mit

$$\mathbf{y} = \mathbf{W}^T \mathbf{x}. \quad (3.33)$$

Die Spaltenvektoren der $n \times m$ -Matrix \mathbf{W} müssen linear unabhängig aber nicht orthonormal sein. Da alle Scattermatrizen die Form einer Kovarianzmatrix besitzen, können die entsprechenden Scattermatrizen \mathbf{S}_w und \mathbf{S}_b des Prozesses \mathbf{y} aus den Scattermatrizen des Prozesses \mathbf{x} abgeleitet werden [31]:

$$\mathbf{S}_w = \mathbf{W}^T \mathbf{S}_W \mathbf{W}, \quad (3.34)$$

$$\mathbf{S}_b = \mathbf{W}^T \mathbf{S}_B \mathbf{W}. \quad (3.35)$$

Die Matrix \mathbf{W} muß jetzt so gewählt werden, daß sie das Verhältnis vom Interklassenabstand zum Intraklassenabstand optimiert. Als Maß dieses Verhältnisses finden sich in der Literatur verschiedene Kriterien.

Ein mögliches Kriterium liegt darin, daß die Klassenzentren im transformierten Merkmalsraum möglichst weit auseinander liegen. Bei der Maximierung des Interklassenabstandes dürfen allerdings die Intraklassenabstände nicht verändert werden. Dieses gelingt dadurch, daß die Intraklassen-Scattermatrix mit einer Whitening-Transformation in einen weißen Prozeß überführt wird. In diesem Fall ist die Matrix invariant gegen jede orthonormale Transformation. Das Kriterium läßt sich folgendermaßen definieren:

$$J(\mathbf{W}) = \text{spur} \{ [\mathbf{W}^T \mathbf{S}_W \mathbf{W}]^{-1} [\mathbf{W}^T \mathbf{S}_B \mathbf{W}] \} = \text{spur} \{ \mathbf{S}_b \} \stackrel{!}{=} \max \quad (3.36)$$

unter der Nebenbedingung

$$\mathbf{S}_w = \mathbf{W}^T \mathbf{S}_W \mathbf{W} = \mathbf{I}. \quad (3.37)$$

Dieses sogenannte Fisher-Kriterium (*Fisher discriminant ratio*) läßt sich für die Auswahl der Achsen im transformierten Raum einsetzen.

Aus der Whitening-Transformation ergibt sich

$$\mathbf{W} = \mathbf{U} \mathbf{\Lambda}_W^{-1/2}. \quad (3.38)$$

Die Matrix \mathbf{U} enthält die Eigenvektoren und $\mathbf{\Lambda}_W$ ist die Diagonalmatrix der Eigenwerte der Scattermatrix \mathbf{S}_W . Eine optimale Transformation bezüglich der Klassenmittelpunkte kann durch Lösung der Eigenwertaufgabe für die Matrix \mathbf{S}_b gefunden werden:

$$\mathbf{S}_b \mathbf{V} = \mathbf{V} \mathbf{\Lambda}_b. \quad (3.39)$$

Die relevante Information der Klassenmittelpunkte wird durch diese Transformation in den $K - 1$ Eigenvektoren $\mathbf{V} = [\mathbf{v}_1, \dots, \mathbf{v}_{K-1}]$ komprimiert. Die Matrix $\mathbf{\Lambda}_b$ ist wiederum die Diagonalmatrix der Eigenwerte der Matrix \mathbf{S}_b .

Die gesamte Transformation läßt sich als

$$\mathbf{A} = \mathbf{W} \mathbf{V} = \mathbf{U} \mathbf{\Lambda}_W^{-1/2} \mathbf{V} \quad (3.40)$$

zusammenfassen. Ein Datenvektor \mathbf{z} wird dann durch

$$\mathbf{y} = \mathbf{A}^T (\mathbf{z} - \boldsymbol{\mu}) \quad (3.41)$$

in das neue Koordinatensystem überführt. Man kann zeigen, daß die Matrix \mathbf{A} aus den Eigenvektoren der Produktmatrix $\mathbf{S}_W^{-1} \mathbf{S}_B$ besteht [30].

Alternativ kann die Merkmalsextraktion ebenso bezüglich des mittleren quadratischen Abstands aller Vektoren optimiert werden oder das Verhältnis des gesamten Interklassenabstands zum gesamten Interklassenabstand kann maximiert werden [30], [31]. Interessanterweise führen auch diese Kriterien auf die gleiche Lösung wie das Kriterium (3.36).

Die Qualität der durch diese optimalen Transformationen ausgewählten Merkmale hängt davon ab, wie genau die genannten Kriterien eine Trennbarkeit der Klassen angeben können. Die Kriterien arbeiten sehr gut, wenn die Klassenverteilungsdichten unimodal sind und durch die

Streuung der Mittelwertsvektoren gut getrennt sind. Sind die Mittelwertsvektoren ähnlich und die Verteilungen multimodal, führt eine Clusteranalyse und eine nachträgliche Aufteilung der Klassen in mehrere Unterklassen noch zu guten Ergebnissen. Im Falle von unimodalen Verteilungen mit ähnlichen Mittelwertsvektoren führt die Diskriminanzanalyse zu unbefriedigenden Ergebnissen.

Da die Verteilungsdichte der Merkmale unbekannt ist, kann von vornherein keine Aussage darüber gemacht werden, welche Art der Merkmalsextraktion sinnvoll ist. Vielmehr müssen die möglichen Transformationen zusammen mit den möglichen Klassifikatoren für ein vorliegendes Erkennungsproblem getestet und bewertet werden.

4 Fahrzeugklassifikation

Die automatische Klassifikation von Fahrzeugen im Straßenverkehr wird seit zwanzig Jahren in Verkehrsleitsystemen zur statistischen Auswertung des Fahrzeugverkehrs durchgeführt. Waren die ersten Ansätze aufgrund der geringen Rechenleistung in den siebziger Jahren noch relativ einfach [57], [28], [68], so wurden Mitte der achtziger Jahre schon aufwendigere Verfahren der Merkmalsextraktion zur Klassifikation eingesetzt, z.B. [11]. Grundlage der Fahrzeugklassifikation sind vorwiegend Induktionsschleifensignale, in den letzten Jahren wurden allerdings auch Verfahren mit Mikrowellensignalen und mit digitalen Bildsignalen untersucht und implementiert, z.B. [70], [86], [60].

Die Klassifikation von Fahrzeugen ist ein typisches Beispiel eines eindimensionalen Mustererkennungsproblems. Anhand der erfaßten Fahrzeugmuster wird ein detektiertes Fahrzeug einer vorgegebenen Klasse zugeordnet. Hierbei eignen sich die Ausgangssignale von Induktionsschleifen zur Zuordnung der Fahrzeuge. Das Problem der Klassifikation liegt neben der Merkmalsextraktion vorwiegend in der Festlegung des Klassifikators. Bei den Klassifikatoren unterscheidet man zwischen *parametrischen* und *nichtparametrischen* Verfahren. Zu letzteren gehören auch die *neuronalen Netzwerke*, die in den letzten zehn Jahren durch den Einsatz von Parallelrechnern Einzug in zahlreiche technische Applikationen erhalten haben. Im folgenden soll die Effizienz verschiedener Fahrzeugklassifikatoren untersucht und miteinander verglichen werden.

Zu Beginn des Kapitels wird die Struktur eines Mustererkennungs-systems zur Fahrzeugklassifikation dargestellt. Die Vorverarbeitung und Merkmalsextraktion der Fahrzeugsignale zur Klassifikation wird diskutiert und anschließend werden die theoretischen Grundlagen der Mustererkennung sowie der parametrischen und nichtparametrischen Verfahren aufgeführt. Ein Vergleich und eine Schätzung des minimalen Klassifikationsfehlers bilden den Abschluß dieses Kapitels.

4.1 Mustererkennungssystem

Die Fahrzeugklassifikation gelingt mit Hilfe der Theorie der Mustererkennung. Mustererkennungssysteme sind technische Einrichtungen, die unbekannte Objekte oder Muster mit Hilfe von Äquivalenzregeln in vorgegebene Objektklassen einordnen. Bild 4.1 gibt einen Überblick über die Struktur eines Mustererkennungssystems zur Fahrzeugklassifikation. Nach der Signalerfassung und Vorverarbeitung müssen aus dem normierten Fahrzeugsignal geeignete Merkmale extrahiert werden. Diese werden zum einen durch die in Kapitel 3 beschriebenen Verfahren aus dem Fahrzeugsignal gewonnen und zum anderen aus den Überfahrkurven als Sondermerkmale selektiert. Nach der Merkmalsextraktion sollten klassengleiche Signale möglichst ähnlich sein und die Klassenzentren möglichst weit auseinander liegen. Anschließend müssen für jede Klasse Entscheidungsfunktionen bestimmt werden, aus denen dann durch Auswahl der maximalen Funktion die Fahrzeugklasse bestimmt werden kann. Die Bestimmung der Entscheidungsfunktionen und die Merkmalsextraktion müssen in der Regel kombiniert betrachtet werden.

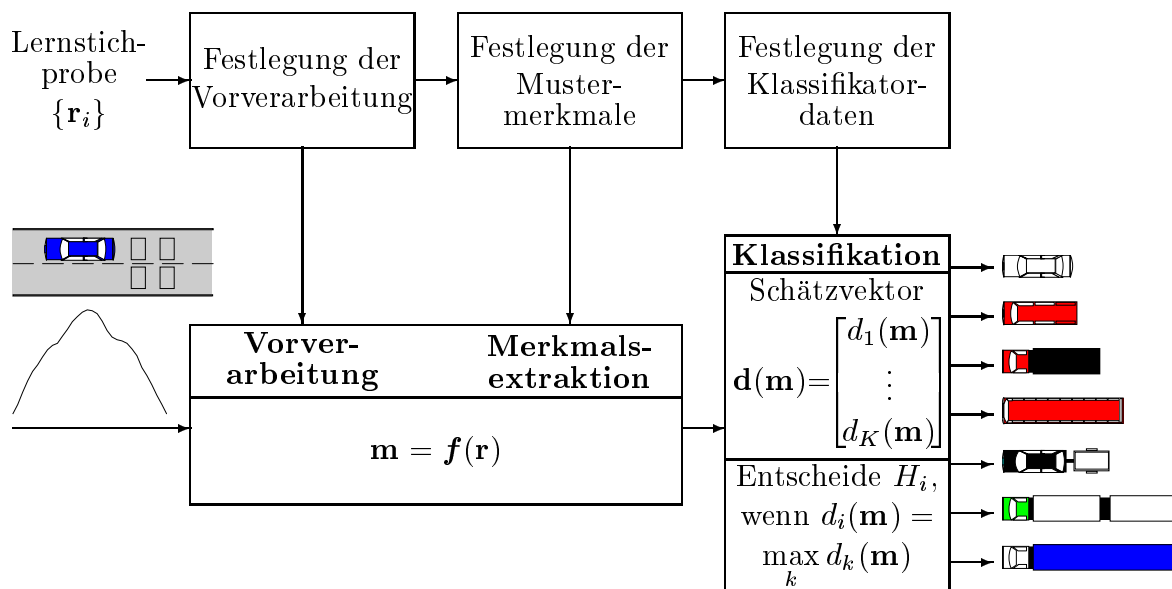









Bild 4.1: Struktur eines Mustererkennungssystems zur Fahrzeugklassifikation

Tabelle 4.1: Mögliche Einteilung der Fahrzeugklassen

k_1	PKW	
k_2	Kleintransporter	
k_3	LKW	
k_4	Bus	
k_5	PKW mit Anhänger	
k_6	LKW mit Anhänger	
k_7	Sattelzug	
k_8	Rückweisung	

Eine mögliche Einteilung der Fahrzeugklassen findet sich in Tabelle 4.1. Die Aufteilung der Fahrzeugklassen richtet sich nach der Anwendung und somit kann es sein, daß z.B. für statistische Auswertungen mehrere Fahrzeugklassen benötigt werden, für andere Anwendungen hingegen eine einfachere Klasseneinteilung ausreicht. Da in einem Mustererkennungssystem nicht eindeutig festgelegt werden kann, ob ein gemessener Merkmalsvektor überhaupt einer der vorgegebenen Klassen entstammt, wird häufig noch eine zusätzliche Rückweisungsklasse definiert.

4.2 Signalvorverarbeitung und Merkmalsextraktion

Bild 4.2 gibt einen Überblick über die Methoden und die Signalvektoren der Vorverarbeitung und Merkmalsextraktion zur Fahrzeugklassifikation.

Die Eingangssignale des Mustererkennungssystems sind die mit den Induktionsschleifen erfaßten Abtastwerte eines Fahrzeugsignals

$$\mathbf{r}(kT) = [\Delta L(kT), \Delta L([k + 1]T), \dots, \Delta L([k + \ell - 1]T)]^T. \quad (4.1)$$

Jedes Fahrzeugsignal stellt eine Realisation eines stochastischen Prozesses \mathbf{r} dar. Zur Festlegung des Klassifikators und der Transformationsmatrizen der Merkmalsextraktion erfolgt eine Analyse einer Lernstichprobe bestehend aus den Realisationen \mathbf{r}_i , $i = 1, \dots, L$. L gibt den Umfang der Lernstichprobe an.

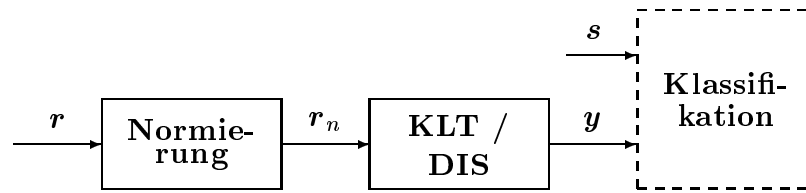


Bild 4.2: Signalvorverarbeitung und Merkmalsextraktion zur Fahrzeugklassifikation

Nach der Signalnormierung erhält man nach Abschnitt 2.2 die Signale

$$\mathbf{r}_n(kT_n) = [r_n(kT_n), r_n([k + 1]T_n), \dots, r_n([k + n - 1]T_n)]^T. \quad (4.2)$$

Die anschließende Merkmalsextraktion kann mit der Karhunen-Loève-Transformation oder mit einer Diskriminanzanalyse durchgeführt werden und der neue Signalvektor \mathbf{y} ergibt sich aus Gleichung (3.25) bzw. (3.41).

Die zur Klassifikation eingesetzten Merkmalsvektoren setzen sich aus den Realisationen der Prozesse \mathbf{y} und \mathbf{s} zusammen. Der Prozeß \mathbf{s} beinhaltet die aus den Überfahrkurven selektierten Sondermerkmale, auf die später noch eingegangen wird. Dem Klassifikator werden somit die Realisationen des Prozesses $\mathbf{m} = [\mathbf{y}^T, \mathbf{s}^T]^T$ zugeführt.

4.3 Klassifikation

Die Vorgehensweisen in der Mustererkennung sind ausführlich in [30], [31], [36], [37], [72], [77], [78], [85] beschrieben. In den nächsten Abschnitten werden lediglich die für diese Arbeit eingesetzten Klassifikationsalgorithmen erläutert.

Die Aufgabe der Klassifikation besteht darin, einen gemessenen Merkmalsvektor \mathbf{m} einer bestimmten Klasse k_i von K möglichen Klassen zuzuordnen. Der Merkmalsvektor muß hierfür die zur Klassifikation notwendige Information anhand geeigneter Merkmale enthalten.

Am Eingang des Klassifikators steht ein gestörtes Signal \mathbf{m} zur Verfügung, für welches eine Entscheidung für eine der Hypothesen

$$\begin{array}{lll} H_1 & : & k_1 \text{ eingetreten} \\ \vdots & & \vdots \\ H_K & : & k_K \text{ eingetreten} \end{array}$$

herbeigeführt werden muß. Hierbei können Kosten für richtige und falsche Entscheidungen aufgestellt werden, und die Entscheidung fällt zugunsten der Hypothese, die zu den geringsten Kosten geführt hat (Bayes-Klassifikation). Eine Entscheidung für diejenige Hypothese, die mit maximaler Wahrscheinlichkeit richtig ist, ist ebenfalls möglich (Maximum-a-posteriori-Klassifikation).

In Bild 4.3 wird die Problematik der Entscheidungsfindung verdeutlicht. Der Merkmalsraum wird in K Raumgebiete R_k aufgeteilt. Die Zuordnung des Klassifikators läßt sich als Abbildung des N -dimensionalen Merkmalsraums in einen K -dimensionalen Entscheidungsraum auffassen. Liegt der Merkmalsvektor \mathbf{m} im Gebiet R_k , fällt die Entscheidung für die Hypothese H_k . Liegt der Vektor auf der Trennfläche zwischen zwei Raumgebieten, fällt die Entscheidung entsprechend den A-priori-Wahrscheinlichkeiten der Ereignisse.

Bei statistischen Mustererkennungsverfahren wird die Klassifikation eines Musters auf sogenannte Diskriminanz- oder Entscheidungsfunktionen zurückgeführt. Für jede Klasse $k = 1, \dots, K$ wird eine Entscheidungsfunktion $d_k(\mathbf{m})$ bestimmt. Die Zuordnung des Merkmalsvektors zu einer der K Klassen geschieht dann z.B. durch Auswahl der größten Entscheidungsfunktion. Die Festlegung der Entscheidungsfunktionen beinhaltet die Hauptaufgabe bei einer Klassifikatorentwicklung.

4.3.1 Bayes-Klassifikation

In Mustererkennungssystemen kann keine absolute Fehlerfreiheit verlangt werden. Vielmehr müssen die Entscheidungsregeln eines Klassifikators so festgelegt werden, daß die Fehlerrate des Systems möglichst gering gehalten wird. Beim Bayes-Klassifikator geht man zur Festlegung der optimalen Entscheidungsfunktion davon aus, daß jede Entscheidungsfindung mit

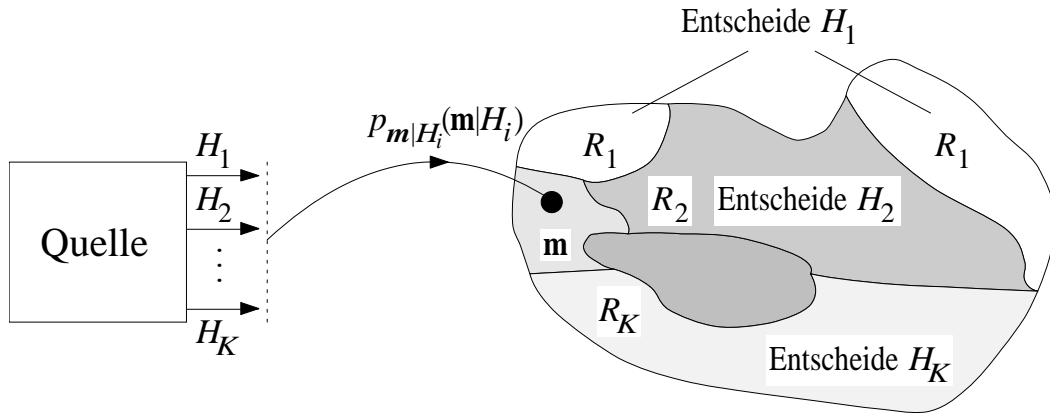


Bild 4.3: Entscheidungsbereiche für zweidimensionale Empfangsvektoren

einem *Risiko* verbunden ist, welches nach [79] mit

$$\mathcal{R} = \sum_{i=1}^K \sum_{k=1}^K C_{ik} P(H_i, k_k) \quad (4.3)$$

festgelegt werden kann. Die Größen C_{ik} können als Kosten verstanden werden, die durch ein Verbundereignis $\{H_i, k_k\}$ mit der Wahrscheinlichkeit $P(H_i, k_k)$ verursacht werden. Im folgenden wird davon ausgegangen, daß die Verbundereignisse $\{H_i, k_k\}$ für $i = k$ eine richtige Entscheidung und für $i \neq k$ eine falsche Entscheidung bedeuten und die Kosten C_{ik} für richtige Entscheidungen kleiner sind als für falsche Entscheidungen. Das Ziel bei der Klassifikatorentwicklung besteht nun darin, das mit dem Klassifikationsproblem verbundene Risiko zu minimieren. Bei vorgegebenem Prozeß und festgelegten Kosten ist das Risiko nur durch die Wahl der Raumgebiete bestimmt, so daß eine Minimierung des Risikos nur durch eine Optimierung der Raumgebiete erfolgen kann.

Aus diesen Überlegungen kann die Entscheidungsregel für den Bayes-Klassifikator abgeleitet werden, siehe z.B. [79], [54]. Die Entscheidung beim Bayes-Klassifikator fällt zugunsten der Hypothese H_i mit dem ge-

ringsten bedingten Risiko $\mathcal{R}_i(\mathbf{m})$, das mit

$$\mathcal{R}_i(\mathbf{m}) = \sum_{k=1}^K C_{ik} P(k_k | \mathbf{m}) \quad (4.4)$$

definiert ist. Die Größen $P(k_k | \mathbf{m})$ sind die A-posteriori-Wahrscheinlichkeiten für die Ereignisse k_k nach Beobachtung des Merkmalsvektors \mathbf{m} . Das in (4.3) definierte Risiko läßt sich als $\mathcal{R} = E\{\sum_{i=1}^K \mathcal{R}_i\}$ interpretieren und gibt das sogenannte *Bayes-Risiko* des Klassifikators an.

4.3.2 Maximum-a-posteriori- und Maximum-Likelihood-Klassifikation

Werden die Kosten für richtige Entscheidungen zu Null gesetzt ($C_{ik} = 0$ für $i = k$) und für falsche Entscheidungen mit Eins festgelegt ($C_{ik} = 1$ für $i \neq k$), so ergibt sich die *Maximum-a-posteriori-Klassifikation*. Bei dieser Klassifikation fällt die Entscheidung zugunsten derjenigen Hypothese, die mit größter A-posteriori-Wahrscheinlichkeit richtig ist. Die Entscheidungsregel lautet damit:

$$\text{Entscheide } H_i, \text{ falls } d_i(\mathbf{m}) = P(k_i | \mathbf{m}) = \max_{k=1, \dots, K} P(k_k | \mathbf{m}). \quad (4.5)$$

Mit der gemischten Bayes-Regel

$$P(k_k | \mathbf{m}) = \frac{p_k p_{\mathbf{m}|k_k}(\mathbf{m}|k_k)}{p_{\mathbf{m}}(\mathbf{m})} \quad (4.6)$$

erhält man alternativ die Entscheidungsregel:

$$\text{Entscheide } H_i, \text{ falls } d_i(\mathbf{m}) = p_i p_{\mathbf{m}|k_i}(\mathbf{m}|k_i) = \max_{k=1, \dots, K} p_k p_{\mathbf{m}|k_k}(\mathbf{m}|k_k). \quad (4.7)$$

Darin sind p_k die A-priori-Wahrscheinlichkeiten der Ereignisse k_k und $p_{\mathbf{m}|k_k}(\mathbf{m}|k_k)$ die bedingten Wahrscheinlichkeitsdichten für den Merkmalsvektor \mathbf{m} unter der Beobachtung des Ereignisses k_k . Die Wahrscheinlichkeitsdichte $p_{\mathbf{m}}(\mathbf{m})$ ist klassenunabhängig und braucht deshalb bei der

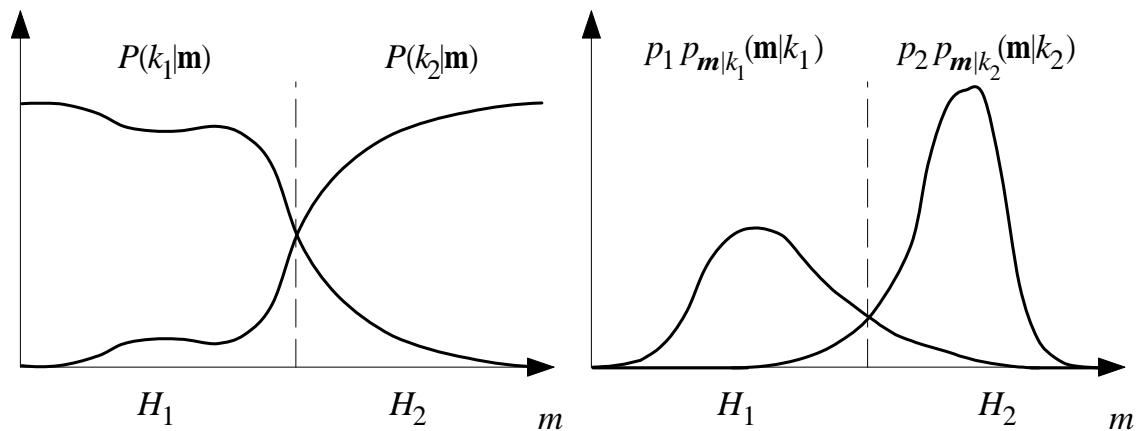


Bild 4.4: Entscheidungsbereiche aus den A-posteriori-Wahrscheinlichkeiten und den Wahrscheinlichkeitsdichtefunktionen

Entscheidungsregel nicht berücksichtigt zu werden. Falls die A-priori-Wahrscheinlichkeiten der Ereignisse gleich sind oder als gleich angenommen werden, ergibt sich aus (4.7) die Entscheidungsregel der *Maximum-Likelihood-Klassifikation*:

$$\text{Entscheide } H_i, \text{ falls } d_i(\mathbf{m}) = p_{\mathbf{m}|k_i}(\mathbf{m}|k_i) = \max_{k=1,\dots,K} p_{\mathbf{m}|k_k}(\mathbf{m}|k_k). \quad (4.8)$$

Bild 4.4 verdeutlicht die Einteilung der Entscheidungsbereiche für die Maximum-a-posteriori-Klassifikation anhand der Entscheidungsregeln (4.5) und (4.7).

In der Entscheidungsfunktion (4.5) müssen die A-posteriori-Wahrscheinlichkeiten $P(k_k|\mathbf{m})$ der verschiedenen Klassen bestimmt werden. Die Schätzung dieser Wahrscheinlichkeiten erweist sich als sehr schwierig, so daß in der Regel bei einer Klassifikatorentwicklung die Schätzung der bedingten Dichten $p_{\mathbf{m}|k_k}(\mathbf{m}|k_k)$ sowie unter Umständen der Wahrscheinlichkeiten p_k aus der Entscheidungsregel (4.7) das Hauptproblem ist. Zur Bestimmung der bedingten Dichten findet man in der Literatur eine Vielzahl von Möglichkeiten, die in *parametrische* und *nichtparametrische* Verfahren eingeteilt werden und im folgenden kurz dargestellt werden sollen.

4.4 Parametrische Klassifikationsverfahren

Eine Möglichkeit der Schätzung der bedingten Dichten liegt in der Annahme einer *gaußschen* Dichte. In diesem Fall lassen sich die bedingten Dichten als

$$p_{\mathbf{m}|k_k}(\mathbf{m}|k_k) = [(2\pi)^{\frac{n}{2}} |\mathbf{R}_k|^{\frac{1}{2}}]^{-1} e^{-\frac{1}{2}[\mathbf{m}-\boldsymbol{\mu}_k]^T \mathbf{R}_k^{-1} [\mathbf{m}-\boldsymbol{\mu}_k]}, \quad k = 1, \dots, K, \quad (4.9)$$

angeben. Die Festlegung der Entscheidungsregel beschränkt sich bei diesem Verfahren auf die Schätzung der klassenspezifischen Mittelwertsvektoren $\boldsymbol{\mu}_k$ und der klassenspezifischen Kovarianzmatrizen \mathbf{R}_k der Merkmalsvektoren. In Abhängigkeit der Kovarianzmatrizen ergeben sich verschiedene Klassifikatoren, die in der Literatur als *Gaußklassifikatoren* zusammengefaßt werden.

Quadratischer Klassifikator

Einsetzen der gaußschen Dichte (4.9) in die Entscheidungsregel (4.7) gibt

$$d_i(\mathbf{m}) = \max_{k=1, \dots, K} p_k [(2\pi)^{\frac{n}{2}} |\mathbf{R}_k|^{\frac{1}{2}}]^{-1} e^{-\frac{1}{2}[\mathbf{m}-\boldsymbol{\mu}_k]^T \mathbf{R}_k^{-1} [\mathbf{m}-\boldsymbol{\mu}_k]}. \quad (4.10)$$

Zur Vereinfachung ist ein Logarithmieren sinnvoll und es ergibt sich:

$$d_i(\mathbf{m}) = \max_{k=1, \dots, K} 2 \ln p_k - \ln |\mathbf{R}_k| - [\mathbf{m} - \boldsymbol{\mu}_k]^T \mathbf{R}_k^{-1} [\mathbf{m} - \boldsymbol{\mu}_k]. \quad (4.11)$$

Der Summand $-n \ln(2\pi)$ ist klassenunabhängig und braucht bei der Entscheidungsfindung nicht berechnet zu werden. Für verschiedene Kovarianzmatrizen \mathbf{R}_k ergibt sich eine quadratische Abhängigkeit der Trennfläche von den Komponenten des Merkmalsvektors. Bild 4.5a zeigt ein Beispiel für die Entscheidungsbereiche eines quadratischen Klassifikators für zweidimensionale Merkmalsvektoren.

Linearer Klassifikator

Bei klassenunabhängigen Kovarianzmatrizen gilt $\mathbf{R} = \mathbf{R}_k$. In diesem Fall lautet die Entscheidungsregel

$$d_i(\mathbf{m}) = \max_{k=1, \dots, K} p_k [(2\pi)^{\frac{n}{2}} |\mathbf{R}|^{\frac{1}{2}}]^{-1} e^{-\frac{1}{2}[\mathbf{m}-\boldsymbol{\mu}_k]^T \mathbf{R}^{-1} [\mathbf{m}-\boldsymbol{\mu}_k]}. \quad (4.12)$$

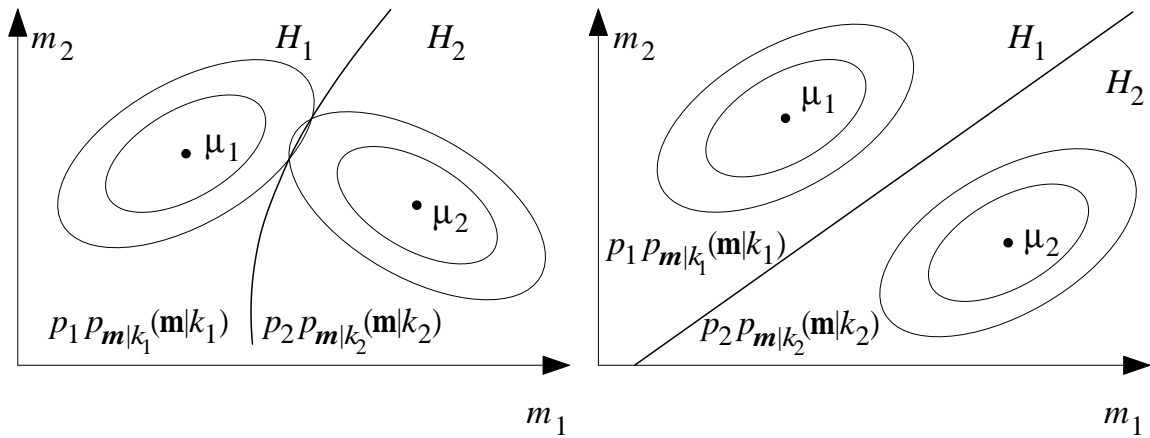


Bild 4.5: Entscheidungsbereiche für Merkmalsvektoren mit verschiedenen und mit gleichen Kovarianzmatrizen

Logarithmieren ergibt:

$$d_i(\mathbf{m}) = \max_{k=1,\dots,K} 2 \ln p_k - \boldsymbol{\mu}_k^T \mathbf{R}^{-1} \boldsymbol{\mu}_k - \boldsymbol{\mu}_k^T \mathbf{R}^{-1} \mathbf{m}. \quad (4.13)$$

Die Summanden $-n \ln(2\pi)$, $-\ln|\mathbf{R}|$ und $\mathbf{m}^T \mathbf{R}^{-1} \mathbf{m}$ sind für alle Klassen gleich und brauchen deshalb nicht berücksichtigt zu werden. Man erkennt, daß die Entscheidungsfunktion (4.13) in diesem Fall nur linear vom Merkmalsvektor abhängig ist. Es ergeben sich damit ebene Trennflächen, für zweidimensionale Merkmalsvektoren erhält man eine Gerade als Trennfunktion, siehe Bild 4.5b.

Bei gleichen A-priori-Wahrscheinlichkeiten ergibt sich der *Mahalanobis-Abstands-Klassifikator* mit dem Quadrat des Mahalanobis-Abstands als Entscheidungsfunktion:

$$d_i(\mathbf{m}) = \min_{k=1,\dots,K} [\mathbf{m} - \boldsymbol{\mu}_k]^T \mathbf{R}^{-1} [\mathbf{m} - \boldsymbol{\mu}_k]. \quad (4.14)$$

Bild 4.6a zeigt beispielhaft die Wahrscheinlichkeitsdichte des Merkmals *Fahrzeuglänge* für die Klassen *LKW* und *LKW mit Anhänger*, die mit Hilfe einer *Parzen-Schätzung* [37] aus einer Lernstichprobe ermittelt wurden. Bild 4.6b gibt die dazugehörigen gaußschen Dichten wieder.

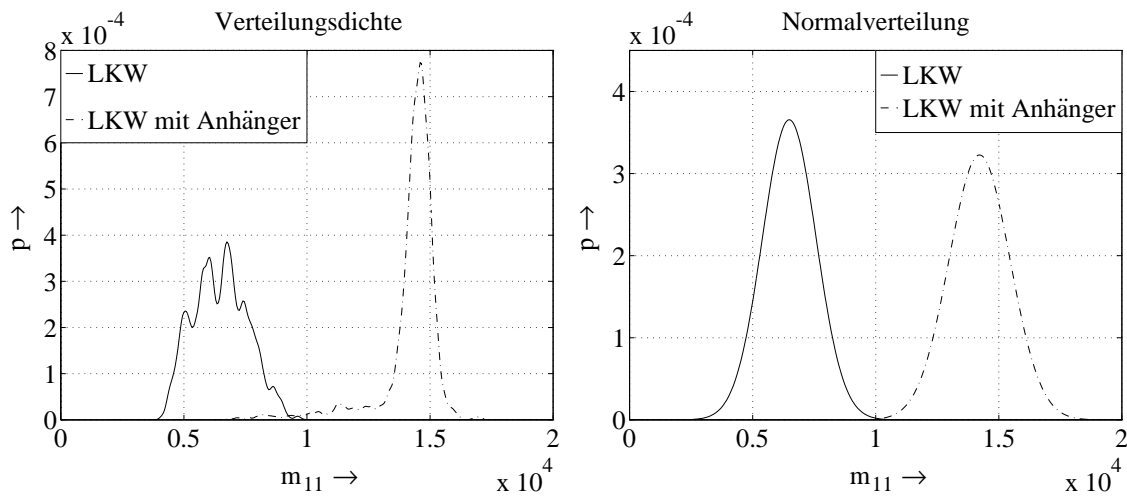


Bild 4.6: Wahrscheinlichkeitsdichten der Fahrzeuglängen für die Klassen *LKW* und *LKW mit Anhänger*: a) Parzen-Schätzung, b) Annäherung durch Normalverteilungen

4.5 Nichtparametrische Klassifikationsverfahren

Die Approximation der klassenspezifischen Verteilungen über Modellannahmen und Schätzung der Modellparameter ist nur eine Möglichkeit, die Entscheidungsfunktionen eines Klassifikators festzulegen. Nichtparametrische Klassifikatoren lassen die Struktur der zugrundeliegenden Dichtefunktion unberücksichtigt. Im folgenden sollen vier in der Praxis häufig eingesetzte Methoden der nichtparametrischen Klassifikation aufgezeigt werden.

4.5.1 Nächster-Nachbar-Klassifikator

Das Prinzip dieses Klassifikators liegt darin, den Abstand eines Merkmalsvektors zu einem bereits klassifizierten Vektor zu ermitteln und dem unbekanntem Merkmalsvektor die Klasse des nächsten Nachbarn zuzuordnen. Dieses Prinzip läßt sich auf die k nächsten Nachbarn erweitern,

der Merkmalsvektor wird dabei der am häufigsten auftretenden Klasse zugewiesen.

Der Vergleich eines Merkmalsvektors \mathbf{m} mit einem Repräsentanten $\mathbf{m}_{\ell k}$ der Klasse k_k erfolgt mit einer Metrik, z.B. mit der *euklidischen Metrik*. In diesem Fall ist der kleinste Abstand zu allen L_k Prototypen einer Klasse mit

$$d_k^2(\mathbf{m}, \mathbf{m}_{\ell k}) = \min_{\ell=1, \dots, L_k} \|\mathbf{m} - \mathbf{m}_{\ell k}\|^2 \quad (4.15)$$

gegeben und für K Klassen gilt dann folgende Entscheidungsregel:

$$\text{Entscheide } H_i, \text{ falls } d_i^2(\mathbf{m}, \mathbf{m}_{\ell k}) = \min_{k=1, \dots, K} d_k^2(\mathbf{m}, \mathbf{m}_{\ell k}). \quad (4.16)$$

Das Abstandsmaß führt auf Trennfunktionen, die aus allen Mittelsenkrechten zwischen jeweils zwei Prototypen zusammengesetzt sind, d.h. auf stückweise lineare Trennfunktionen, siehe Bild 4.7. Der Nachteil dieses Klassifikators liegt darin, daß im allgemeinen eine große Anzahl an Prototypen gespeichert werden müssen und in der Anwendungsphase die Abstände zu allen Prototypen berechnet werden müssen. Der Vorteil wiederum ist darin zu sehen, daß die Gebietsaufteilung im Merkmalsraum auch kompliziertere Formen zuläßt. Bei genügend großem Stichprobenumfang kann mit dem k -NN-Klassifikator eine Abschätzung der Dichtefunktion der Muster im Merkmalsraum und damit eine Abschätzung der zu erwartenden Fehlerrate des Klassifikators vorgenommen werden [20], [37]. Hierauf wird in Abschnitt 4.8 noch eingegangen.

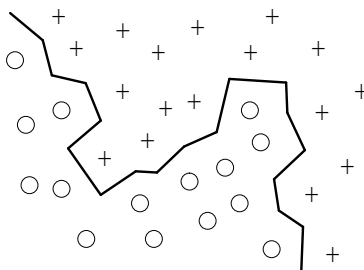


Bild 4.7: Trennfunktionen des NN-Klassifikators

4.5.2 Polynomklassifikator

Ein pragmatisches Konzept zur Bestimmung der Verbunddichten liegt darin, die unbekannte Dichte $p_{\mathbf{m}|k_k}(\mathbf{m}|k_k)$ durch eine Funktion $d_k(\mathbf{m})$ anzunähern. Dieses führt auf ein Variationsproblem.

Gesucht ist der Vektor $\mathbf{d}(\mathbf{m}) = [d_1(\mathbf{m}), \dots, d_K(\mathbf{m})]^T$, der den Zielvektor

$$\mathbf{c} = [c_1, \dots, c_K]^T \quad (4.17)$$

mit

$$c_i = \begin{cases} 1 & \text{falls } k_i \text{ zum Vektor } \mathbf{m} \text{ geführt hat,} \\ 0 & \text{falls } k_j, j \neq i, \text{ zum Vektor } \mathbf{m} \text{ geführt hat,} \end{cases} \quad (4.18)$$

im Sinne des Quadratmittelansatzes

$$J(\mathbf{d}) = E\{|\mathbf{c} - \mathbf{d}|^2\} \stackrel{!}{=} \min \quad (4.19)$$

am besten annähert.

Diese Vorgehensweise führt auf den Regressionsklassifikator, bei dem der Schätzvektor $\mathbf{d}(\mathbf{m})$ durch

$$\mathbf{d}(\mathbf{m}) = \mathbf{A}^T \mathbf{w}(\mathbf{m}) \quad (4.20)$$

berechnet wird. Ein Sonderfall dieser Klassifikatoren ist der *Polynomklassifikator* [72], bei dem die Entscheidungsfunktionen durch einen Polynomvektor $\mathbf{w}(\mathbf{m})$ gebildet werden. Beim *linearen* Polynomansatz wird der Polynomvektor mit

$$\mathbf{w}(\mathbf{m}) = [1, m_1, \dots, m_N]^T. \quad (4.21)$$

vorgegeben, beim vollständig *quadratischen* Polynomansatz hat der Polynomvektor folgendes Format:

$$\mathbf{w}(\mathbf{m}) = [1, m_1, \dots, m_N, m_1^2, m_1 m_2, \dots, m_{N-1} m_N, m_N^2]^T. \quad (4.22)$$

Die Grenzflächen zwischen zwei Klassen sind Flächen zweiter Ordnung.

Die Koeffizientenmatrix \mathbf{A} wird im Sinne des quadratischen Fehlers optimiert,

$$J(\mathbf{A}) = E\{|\mathbf{c} - \mathbf{A}^T \mathbf{w}(\mathbf{m})|^2\} \stackrel{!}{=} \min, \quad (4.23)$$

und kann nach [72] mit

$$\mathbf{A} = E\{\mathbf{w}\mathbf{w}^T\}^{-1} E\{\mathbf{w}\mathbf{c}^T\} \quad (4.24)$$

bestimmt werden. Die Matrizen $E\{\mathbf{w}\mathbf{w}^T\}$ und $E\{\mathbf{w}\mathbf{c}^T\}$ werden aus einer Lernstichprobe ermittelt.

Die Dimension des Polynomvektors steigt sehr schnell mit dem Grad des Polynoms und der Dimension des Merkmalsvektors. Deshalb wird in der Regel der Grad des Polynoms selten größer als zwei gewählt, oder es wird ein unvollständiger Polynomansatz gewählt [72].

Bild 4.8 zeigt beispielhaft die Aufteilung des Entscheidungsraums für die Klassen *LKW* und *LKW mit Anhänger* für einen vollständig quadratischen Polynomklassifikator mit $N = 16$ Merkmalen.

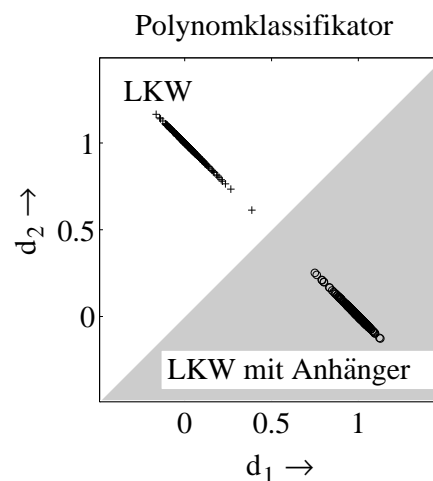


Bild 4.8: Aufteilung des Entscheidungsraums für die Klassen *LKW* und *LKW mit Anhänger* bei $N = 16$ Merkmalen

4.5.3 Histogrammklassifikator

Ein einfacher Ansatz zur Schätzung der bedingten Dichten ist die Ermittlung einer Häufigkeitsverteilung aus der Lernstichprobe. Diese Verteilungen können für jedes Merkmal als Histogramme $\hat{p}_{\mathbf{m}|k_k}(m_j|k_k)$, $j = 1, \dots, N$, $k = 1, \dots, K$, aufgetragen werden.

Im Fall statistisch unabhängiger Merkmale kann die bedingte Dichte durch

$$\hat{p}_{\mathbf{m}|k_k}(\mathbf{m}|k_k) = \prod_{j=1}^N \hat{p}_{m_j|k_k}(m_j|k_k) \quad (4.25)$$

angenähert werden und die Entscheidung nach Gleichung (4.7) erfolgen.

Eine praktische Schwierigkeit liegt bei dieser Technik in der geeigneten Wahl der Intervallaufteilung. Die Intervalle müssen einerseits hinsichtlich einer guten Approximation klein genug sein, andererseits müssen sie so groß gewählt werden, daß die Anzahl der in ein Intervall fallenden Stichprobenelemente eine hinreichend zuverlässige Schätzung der Wahrscheinlichkeitsdichte an der betrachteten Stelle gestattet. Zusätzlich müssen die statistischen Abhängigkeiten der Merkmale erfaßt werden, so daß häufig eine sehr große Lernstichprobe benötigt wird. Der Vorteil dieses Klassifikators liegt darin, daß keine Einschränkungen bezüglich der Form der Trennflächen zwischen den Raumgebieten bestehen.

In Bild 4.9 sind die zu den Wahrscheinlichkeitsdichten aus Bild 4.6a gehörenden Häufigkeitsverteilungen des Merkmals *Fahrzeuglänge* der Klassen *LKW* und *LKW mit Anhänger* aufgetragen.

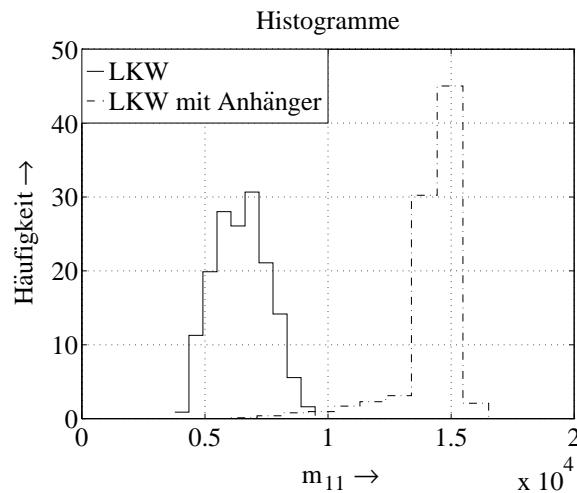


Bild 4.9: Annäherung der Wahrscheinlichkeitsdichten der Fahrzeuglängen durch Histogramme für die Klassen *LKW* und *LKW mit Anhänger*

4.5.4 Klassifikation mit neuronalen Netzwerken

In den letzten Jahren haben *neuronale Netzwerke* Einzug in eine Vielzahl von Klassifikationsaufgaben erhalten. Der Grund hierfür liegt vor allem im Auffinden effizienter Lernverfahren. Kompakte Übersichten über neuronale Netzwerke finden sich in den Artikeln [51], [52]. Detaillierte Beschreibungen können [41], [48] oder [58] entnommen werden.

Im vorliegenden Problem reicht eine Beschränkung auf sogenannte *Multi-Layer-Feedforward-Netzwerke*, die in der Literatur häufig auch als *Multi-Layer-Perceptron* (MLP) bezeichnet werden. Diese Netzwerke zeichnen sich dadurch aus, daß sie aus mehreren Schichten (*Layer*) bestehen können, die durch Neuronen* miteinander verknüpft sind und nur in Vorwärtsrichtung betrieben werden, siehe Bild 4.10. Die Merkmale des Merkmalsvektors liegen an der Eingangsschicht an. In der Ausgangsschicht wird die Entscheidungsfunktion des Netzwerks gebildet und damit die Klassifikation durchgeführt. Dazwischen können sich eine oder mehrere verborgene Schichten (*hidden Layer*) befinden. Entfallen die verborgenen Schichten, so handelt es sich um ein *Single-Layer-Perceptron*.

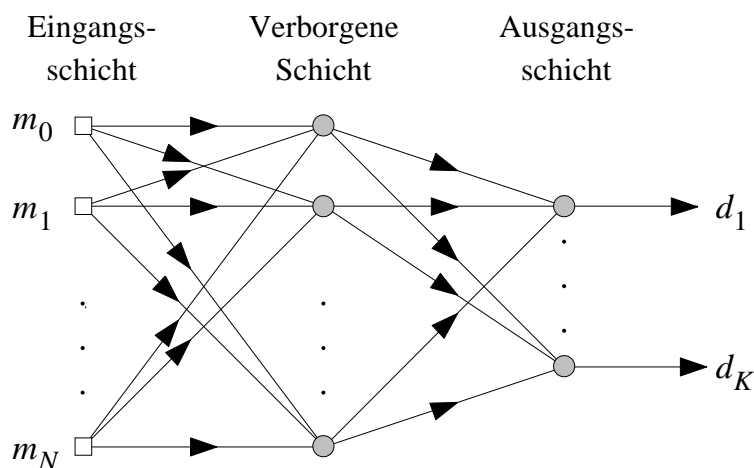


Bild 4.10: Multi-Layer-Perceptron

*Die Bezeichnung Neuron ist in sehr grober Anlehnung an die Funktionalität von Nervenzellen gewählt

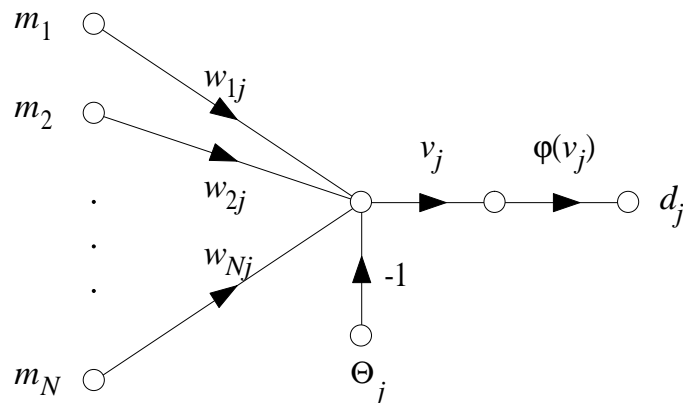


Bild 4.11: Signalflußgraph eines Single-Layer-Perceptrons

Die Funktionsweise dieses Netzwerks läßt sich am einfachsten anhand eines Single-Layer-Perceptrons mit einem Ausgang erläutern. Der Signalflußgraph eines derartigen Netzwerks ist in Bild 4.11 dargestellt.

Die Eingangsschicht stellt die Merkmale $\mathbf{m} = [m_1, \dots, m_N]$ zur Verfügung. Bei der Übergabe zur nächsten Schicht werden alle Merkmale mit den Faktoren w_{ij} gewichtet. Der Trick der neuronalen Netzwerke liegt in den nachfolgenden Neuronen, die in einem ersten Schritt die gewichteten Eingangsdaten summieren,

$$v_j = \sum_{i=1}^N w_{ij} m_i - \Theta_j = \sum_{i=0}^N w_{ij} m_i. \quad (4.26)$$

Die Konstante Θ_j ist ein Schwellwert oder *Bias*, der häufig auch als zusätzliches Eingangssignal $w_{0j} m_0$ mit $m_0 = -1$ in die Summierung eingeht. In einem zweiten Schritt wird die Summe über eine *nichtlineare* Aktivierungsfunktion an den Ausgang weitergeleitet. Die Nichtlinearität muß eine differenzierbare Funktion sein, wie z.B. die *Sigmoidfunktion*

$$d_j = \varphi(v_j) = \frac{1}{1 + \exp(-v_j)}, \quad -\infty < v_j < \infty. \quad (4.27)$$

Die Ausgangsdaten können jetzt wieder die Eingangsdaten einer verborgenen Schicht sein. Durch die Aktivierungsfunktionen erhält man eine nichtlineare Abbildung des Merkmalsraums auf den Entscheidungsraum.

Neuronale Netzwerke müssen ebenso wie andere Klassifikatoren trainiert werden. Hierzu dienen vorklassifizierte Merkmalsvektoren einer Lernstichprobe. Die Merkmalsvektoren liegen als Eingangsvektoren \mathbf{m} vor, die Fahrzeugklassen bilden die Ausgangssignale \mathbf{d} und sollen sich einem Zielvektor \mathbf{c} annähern, dessen Elemente sämtlichst Null sind. Lediglich das Element c_j soll den Wert Eins annehmen, wenn das Eingangssignal der Klasse k_j entstammt. Hieraus wird die Verwandtschaft der neuronalen Netzwerke mit dem Polynomklassifikator ersichtlich, die die gleiche Zielfunktion besitzen. Die Bestimmung der unbekanntenen Gewichtungsfaktoren erfolgt im Gegensatz zum Polynomklassifikator allerdings iterativ mit Hilfe des sogenannten *Backpropagation-Algorithmus* [71]. Dieser Algorithmus basiert auf einer iterativen Gradientenmethode, bei der das Fehlersignal des Ausgangs eines Neurons j für das ℓ -te Trainingsmuster,

$$e_j(\ell) = d_j(\ell) - c_j(\ell), \quad (4.28)$$

durch wiederholte Änderung der Gewichte in Richtung des negativen Gradienten minimiert wird. Es ergibt sich ein Minimierungsproblem für den Gesamtfehler aller Lernstichprobenelemente

$$J(\mathbf{w}) = \frac{1}{L} \sum_{\ell=1}^L \frac{1}{2} \sum_j e_j^2(\ell) \stackrel{!}{=} \min. \quad (4.29)$$

Auf die Einzelheiten des Algorithmus soll hier nicht näher eingegangen werden, eine genaue Beschreibung findet sich z.B. in [41].

Der Vorteil der MLP-Netzwerke liegt in der Fähigkeit, komplexe Trennflächen zu generieren. Man kann zeigen, daß die Trennflächen eines Single-Layer-Perceptrons linear sind, ein Two-Layer-Perceptron konvexe Trennflächen zweiter Ordnung aber zum Teil auch nicht zusammenhängende Entscheidungsgebiete erzeugen kann und mit einem Three-Layer-Perceptron Trennflächen jeglicher Art erzeugt werden können [51], [38]. Zudem kann gezeigt werden, daß derartige neuronale Netzwerke bei erfolgreicher Minimierung der Fehlerfunktion (4.29) eine Schätzung für die unbekanntenen A-posteriori-Wahrscheinlichkeiten der Merkmalsvektoren liefern [69].

Die Aufteilung des Entscheidungsraums bei der binären Klassifikation der Klassen *LKW* und *LKW mit Anhänger* mit neuronalen Netz-

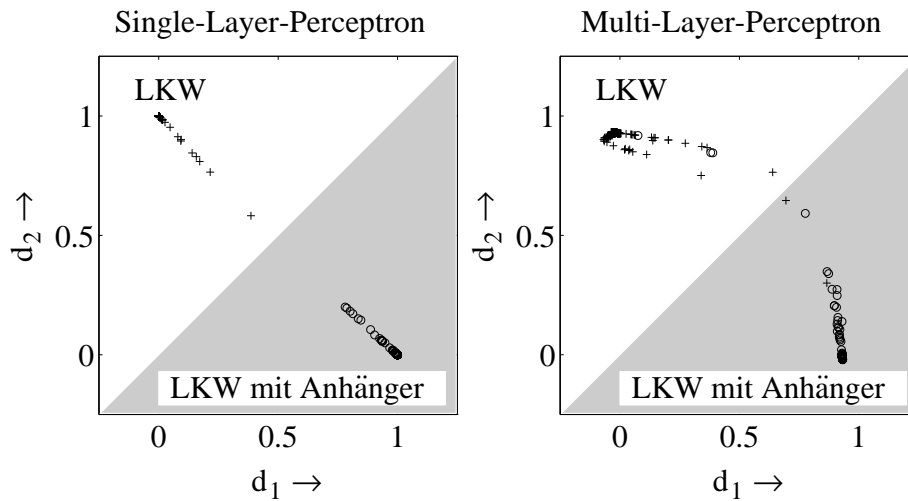


Bild 4.12: Aufteilung des Entscheidungsraums für die Klassen *LKW* und *LKW mit Anhänger* für ein Single-Layer-Perceptron mit $N = 16$ Merkmalen (links) und ein Multi-Layer-Perceptron mit $N = 2$ Merkmalen (rechts)

werken ist in Bild 4.12 dargestellt. Die getestete Stichprobe entspricht der Lernstichprobe. Man erkennt, daß mit einem Single-Layer-Perceptron mit Merkmalsvektoren der Dimension $N = 16$ eine fehlerfreie Klassifikation durchgeführt werden kann. Die beiden Klassen sind linear separierbar. Besitzt man nur sehr wenige Merkmale, so kann mit einem Multi-Layer-Perceptron trotzdem ein gutes Klassifikationsergebnis erzielt werden. Als Eingangssignal eines Multi-Layer-Perceptrons mit einer verborgenen Schicht wurden lediglich zwei Merkmale der Fahrzeugsignale benutzt. Die Anzahl der verborgenen Knoten, die die Zahl der Freiheitsgrade des Klassifikators bestimmt, beträgt ebenfalls 16.

4.6 Mehrstufige Klassifikationsverfahren

Für Mehrklassenprobleme können mit linearen Klassifikatoren je nach Wahl der Entscheidungsregeln verschiedenartige Entscheidungsbereiche erzeugt werden [31]. Die Entscheidungsfunktion eines linearen Klassifika-

tors läßt sich mit

$$d(\mathbf{m}) = \mathbf{w}_i^T \mathbf{m} \quad (4.30)$$

mit

$$\mathbf{w}_i = [w_0, w_1, \dots, w_N]^T \quad \text{und} \quad \mathbf{m} = [1, m_1, \dots, m_N]^T \quad (4.31)$$

beschreiben. Mit der Entscheidungsregel (4.7) ergeben sich Trennfunktionen, wie sie in Bild 4.13a am Beispiel eines Dreiklassenproblems dargestellt sind. Die Trennfunktionen zweier Klassen k_i und k_j sind in diesem Fall durch die Gleichungen

$$d_i(\mathbf{m}) = d_j(\mathbf{m}) \quad (4.32)$$

gegeben, und es wird jedem unbekanntem Merkmalsvektor damit eine Klasse zugewiesen, es existiert kein undefiniertes Gebiet.

Alternativ können auch $K(K-1)/2$ lineare Entscheidungsfunktionen benutzt werden, mit denen jede Klasse paarweise von jeder anderen Klasse getrennt wird. Man erhält Entscheidungsfunktionen $d_{ij}(\mathbf{m})$ und die Entscheidungsregel ergibt sich zu:

$$\text{Entscheide } H_i, \text{ falls } d_i(\mathbf{m}) = \mathbf{w}_{ij}^T \mathbf{m} > 0 \quad \text{für alle } j = 1, \dots, K, j \neq i. \quad (4.33)$$

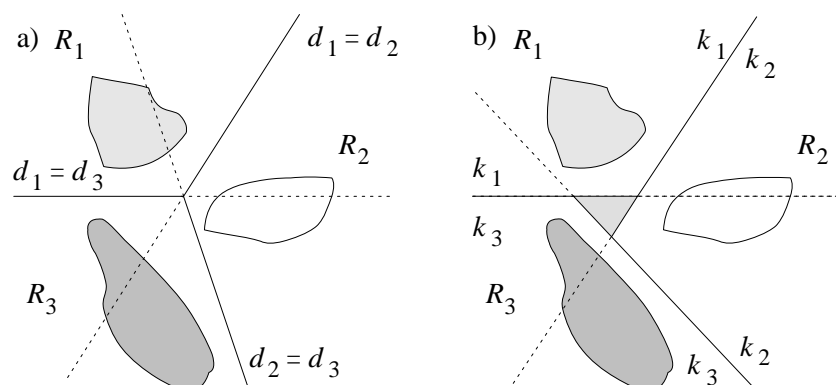


Bild 4.13: Entscheidungsgebiete eines linearen Klassifikators für ein Dreiklassenproblem

In diesem Fall können auch Gebiete entstehen, in denen keine eindeutige Klassifikation durchgeführt werden kann, siehe Bild 4.13b. Der durch diese Entscheidungsregel verursachte Mehraufwand kann durch eine zweistufige Klassifikation umgangen werden. In der Regel muß nämlich kein Vergleich zwischen allen Klassen stattfinden. Relativ schnell kann durch eine Vorklassifikation festgestellt werden, ob ein zu klassifizierender Merkmalsvektor überhaupt einer der möglichen Klassen entstammen kann. Vergleicht man die Merkmale des unbekannten Musters mit den aus einer Lernstichprobe ermittelten minimalen und maximalen Werten für die einzelnen Klassen, so reduziert sich der Vergleich für das vorliegende Klassifikationsproblem auf maximal vier Klassen. Bild 4.14 verdeutlicht diese Vorgehensweise am Beispiel der Fahrzeuglänge als Merkmal. Besitzt ein unbekanntes Muster eine Fahrzeuglänge von z.B. 4800 mm, so kann man der Dichteverteilung dieses Merkmals entnehmen, daß lediglich zwischen den Klassen *PKW* und *Kleintransporter* entschieden werden muß.

Die Gewichtsvektoren \mathbf{w}_{ij} können z.B. durch einen Polynomansatz oder mit Hilfe des *Backpropagation-Algorithmus* für ein Single-Layer-Perceptron bestimmt werden.

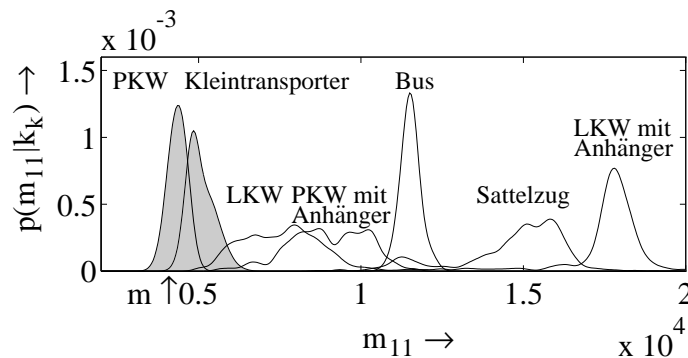


Bild 4.14: Vorklassifikation am Beispiel der Fahrzeuglänge als Merkmal

4.7 Verfahrensvergleich

Bei der Untersuchung der verschiedenen Klassifikationsverfahren liegt der Schwerpunkt darin, einen möglichst einfachen Klassifikator zu finden, der mit wenigen Parametern ein gutes Klassifikationsergebnis liefert und als Echtzeitsystem implementiert werden kann. Aus diesem Grund wurde auf eine feinere Aufteilung der Fahrzeugklassen verzichtet, wie sie z.B. durch eine Cluster-Bildung [31] erreicht werden kann.

Die Bewertung eines Klassifikators ist in der Regel weniger trivial als es zunächst erscheint. Die Bestimmung des Bayes-Fehlers, der das bestmögliche Klassifikationsergebnis für das vorliegende Problem angibt, ist in der Praxis äußerst schwierig, auf sie wird in Abschnitt 4.8 noch näher eingegangen. Die verschiedenartige Einteilung der Raumgebiete bei den in den vorangegangenen Abschnitten aufgeführten Klassifikatoren führt zu verschiedenen Fehlerraten der jeweiligen Klassifikatoren, die bestenfalls so gering wie der Bayes-Fehler sind.

Grundlage zur Bestimmung der Entscheidungsfunktionen und damit auch der Fehlerrate eines Klassifikators ist die Lernstichprobe. Die Güte eines Klassifikators kann durch die *Reklassifikation* der Lernstichprobe bestimmt werden. Allerdings kann es bei der Reklassifikation zu einer Überadaptation des Klassifikators kommen. In diesem Fall paßt er sich den Eigenschaften der Lernstichprobe zu sehr an und suggeriert ein zu gutes Klassifikationsergebnis. Dieses geschieht vor allem dann, wenn die Lernstichprobe einen geringen Umfang besitzt. In solchen Fällen kann es sein, daß die Erkennungsraten des Klassifikators für unbekannte Muster (*Fremdklassifikation*) sehr schlecht sind. Insofern ist es wichtig, eine zweite unabhängige Teststichprobe zu verwenden, die stellvertretend für den Anwendungsfall steht. Probleme entstehen hier, wenn nur sehr wenige Stichproben zur Verfügung stehen. In diesem Fall müssen Verfahren herangezogen werden, mit denen die Entscheidungsfunktionen optimal eingestellt werden können und trotzdem eine Aussage über die Erkennungsrate im Anwendungsfall gemacht werden kann.

Vorerst sollen die Fehlerraten[†] der verschiedenen Klassifikatoren für die Reklassifikation präsentiert werden. Abschließend folgt ein Vergleich der verschiedenen Klassifikatoren anhand ihrer Fehlerraten bei der Klassifikation der Lernmenge sowie bei einer Fremdklassifikation.

Stichprobe

Die Lernstichproben der Fahrzeugklassifikation bestehen aus ca. 4000 Fahrzeugsignalen einer 2,5m-Schleife und ca. 6000 Fahrzeugsignalen einer 1m-Schleife. Über die Anzahl der notwendigen Stichproben pro Klasse kann a priori keine Aussage gemacht werden, die Anzahl ist vielmehr abhängig von der sogenannten inneren Dimensionalität (*intrinsic dimensionality*) der Muster [37]. Im vorliegenden Fall kann von einer ausreichenden Lernstichprobe ausgegangen werden, lediglich die Musteranzahl der Klasse *Bus* muß für beide Schleifentypen als zu gering angesehen werden.

Auswahl der Merkmale

Die Fahrzeugsignale werden auf $n = 32$ Werte normiert, nach der Transformation ergeben sich bei der Diskriminanzanalyse $m = K - 1 = 6$ Merkmale, bei der Karhunen-Loève-Transformation kann die Anzahl m der Eigenvektoren der Transformationsmatrix variabel gehalten werden. Zusätzlich enthält der Merkmalsvektor noch fünf aus den Überfahrkurven selektierte Sondermerkmale, die sich aus der Fahrzeuglänge, einer Anhängerdetektion sowie aus einer Minimumbestimmung ergeben [55].

Ein Vergleich einiger Klassifikatoren bezüglich der gewählten Transformation der Fahrzeugsignale ist in Bild 4.15 wiedergegeben. Zum Vergleich ist die Anzahl der Merkmale m aus den normierten Signalen für beide Verfahren identisch. Es zeigt sich, daß die Fehlerraten der Klassifikatoren bei einer Diskriminanzanalyse geringfügig größer sind als bei einer Karhunen-Loève-Transformation, was sich mit einer ungünstigen Aufteilung der Klassenschwerpunkte und der Verteilungsdichten im vorliegenden Problem begründen läßt.

[†]Die im folgenden dargestellten Fehlerraten stellen jeweils die Gesamtfehlerrate des Klassifikators dar. Auf eine detaillierte Darstellung der Fehlerraten zwischen den einzelnen Klassen wird hier verzichtet, da der Vergleich der einzelnen Klassifikatoren im Vordergrund steht.

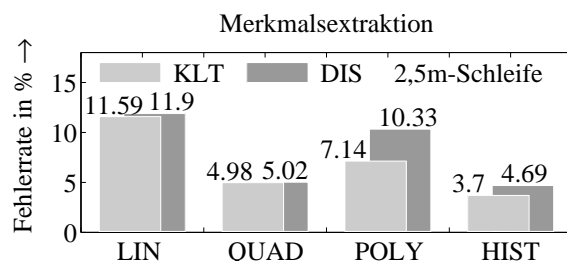


Bild 4.15: Fehlerraten einiger Klassifikatoren bei einer Merkmalsextraktion mit der Karhunen-Loève-Transformation und mit einer Diskriminanzanalyse

Gaußklassifikator

Bei der Nachbildung der klassenweisen Verteilungsdichten mit gaußschen Dichten müssen die Klassenschwerpunkte und Kovarianzmatrizen für beide Schleifentypen anhand der Lernstichprobe bestimmt werden. Die aufwendige Inversion der Kovarianzmatrizen muß nur einmalig in der Lernphase durchgeführt werden und ist hier unproblematisch. Die sich aus der Klassifikation der Lernstichprobe ergebenden Fehlerraten der Klassifikatoren sind in Bild 4.16 in Abhängigkeit der Dimension der Merkmalsvektoren bei einer Karhunen-Loève-Transformation der Fahrzeugsignale wiedergegeben. Die bessere Erkennungsleistung des quadratischen Klassifikators, bei dem die Kovarianzmatrizen klassenabhängig gebildet werden, ist ersichtlich. Gleichzeitig zeigt sich, daß bei einer größeren Anzahl von Merkmalen auch die Erkennungsleistung beider Klassifikatoren steigt.

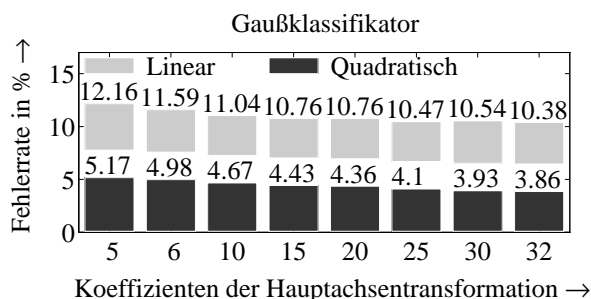


Bild 4.16: Fehlerraten der Gaußklassifikatoren in Abhängigkeit der Koeffizienten der Hauptachsentransformation (2,5m-Schleife)

Nächster-Nachbar-Klassifikator

Da bei diesem Klassifikator die Verteilung der Muster im Merkmalsraum anhand einer Vielzahl von Prototypen beschrieben wird, erzielt man mit dem NN-Klassifikator eine sehr gute Erkennungsrate. Bezieht man mehrere Prototypen bei der Klassifikation ein, so lassen sich theoretisch noch bessere Erkennungsraten erzielen. Bild 4.17 gibt die Fehlerraten des k -NN-Klassifikators für die 1m- und die 2,5m-Schleife wieder. Es zeigt sich, daß für das vorliegende Problem eine Erhöhung der Nachbaranzahl keine Vorteile bringt, was sich durch die für diesen Fall zu kleine Lernstichprobe begründen läßt. Bei den aufgeführten Fehlerraten handelt es sich um die Fehlerraten einer Fremdklassifikation, da bei diesem Klassifikator jedes Muster als Testmuster mit den restlichen Mustern als Lernmuster klassifiziert wird.

Zu berücksichtigen ist, daß der NN-Klassifikator aus Aufwandsgründen nur als Referenzklassifikator benutzt werden kann und eine praktische Implementierung nicht in Frage kommt.

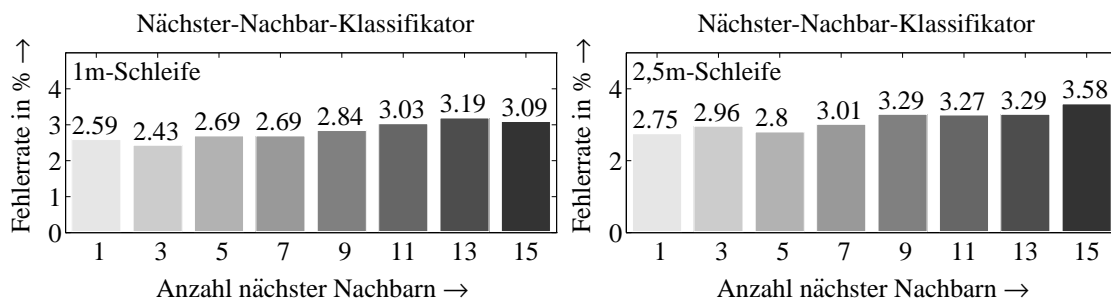


Bild 4.17: Fehlerraten des k -NN-Klassifikators in Abhängigkeit der Anzahl der nächsten Nachbarn

Polynomklassifikator

Bild 4.18 zeigt die Fehlerraten der Reklassifikation für einen linearen und einen quadratischen Polynomklassifikator. Die Erkennungsrate des quadratischen Klassifikators ist deutlich besser als die des linearen Klassifikators. Dieses begründet sich in der intern verwendeten Komponentenzahl, die die Zahl der Freiheitsgrade des Klassifikators bestimmt. Im Fall von 10 Merkmalen bildet der quadratische Klassifikator aufgrund der

Produktterme in Gleichung (4.22) intern 66 Merkmale, die die Trennbarkeit der einzelnen Klassen verbessern. Die maximale Anzahl an Merkmalen wurde beim quadratischen Klassifikator auf 25 begrenzt. Eine weitere Erhöhung führt zu numerischen Problemen bei der Bestimmung der Koeffizientenmatrix infolge der Invertierung nach Gleichung (4.24). Zudem steigt der Aufwand zur Speicherung der Koeffizientenmatrix quadratisch mit der Anzahl der Merkmale.

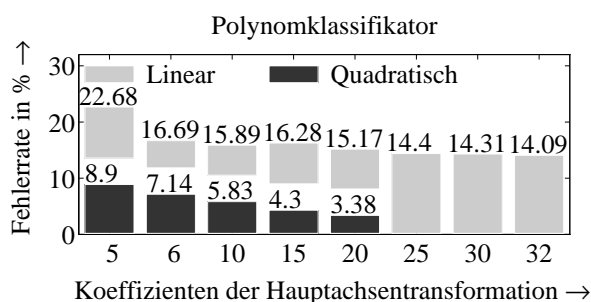


Bild 4.18: Fehlerraten des linearen und quadratischen Polynomklassifikators in Abhängigkeit der Merkmalsanzahl (1m-Schleife)

Histogrammklassifikator

Eine Annäherung der bedingten Dichten durch eine aus der Lernstichprobe abgeleitete Häufigkeitsverteilung führt zum Histogrammklassifikator, bei dem die Wahl der Intervallaufteilung, wie in Abschnitt 4.5.3 angedeutet, problematisch ist. Bild 4.19 gibt die Fehlerraten des Histogrammklassifikators in Abhängigkeit der Intervallanzahl für beide Schleifen wieder. Man erkennt, daß eine zu grobe Intervallaufteilung die bedingte Dichte nur unzureichend nachbilden kann und dementsprechend die Fehlerraten höher sind. Bei kleiner werdenden Intervallen sinkt auch die Fehlerrate, allerdings muß berücksichtigt werden, daß es bei der Klassifikation der Lernstichprobe zu einer Überadaption des Klassifikators kommen kann.

Klassifikation mit neuronalen Netzwerken

Lernverfahren für neuronale Netzwerke sind sehr aufwendig. Da die Gewichtungsfaktoren allerdings nur einmalig in der Lernphase bestimmt werden müssen, ergibt sich in der Anwendungsphase ein sehr einfacher Klassifikator. Die erforderliche Anzahl an verborgenen Schichten und Knoten kann für ein gegebenes Problem a priori nicht bestimmt werden, sondern

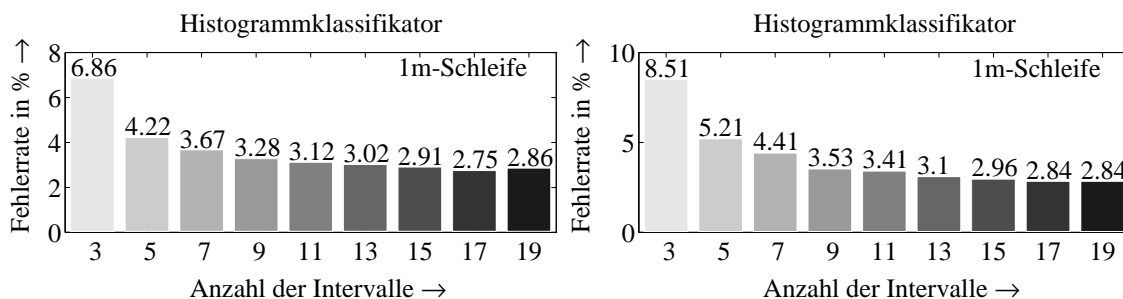


Bild 4.19: Fehllraten des Histogrammklassifikators in Abhängigkeit der Intervallanzahl

muß anhand der Klassifikationsrate des Netzwerks festgelegt werden [41]. Aufgrund von Konvergenzproblemen des Backpropagation-Algorithmus ist die maximale Anzahl an verborgenen Schichten für den Anwendungsfall der Fahrzeugklassifikation auf Eins beschränkt. Bild 4.20 stellt die Fehlerrate eines Two-Layer-Perceptrons in Abhängigkeit der Anzahl verborgener Knoten dar. Es zeigt sich, daß sich bei einer Erhöhung der Knotenanzahl die Klassifikationsrate verbessern läßt. Allerdings nimmt der Speicheraufwand für die Gewichtsvektoren auch linear mit der Anzahl der Knoten zu.

Sehr gute Ergebnisse lassen sich mit einem zweistufigen Klassifikator mit einem Single-Layer-Perceptron erzielen. Hier müssen die Gewichte für binäre Entscheidungsfunktionen gelernt werden, der zugehörige Algorithmus konvergiert dabei sehr schnell, da die Anzahl der Gewichte die Anzahl der Merkmale lediglich um Eins übersteigt. Der Vorteil dieses Verfahrens liegt darin, daß bei einer binären Klassifikation viele Fahrzeugklassen linear separierbar sind. Von $K(K-1)/2 = 21$ möglichen Entscheidungen müssen bei 15 Merkmalen lediglich acht Entscheidungsfunktionen für die Klassenpaare *PKW/Kleintransporter*, *PKW/LKW*, *Kleintransporter/LKW*, *LKW/Bus*, *LKW/LKW mit Anhänger*, *LKW/Sattelzug* und *LKW mit Anhänger/Sattelzug* berechnet und gespeichert werden. Eine Verbesserung des Verfahrens kann zusätzlich erreicht werden, wenn für schlecht separierbare Klassenpaare quadratische Trennfunktionen mit einem Multi-Layer-Perceptron generiert werden.

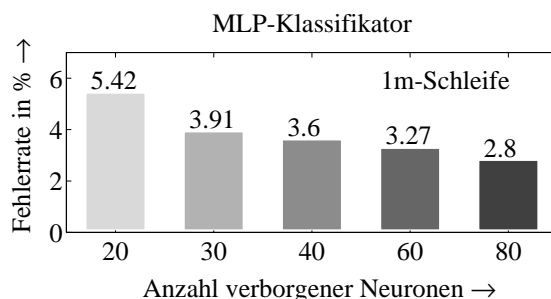


Bild 4.20: Fehlerraten des Multi-Layer-Perceptrons in Abhängigkeit der Anzahl verborgener Neuronen

Verfahrensvergleich

Vergleicht man die in dieser Arbeit untersuchten Klassifikatoren, so muß neben der Fehlerrate auch der Rechen- und Speicheraufwand berücksichtigt werden. Aus diesem Grund ist eine Klassifikation mit dem NN-Klassifikator und mit dem quadratischen Polynomklassifikator ineffizient. Am geeignetesten ist die zweistufige lineare Klassifikation mit einem neuronalen Netzwerk. Die Fehlerrate liegt sehr nahe am Bayes-Fehler und der Speicheraufwand beschränkt sich auf maximal $K(K - 1)/2$ Gewichtsvektoren der Dimension des Merkmalsvektors. Unter den übrigen Klassifikatoren erweist sich der Histogrammklassifikator als günstig, da auch hier der Rechen- und Speicheraufwand gering ist und die Fehlerrate anderen Klassifikatoren gleichzusetzen ist, wie Bild 4.21 zeigt.

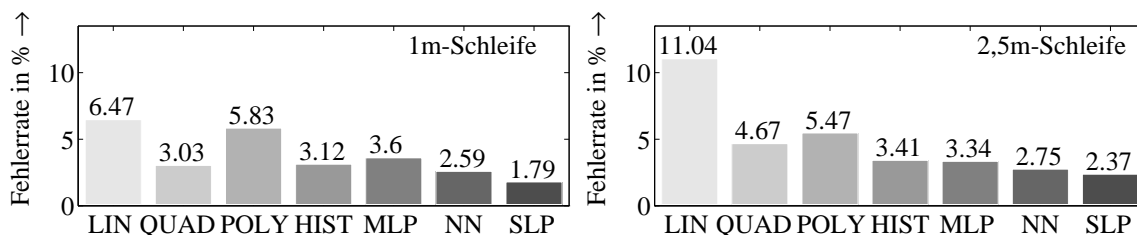


Bild 4.21: Fehlerraten bei der Reklassifikation mit 15 Merkmalen

Ein Verfahren zur Klassifikation einer unabhängigen Teststichprobe bei geringem Lernstichprobenumfang ist mit der sogenannten *Leave-One-Out*-Methode [37] gegeben. Bei diesem Verfahren wird ein Muster der Stichprobe als Testmuster herausgenommen. Der entsprechende Klassi-

ifikator wird mit den übrigen Stichprobenmustern trainiert. Die Klassifikation des Testmusters ist entweder erfolgreich oder falsch. Diese Vorgehensweise wird für alle Elemente der Lernstichprobe wiederholt und aus der Anzahl der falsch klassifizierten Fahrzeuge ergibt sich dann eine Fehlerrate für die Fremdklassifikation. Diese Methode ist allerdings sehr rechenintensiv, da für alle Klassifikatoren bei dem gegebenen Stichprobenumfang ca. 10000 Trainingsläufe für beide Schleifen durchgeführt werden müssten. Deshalb werden für Fahrzeugklassen, für die die Stichprobe umfangreicher ist, mehrere Fahrzeuge als Testmuster verwendet. Bild 4.22 gibt die Fehlerraten einiger Klassifikatoren bei einer Fremdklassifikation nach der Leave-One-Out-Methode wieder.

Dieses Verfahren ist für die Lernphase der neuronalen Netzwerke ungeeignet, da die Trainingsphase sehr umfangreich ist. Für derartige Klassifikatoren wird deshalb die Gesamtstichprobe in eine Lern- und Teststichprobe sowie in eine zusätzliche dritte Stichprobe getrennt. Diese Methode ist in der Literatur unter dem Begriff *Cross Validation* bekannt [41]. Mit Hilfe der Erkennungsleistung der zusätzlichen Stichprobe läßt sich während des iterativen Lernverfahrens feststellen, wann eine Überadaptation eintritt. Die vor der Überadaptation ermittelten Gewichtungsfaktoren werden dann zur Klassifikation der unabhängigen Teststichprobe eingesetzt. Anhand der resultierenden Erkennungsrate kann auf die Leistung des Klassifikators in der Anwendungsphase geschlossen werden. Für das vorliegende Problem ist der Umfang der Gesamtstichprobe allerdings zu gering, um eine zuverlässige Schätzung des Klassifikationsfehlers bei einer Fremdklassifikation durchzuführen.

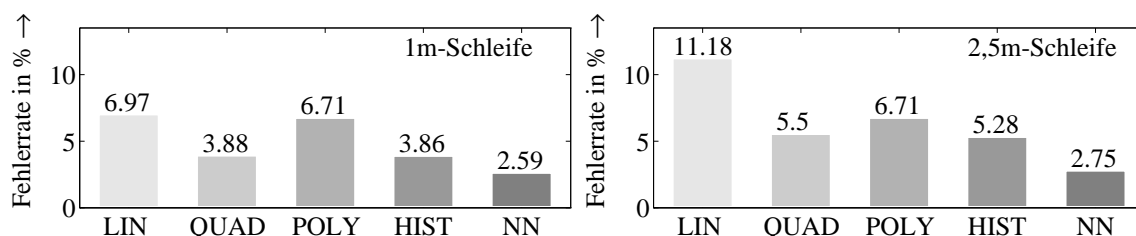


Bild 4.22: Fehlerraten bei der Fremdklassifikation mit 15 Merkmalen

4.8 Schätzung des Bayes-Fehlers

Die Entscheidungsregel der Bayes-Klassifikation ist mit Gleichung (4.5) gegeben. Der Klassifikator entscheidet sich beim Muster \mathbf{m} für die Hypothese H_i , falls die A-posteriori-Wahrscheinlichkeit $P(k_i|\mathbf{m})$ für alle $i = 1, \dots, K$ maximal ist. Ein Fehler tritt dann auf, wenn das Muster zu einer Klasse gehört, deren A-posteriori-Wahrscheinlichkeit nicht den größten Wert annimmt. Die Wahrscheinlichkeit hierfür kann durch die bedingte Bayessche Fehlerwahrscheinlichkeit

$$e(\mathbf{m}) = 1 - \max_k P(k_k|\mathbf{m}) \quad (4.34)$$

ausgedrückt werden. Der Bayes-Fehler ist der Erwartungswert von $e(\mathbf{m})$:

$$\varepsilon_B = E\{e(\mathbf{m})\} = \int_{\mathbf{m}} p(\mathbf{m}) (1 - \max_k P(k_k|\mathbf{m})) d\mathbf{m} \quad (4.35)$$

und kann mit der Bayes-Formel (4.6) auch als

$$\varepsilon_B = E\{e(\mathbf{m})\} = 1 - \int_{\mathbf{m}} \max_k (p_k p_{\mathbf{m}|k_k}(\mathbf{m}|k_k)) d\mathbf{m} \quad (4.36)$$

geschrieben werden. Der Fehler setzt sich aus den Überlappungsflächen der mit den A-priori-Wahrscheinlichkeiten gewichteten Likelihood-Funktionen zusammen, siehe Bild 4.23.

Der Bayes-Fehler ist wegen der Unkenntnis der Verteilungsfunktionen in der Regel analytisch nicht berechenbar. Deshalb werden in der Literatur Gütemaße genannt, die eine Abschätzung des Bayes-Fehlers erlauben.

Eine Eingrenzung des Bayes-Fehlers ist durch die Klassifikationsrate des Nächsten-Nachbar-Klassifikators nach [37] mit

$$\varepsilon_B \leq \varepsilon_{NN} \leq \left(2 - \frac{K\varepsilon_B}{K-1}\right) \varepsilon_B < 2\varepsilon_B \quad (4.37)$$

gegeben.

Für ein Zweiklassenproblem kann der maximale Bayes-Fehler nach [37] mit der *Chernoff-Grenze*

$$\varepsilon_C(s) = \int_{\mathbf{m}} p_{\mathbf{m}|k_1}^s(\mathbf{m}|k_1) p_{\mathbf{m}|k_2}^{1-s}(\mathbf{m}|k_2) d\mathbf{m} \quad \text{für } 0 \leq s \leq 1 \quad (4.38)$$

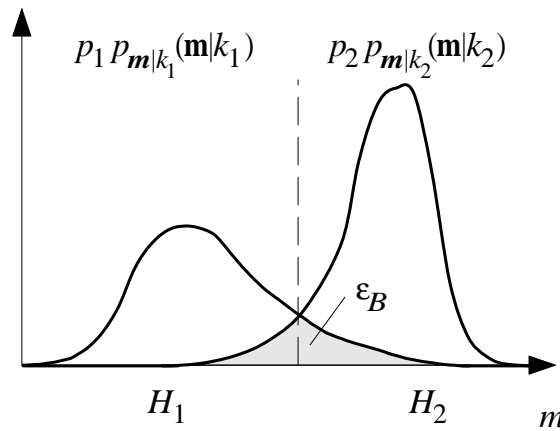


Bild 4.23: Fehlerrate des Bayes-Klassifikators

geschätzt werden:

$$\varepsilon_B \leq p_1^s p_2^{1-s} \varepsilon_C(s). \quad (4.39)$$

Bei Annahme einer Normalverteilung der Dichten kann das Integral in (4.38) geschlossen gelöst werden und es ergibt sich

$$\varepsilon_C(s) = e^{-\mu(s)} \quad (4.40)$$

mit

$$\begin{aligned} \mu(s) &= \frac{s(1-s)}{2} [\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2]^T [s\mathbf{R}_1 + (1-s)\mathbf{R}_2]^{-1} [\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2] \\ &+ \frac{1}{2} \ln \frac{|s\mathbf{R}_1 + (1-s)\mathbf{R}_2|}{|\mathbf{R}_1|^s |\mathbf{R}_2|^{1-s}}. \end{aligned} \quad (4.41)$$

Der Term $\mu(s)$ ist auch als *Chernoff-Abstand* bekannt. Mit dem optimalen Wert s_0 aus dem Intervall $[0, 1]$ kann eine obere Grenze des Bayes-Fehlers für ein Zweiklassenproblem mit

$$\varepsilon_B \leq p_1^{s_0} p_2^{1-s_0} \varepsilon_C(s_0) \quad (4.42)$$

bestimmt werden. Für ein Mehrklassenproblem müssen die einzelnen Chernoff-Grenzen aufsummiert werden.

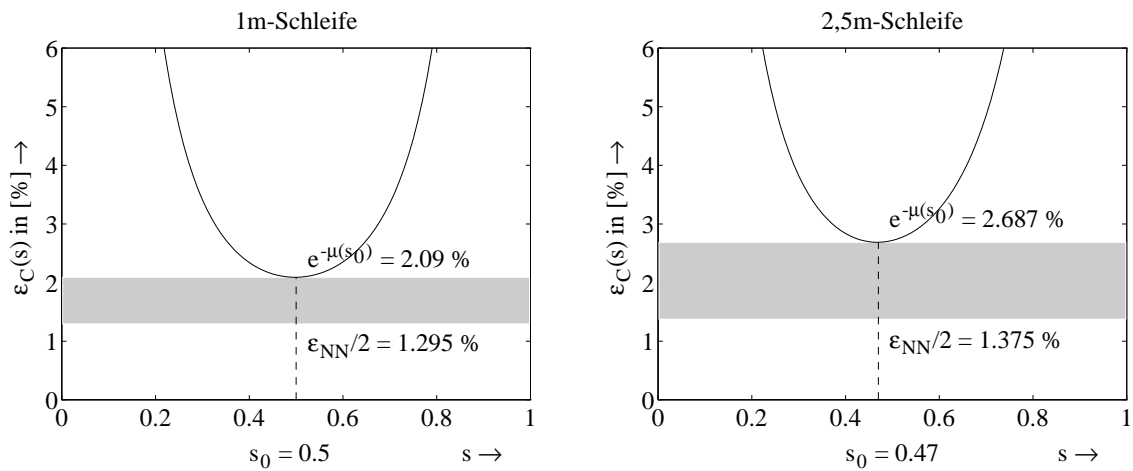


Bild 4.24: Abschätzung des Bayes-Fehlers

Der Bayes-Fehler kann demnach bei gleich angenommenen A-priori-Wahrscheinlichkeiten durch

$$\varepsilon_{NN}/2 \leq \varepsilon_B \leq \varepsilon_C(s_0) \quad (4.43)$$

eingegrenzt werden und liegt unter der Annahme annähernd normalverteilter Muster für die Klassifikation mit 1m-Schleifen im Intervall [1,30 %, 2,09 %] und für die Klassifikation mit 2,5m-Schleifen im Intervall [1,38 %, 2,69 %], siehe Bild 4.24.

5 Streckenbezogene Verkehrsdatenerfassung

Die Erfassung streckenbezogener Verkehrsdaten ist ein Forschungsgebiet, das vor allem in Europa seit Beginn der achtziger Jahre bearbeitet wird und dessen Ergebnisse heute in verschiedenen Verkehrsleitsystemen eingesetzt werden [11], [62], [47]. Die Verfahren basieren auf einer genauen Klassifizierung von Einzelfahrzeugen aus den gemessenen Ausgangssignalen von Induktionsschleifen, wodurch Fahrzeuge oder Fahrzeugkollektive identifiziert und am nächsten Meßquerschnitt wiedererkannt werden können. Ein Problem stellt sich hierbei in den Anforderungen an die Detektoren. Sie müssen auch bei ungünstigen Materialbedingungen einer Fahrbahn oder Schrägfahrten der Fahrzeuge noch genügend genaue und mit anderen Querschnitten vergleichbare Signalverläufe liefern. Dieses Problem kann durch geeignete Verfahren zur Vorverarbeitung und Merkmalsextraktion der Fahrzeugsignale reduziert werden. Gleichzeitig müssen die Merkmale im Vergleich zu den Merkmalen der übrigen Fahrzeuge eine eindeutige Wiedererkennung ermöglichen.

Zur Zuordnung von Ereignissen existieren verschiedene Ansätze, von denen die in [61] hergeleitete Korrelationsanalyse am häufigsten eingesetzt wird. In diesem Verfahren werden Fahrzeugkollektive wiedererkannt und aus der Verzögerungszeit der Detektion an zwei Meßquerschnitten wird auf die mittlere Reisezeit des Fahrzeugkollektivs im Streckenabschnitt geschlossen. Der Nachteil hierbei liegt allerdings darin, daß ein Fahrzeugkollektiv nach einer Durchmischung der Fahrzeuge im Streckenabschnitt durch Überholvorgänge nicht mehr eindeutig wiedererkannt werden kann. Zwei neue Ansätze, die mit einer Wiedererkennung von Individualfahrzeugen im Fahrzeugverbund arbeiten, sollen in dieser Arbeit präsentiert und diesem Verfahren gegenübergestellt werden.

Das erste Verfahren arbeitet mit Kostenfunktionen, die die gesamte Streckeninformation nutzen. Jedem Fahrzeug im Streckenabschnitt wer-

den Kosten zugeordnet, die sich aus den Unterschieden der Merkmalsvektoren an den Meßquerschnitten sowie den Unterschieden von prognostizierten und wahren Meßdaten der Fahrzeuge an den Induktionsschleifen ergeben. Die individuelle Zuordnung eines Fahrzeugs geschieht dann durch Minimierung der Kostenfunktionen. Das Ergebnis dieses Ansatzes führt zu einem weiteren Verfahren, das die statistischen Bindungen der Fahrzeuge im Streckenabschnitt berücksichtigt. Diese zeitlichen Abhängigkeiten können ähnlich wie in der Spracherkennung durch sogenannte *Hidden-Markov-Modelle* modelliert werden. Die Wiedererkennung der Fahrzeuge im Verbund erfolgt effizient durch die dynamische Programmierung mit dem Viterbi-Algorithmus.

Im folgenden soll zunächst die Struktur eines Mustererkennungssystems zur streckenbezogenen Verkehrsdatenerfassung aufgezeigt werden. Die Vorverarbeitung und Merkmalsextraktion der Fahrzeugsignale wird diskutiert und zusammen mit einer Entscheidungsfunktion anhand von Daten aus einer Lernstichprobe fixiert. Anschließend werden die Schätzverfahren zur streckenbezogenen Verkehrsdatenerfassung präsentiert und mit Simulationsergebnissen bewertet. Ein Vergleich bildet den Abschluß dieses Kapitels.

5.1 Mustererkennungssystem

Die Schätzung streckenbezogener Verkehrsdaten gelingt ähnlich wie die Klassifikation von Fahrzeugen in Kapitel 4 mit der Theorie der Mustererkennung. Im Gegensatz zu den Problemen der Fahrzeugklassifikation sind jetzt aber keine festen Objektklassen vorgegeben, sondern der Merkmalsvektor eines am stromabwärtigen Meßquerschnitt erfaßten Fahrzeugs muß einem Fahrzeug im Streckenabschnitt zugeordnet werden. Die Fahrzeuge im Streckenabschnitt bilden demnach die aktuellen Objektklassen und die statistischen Eigenschaften sind durch die am stromaufwärtigen Meßquerschnitt extrahierten Merkmalsvektoren festgelegt.

Bild 5.1 gibt einen Überblick über die Struktur eines Mustererkennungssystems zur streckenbezogenen Verkehrsdatenerfassung. Ausgehend

von einer Lernstichprobe von beiden Meßquerschnitten werden die Vorverarbeitung und Merkmalsextraktion festgelegt. In der Vorverarbeitung müssen sowohl die irrelevanten Eigenschaften der Meßaufnehmer als auch schleifenspezifische Abweichungen an den Meßquerschnitten beseitigt werden. Die Zielsetzung der Merkmalsextraktion bei der Fahrzeugklassifikation liegt darin, klassenweise möglichst ähnliche Signale zu erhalten und die Klassenzentren möglichst weit zu streuen. Für die Wiedererkennung hingegen sollen die Signale sich für alle Fahrzeuge und hierbei insbesondere für Fahrzeuge der gleichen Klasse möglichst gut unterscheiden.

Nach der Signalvorverarbeitung und der Merkmalsextraktion muß eine Metrik festgelegt werden, die ein Maß für den Unterschied verschiedener Merkmalsvektoren liefert. Bei der Wiedererkennung wird der Merkmalsvektor eines Fahrzeugs vom zweiten Meßquerschnitt mit allen Merkmalsvektoren der Fahrzeuge vom ersten Meßquerschnitt verglichen und die Entscheidung fällt auf das Fahrzeug mit dem geringsten Abstand. Die Einbeziehung anderer Kosten oder Wahrscheinlichkeiten ist in Bild 5.1 noch nicht enthalten.

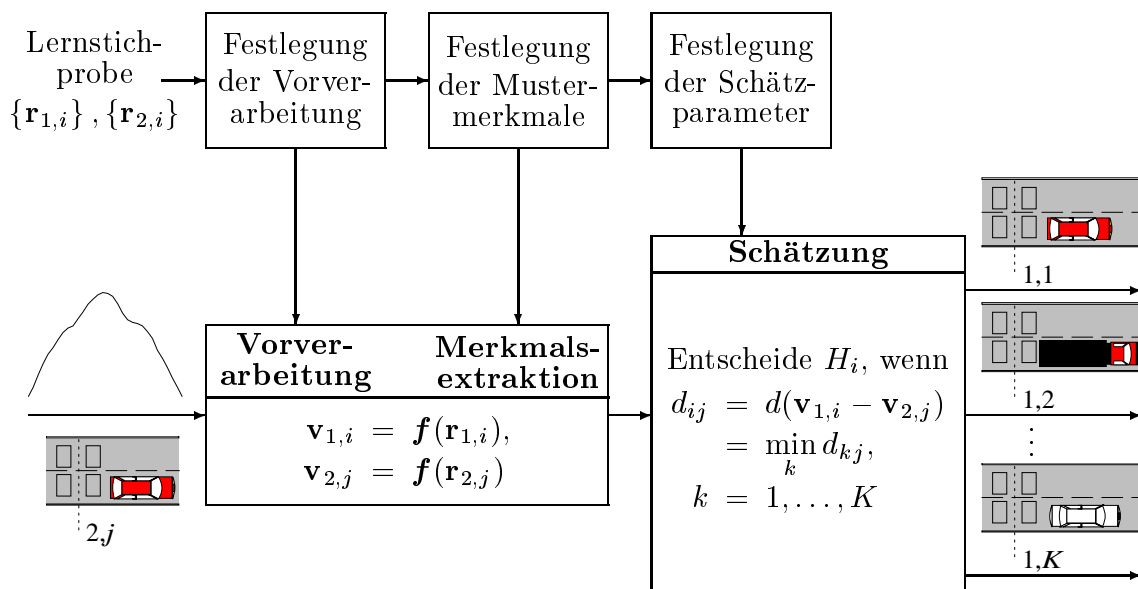


Bild 5.1: Struktur eines Mustererkennungssystems zur streckenbezogenen Verkehrsdatenerfassung

5.1.1 Signalvorverarbeitung und Merkmalsextraktion

Bild 5.2 gibt einen zusammenfassenden Überblick über die Methoden und die Signalvektoren der Vorverarbeitung und Merkmalsextraktion zur streckenbezogenen Verkehrsdatenerfassung.

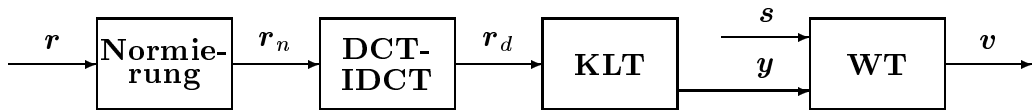


Bild 5.2: Signalvorverarbeitung und Merkmalsextraktion zur streckenbezogenen Verkehrsdatenerfassung

Zur Beschreibung der Schätzverfahren werden die Abtastwerte der Fahrzeugsignale in einem Intervall der Länge ℓT nach Gleichung (2.1) zu einem zeitabhängigen Vektor

$$\mathbf{r}(kT) = [r(kT), r([k+1]T), \dots, r([k+\ell-1]T)]^T \quad (5.1)$$

zusammengefaßt, wobei ℓ eine natürliche Zahl und abhängig von der Fahrzeuggeschwindigkeit und der Fahrzeuglänge ist.

Die Vektoren sollen nach ihrem Entstehungsort unterschieden werden. Signale vom stromaufwärtigen Meßquerschnitt werden analog zur Numerierung des Meßquerschnitts mit \mathbf{r}_1 bezeichnet. Signale vom stromabwärtigen Meßquerschnitt sollen mit \mathbf{r}_2 beschrieben werden.

Jedes Fahrzeugsignal stellt eine Realisation des stochastischen Prozesses \mathbf{r}_1 bzw. \mathbf{r}_2 dar. Zur Erzeugung der bei der streckenbezogenen Verkehrsdatenerfassung eingesetzten Transformationsmatrizen erfolgt eine Analyse einer Lernstichprobe bestehend aus den Realisationen $\mathbf{r}_{1,i}$ und $\mathbf{r}_{2,i}$, $i = 1, \dots, L$, wobei L die Anzahl der Elemente der Lernstichproben angibt.

Ein Fahrzeugsignal wird in einem ersten Schritt zum Ausgleich der unterschiedlichen Detektorempfindlichkeiten in der Signalamplitude normiert. Des weiteren erfolgt zur Elimination der Fahrzeuggeschwindigkeit

eine Abtastratenumwandlung durch Interpolation bzw. Dezimation auf eine feste Signallänge n , siehe Kapitel 2. Die entstehende Abtastperiode T_n ist dann lediglich von der Fahrzeuglänge abhängig.

Es ergeben sich zeitabhängige Signalvektoren

$$\mathbf{r}_n(kT_n) = [r_n(kT_n), r_n([k+1]T_n), \dots, r_n([k+n-1]T_n)]^T \quad (5.2)$$

und damit die stochastischen Prozesse $\mathbf{r}_{1,n}$ und $\mathbf{r}_{2,n}$.

Die schleifenspezifischen Abweichungen können durch eine Tiefpaßfilterung eliminiert werden. Die Filterung kann durch Anwendung der diskreten Cosinus-Transformation [1], [67] gleichzeitig zu einer Merkmalsextraktion eingesetzt werden. Die Transformation des Signals in den Frequenzbereich durch die DCT und die anschließende Rücktransformation des reduzierten Signals in den Zeitbereich durch die inverse DCT kann als

$$\mathbf{r}_{1,d} = \mathbf{K}\mathbf{K}^T \mathbf{r}_{1,n} = \mathbf{D} \mathbf{r}_{1,n} \quad \text{bzw.} \quad \mathbf{r}_{2,d} = \mathbf{K}\mathbf{K}^T \mathbf{r}_{2,n} = \mathbf{D} \mathbf{r}_{2,n} \quad (5.3)$$

geschrieben werden. Die Matrix $\mathbf{K} = [\mathbf{k}_1, \dots, \mathbf{k}_m]$ setzt sich aus den ersten m Basisvektoren der DCT zusammen [67].

Der nächste Schritt der Merkmalsextraktion liegt in der Karhunen-Loève-Transformation der gefilterten Signale. Zur Karhunen-Loève-Transformation wird die Kovarianzmatrix

$$\mathbf{R}_{dd} = E\{\mathbf{r}_d \mathbf{r}_d^T\} \quad (5.4)$$

des gemeinsamen Prozesses

$$\mathbf{r}_d = \frac{1}{2} (\mathbf{r}_{1,d} + \mathbf{r}_{2,d}) \quad (5.5)$$

gebildet. Die optimale Transformation erfolgt aus der Gleichung

$$\mathbf{R}_{dd} \mathbf{u}_j = \lambda_j \mathbf{u}_j, \quad j = 1, \dots, n. \quad (5.6)$$

Das entstehende Basissystem $\mathbf{U} = [\mathbf{u}_1, \dots, \mathbf{u}_k]$ beinhaltet nach Abschnitt 3.2 die zu den k größten Eigenwerten gehörenden Eigenvektoren.

Nach der Karhunen-Loève-Transformation liegen die Prozesse

$$\mathbf{y}_1 = \mathbf{U} \mathbf{r}_{1,d} \quad \text{und} \quad \mathbf{y}_2 = \mathbf{U} \mathbf{r}_{2,d} \quad (5.7)$$

vor.

Die zur Weiterverarbeitung eingesetzten Merkmalsvektoren setzen sich aus den Realisationen der Prozesse \mathbf{y}_1 und \mathbf{s}_1 bzw. \mathbf{y}_2 und \mathbf{s}_2 zusammen, wobei die Prozesse \mathbf{s}_1 und \mathbf{s}_2 die aus den Überfahrkurven selektierten Sondermerkmale beinhalten. Es ergeben sich damit die Prozesse

$$\mathbf{m}_1 = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{s}_1 \end{bmatrix} \quad \text{und} \quad \mathbf{m}_2 = \begin{bmatrix} \mathbf{y}_2 \\ \mathbf{s}_2 \end{bmatrix}. \quad (5.8)$$

Der letzte Schritt der Merkmalsextraktion liegt in einer Whitening-Transformation des Prozesses $\mathbf{m} = \mathbf{m}_1 - \mathbf{m}_2$, die für eine gleiche Varianz und damit auch Bewertung aller Koeffizienten des Prozesses sorgt, der aus der Differenz der Signale an den jeweiligen Detektoren entsteht. Nach Abschnitt 3.3 kann die Transformationsmatrix \mathbf{W} aus der Gleichung

$$\mathbf{R}_{vv} = E\{\mathbf{v}\mathbf{v}^T\} = \mathbf{W} E\{\mathbf{m}\mathbf{m}^T\} \mathbf{W}^T = \mathbf{W} \mathbf{R}_{mm} \mathbf{W}^T = \mathbf{I} \quad (5.9)$$

ermittelt werden. Die Ausgangsprozesse ergeben sich dann zu

$$\mathbf{v}_1 = \mathbf{W} \mathbf{m}_1 \quad \text{und} \quad \mathbf{v}_2 = \mathbf{W} \mathbf{m}_2. \quad (5.10)$$

Werden keine Sondermerkmale eingesetzt, kann die Signalvorverarbeitung und Merkmalsextraktion durch Zusammenfassen der oben beschriebenen Transformationen auch in einem Schritt erfolgen.

Die Realisationen der Prozesse \mathbf{v}_1 und \mathbf{v}_2 werden mit $\mathbf{v}_{1,i}$ und $\mathbf{v}_{2,i}$ bezeichnet.

5.1.2 Entscheidungsfunktion

Der Vergleich eines Merkmalsvektors $\mathbf{v}_{2,j}$ vom zweiten Meßquerschnitt mit den Merkmalsvektoren $\mathbf{v}_{1,i}$, $i = 1, \dots, K$, einer Gruppe von K Fahrzeugen vom ersten Meßquerschnitt erfolgt mit der Metrik

$$d_{ij}(\mathbf{v}_{1,i}, \mathbf{v}_{2,j}) = \left[\sum_{k=1}^N |v_{1,i,k} - v_{2,j,k}|^p \right]^{\frac{1}{p}}, \quad 0 < p < \infty, \quad \mathbf{v}_{1,i}, \mathbf{v}_{2,j} \in \ell_p(1, N), \quad (5.11)$$

wobei der Parameter p im angegebenen Intervall beliebig gewählt werden kann.

Ein Beispiel ist die euklidische Metrik

$$d_{ij}(\mathbf{v}_{1,i}, \mathbf{v}_{2,j}) = \left[\sum_{k=1}^N |v_{1,i,k} - v_{2,j,k}|^2 \right]^{\frac{1}{2}}, \quad \mathbf{v}_{1,i}, \mathbf{v}_{2,j} \in \ell_2(1, N). \quad (5.12)$$

Quadriert man Gleichung (5.12), dann ergibt sich

$$\begin{aligned} d_{ij}^2(\mathbf{v}_{1,i}, \mathbf{v}_{2,j}) &= \|\mathbf{v}_{1,i} - \mathbf{v}_{2,j}\|^2 \\ &= [\mathbf{v}_{1,i} - \mathbf{v}_{2,j}]^T [\mathbf{v}_{1,i} - \mathbf{v}_{2,j}] \\ &= \mathbf{v}_{1,i}^T \mathbf{v}_{1,i} - 2\mathbf{v}_{1,i}^T \mathbf{v}_{2,j} + \mathbf{v}_{2,j}^T \mathbf{v}_{2,j}. \end{aligned} \quad (5.13)$$

als leicht berechenbarer Wert für den Abstand zweier Fahrzeugsignale.

Die Zuordnung eines Fahrzeugs an der zweiten Meßstelle mit dem Merkmalsvektor $\mathbf{v}_{2,j}$ zu einem Fahrzeug an der ersten Meßstelle mit dem Merkmalsvektor $\mathbf{v}_{1,i}$ kann nun für dasjenige Fahrzeug erfolgen, dessen Metrik $d(\mathbf{v}_{1,i}, \mathbf{v}_{2,j})$ minimal ist. Man entscheidet sich für die Hypothese H_i , d.h. der Merkmalsvektor $\mathbf{v}_{1,i}$ stammt von dem gleichen Fahrzeug wie der Merkmalsvektor $\mathbf{v}_{2,j}$, wenn

$$d_{ij}^2 = \min_{1 \leq k \leq K} \left([\mathbf{v}_{1,k} - \mathbf{v}_{2,j}]^T [\mathbf{v}_{1,k} - \mathbf{v}_{2,j}] \right). \quad (5.14)$$

5.1.3 Parameterfestlegung

Freie Parameter aus der Vorverarbeitung und Merkmalsextraktion ergeben sich aus der Länge der normierten Signale (n), der Anzahl extrahierter Merkmale aus der DCT-IDCT (m) und der KLT (k), sowie aus der gewählten Metrik (p). Untersuchungen zur Bestimmung der günstigsten Parameter wurden in [76] durchgeführt. Als Datengrundlage standen Fahrzeugsignale von 1m- und 2,5m-Schleifen zur Verfügung, wobei jeweils ca. 2000 Fahrzeugkurven von einer 1m- und einer 2,5m-Doppelschleife und ca. 2000 Fahrzeugkurven von zwei verschiedenen Meßstellen mit 2,5m-Schleifen aufgenommen wurden. Zur Ermittlung der günstigsten Parameter werden die Wiedererkennungsraten von Fahrzeugen aus einer bestimmten Fahrzeuggruppe herangezogen.

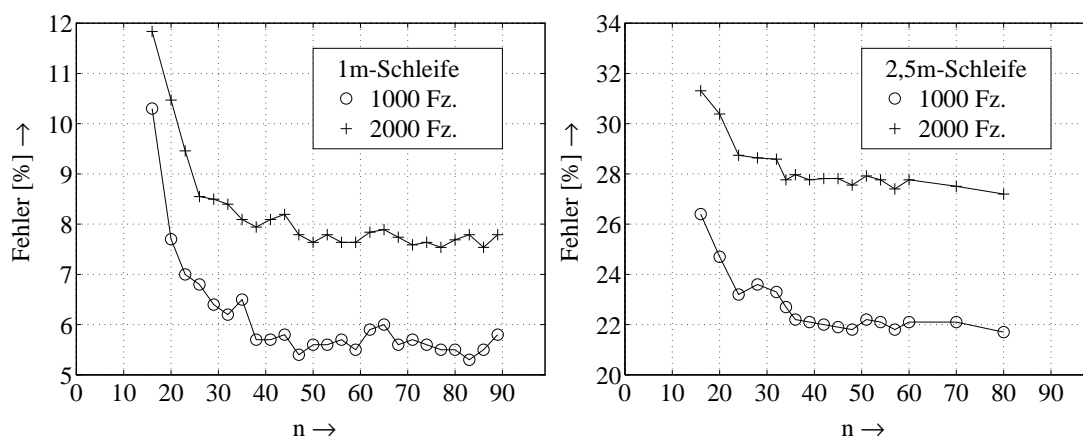


Bild 5.3: Abhängigkeit der Fehlerrate von der Anzahl normierter Werte (n) bei 1000 und 2000 Vergleichsfahrzeugen

Normierung

Bild 5.3 zeigt die Fehlerraten der Wiedererkennung in Abhängigkeit der Signallänge n aus der Längennormierung für Signale von 1m- und 2,5m-Schleifen. Für kleine Werte von n ($n < 40$) erhöht sich die Fehlerrate, da die Auflösung der Signale zur Fahrzeugunterscheidung im Gegensatz zur Fahrzeugklassenunterscheidung verloren geht (hier wurde $n = 32$ gewählt). Es bietet sich daher an, die Längennormierung auf $n = 64$ Werte festzulegen, da diese normierten Signale durch Dezimation um den Faktor zwei gleichzeitig zur Fahrzeugklassifikation eingesetzt werden können.

Merkmalsextraktion

Die Merkmalsextraktion mit Hilfe der diskreten Cosinus-Transformation und der Karhunen-Loève-Transformation führt auf Signale der Länge m bzw. k . Da nach der Karhunen-Loève-Transformation maximal m Eigenwerte ungleich Null sind und beide Transformationen in einem Schritt berechnet werden sollten, ist es sinnvoll, beide Parameter zusammen zu betrachten. Bild 5.4 zeigt die Fehlerraten der Wiedererkennung in Abhängigkeit von m und $k < m$ für Signale von 1m- und 2,5m-Schleifen. Als günstigste Wahl der Parameter ergibt sich für beide Schleifentypen das Intervall $12 < k \leq m < 18$.

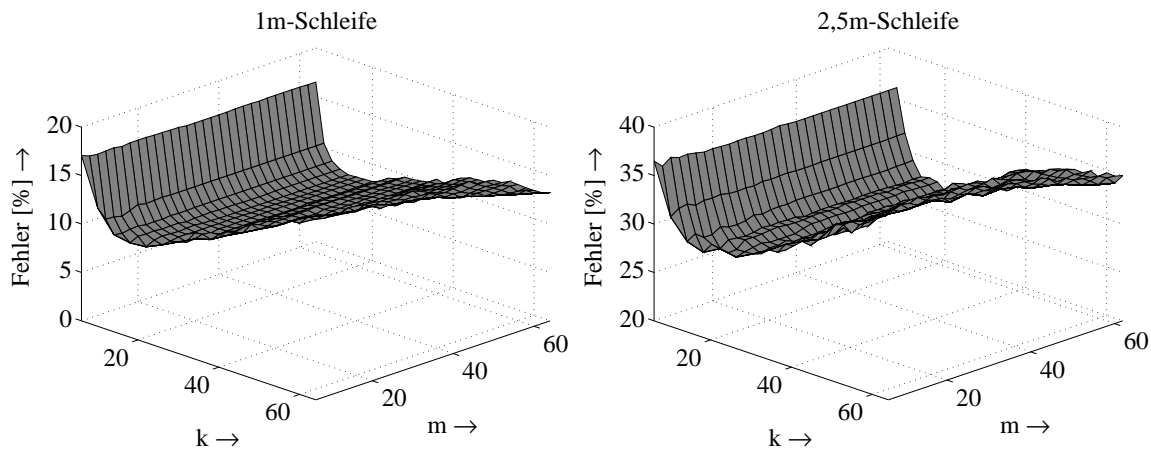


Bild 5.4: Abhängigkeit der Fehlerrate von der Anzahl der Koeffizienten bei der DCT-IDCT (m) und der KLT (k) bei ca. 2000 Vergleichsfahrzeugen

Aus den Untersuchungen in [76] ergibt sich, daß im Gegensatz zur Fahrzeugklassifikation aus der Merkmalsselektion lediglich die Fahrzeuglänge als Sondermerkmal miteinbezogen werden muß. Somit besitzen die resultierenden Merkmalsvektoren die Dimension $N = k + 1$.

Entscheidungsfunktion

Führt man die Wiedererkennung der Fahrzeuge nach der Entscheidungsregel (5.11) durch, so bietet der in der Metrik auftretende Parameter p einen Freiheitsgrad. In den durchgeführten Untersuchungen wurde der Parameter p im Bereich $0 < p < 5$ variiert. In Bild 5.5 sind die Fehlerraten bezüglich des Parameters p dargestellt. Zur Wiedererkennung wurden dabei ca. 2000 Überfahrkurven von 1m-Schleifensignalen eingesetzt.

Für diese Konfiguration liegt der optimale Wert bei $p \approx 1.8$, die Fehleraten sind allerdings im Bereich $1 < p < 3$ nahezu gleich. Somit erscheint eine Wahl von $p = 2$ als günstig, da der Aufwand zur Berechnung der Metrik nach Gleichung (5.13) am geringsten ist.

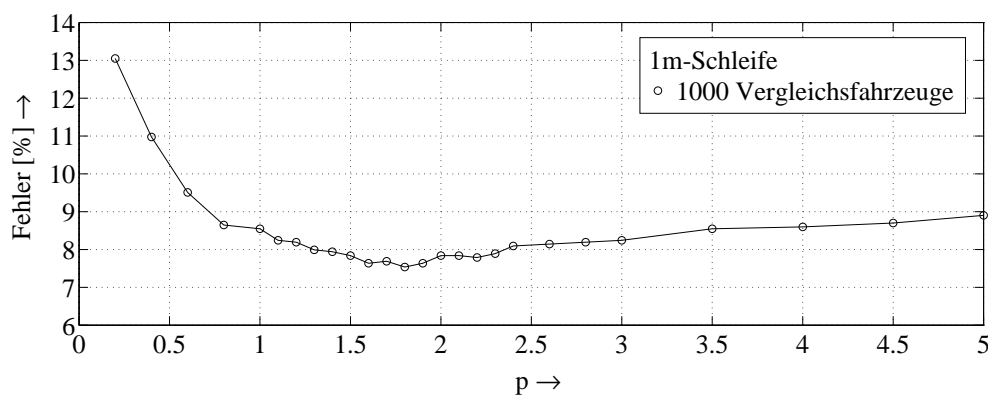


Bild 5.5: Prozentuale Fehlerrate bezüglich des Parameters p der Metrik

5.1.4 Voruntersuchungen

Die Entscheidungsregel (5.14) ordnet Fahrzeuge lediglich nach ihren extrahierten Merkmalsvektoren zu. Die folgenden Voruntersuchungen sollen zeigen, wie groß die Fehlerrate dieser Zuordnung in Abhängigkeit der Anzahl der zu vergleichenden Fahrzeuge ist. Die Anzahl der Fahrzeuge hängt im Verkehrsfluß von der Länge des Streckenabschnitts, der Anzahl der Fahrspuren und der Verkehrsdichte ab. Zudem soll diese Untersuchung Aufschluß über die Schleifeneigenschaften zur Fahrzeugwiedererkennung geben.

Die aus verschiedenen Durchläufen ermittelten prozentualen Fehlerraten sind gemittelt in Bild 5.6 für die vorhandenen Stichproben dargestellt. Die Fehlerrate steigt mit zunehmender Vergleichszahl an und strebt gegen einen Grenzwert, der vom Schleifentyp und von den Meßstellen abhängig ist.

Aus Bild 5.6a geht hervor, daß die Wiedererkennung von Fahrzeugen mit Signalen von 1m-Schleifen deutlich bessere Ergebnisse liefert als mit Signalen von 2,5m-Schleifen. Die Fehlerrate ist für 2,5m-Schleifen je nach Anzahl der Vergleichsfahrzeuge um den Faktor drei bis vier größer.

Dieses Ergebnis stimmt mit theoretischen Untersuchungen von BÖKER überein, der in [10] ein informationstheoretisches Gütemaß für Merk-

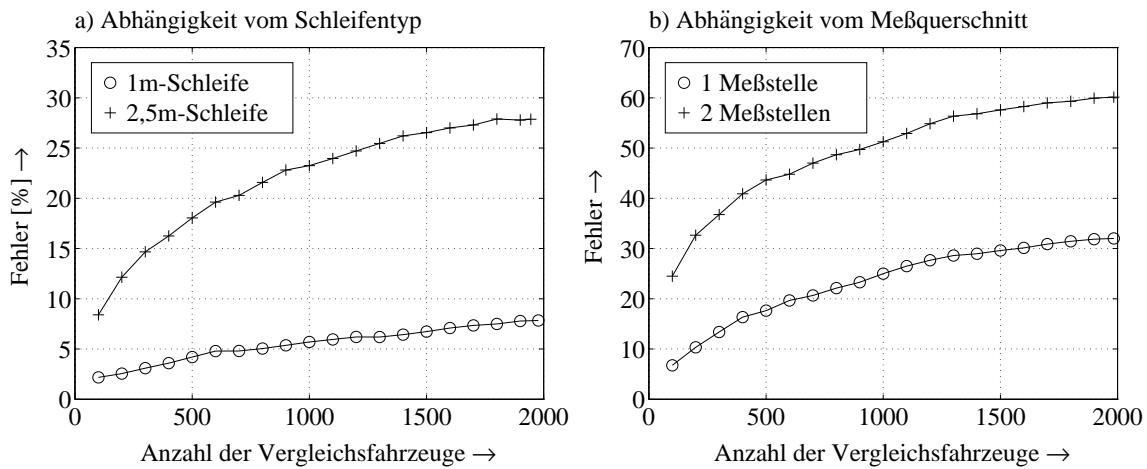


Bild 5.6: Prozentuale Fehlerrate in Abhängigkeit der Anzahl der Vergleichsfahrzeuge: a) 1m- und 2,5m-Doppelschleifen, b) 2,5m-Doppelschleife und 2,5m-Schleifen von verschiedenen Meßquerschnitten

malsvektoren in Abhängigkeit der geometrischen Abmessungen der verwendeten Induktionsschleifen definiert. Die in den Merkmalen enthaltene *Transinformation** ist für Signale von 1m-Schleifen am größten, da dieser Schleifentyp eine bestmögliche Kombination aus der Auflösung und der Reproduzierbarkeit eines Fahrzeugsignals bietet.

Für Fahrzeugsignale, die mit Doppelschleifen an einem Meßquerschnitt aufgenommen wurden, ergibt sich eine doppelt so gute Wiedererkennungsrate, wie bei Fahrzeugsignalen, die von zwei verschiedenen Meßquerschnitten aufgenommen wurden (Bild 5.6b). Die Reproduzierbarkeit der Fahrzeugsignale an Doppelschleifen kann als optimal angesehen werden, da die geometrischen Schleifenabmessungen nahezu identisch sind, das Fahrverhalten über die Schleifen gleich ist und die Fahrbahneigenschaften ebenfalls identisch sind. Die mit Doppelschleifen erzielbaren Wiedererkennungsraten können allerdings nur als obere Schranke angesehen werden.

Da von 1m-Schleifen keine Messungen an verschiedenen Querschnitten

* Als *Transinformation* wird der durch das Eingangssignal verursachte im Ausgangssignal enthaltene Informationsanteil bezeichnet

zur Verfügung standen, kann auch das Ergebnis aus den Doppelschleifenmessungen nur als minimale Fehlerrate aus der Fahrzeugwiedererkennung interpretiert werden. Da die Zusammensetzung des Verkehrs für alle drei Stichproben ähnlich ist, ist auch bei 1m-Schleifen eine doppelt so hohe Fehlerrate zu erwarten, wenn die Fahrzeugsignale von verschiedenen Meßquerschnitten stammen.

Es zeigt sich, daß für Merkmalsvektoren von 1m-Schleifen allein aus der Entscheidungsregel (5.14) schon eine recht gute Wiedererkennungsrate erreicht werden kann, da zur Bestimmung der streckenbezogenen Verkehrsdaten über ein Zeitintervall oder eine Fahrzeuggruppe gemittelt werden muß. Die Wiedererkennungsraten bei 2,5m-Schleifen und einer großen Anzahl von Fahrzeugen im Streckenabschnitt reichen allerdings zur Bestimmung der streckenbezogenen Verkehrsdaten nicht aus. Deshalb müssen weitere Informationen aus dem Verkehrsfluß genutzt werden, um die streckenbezogenen Verkehrsdaten zu schätzen. Derartige Schätzverfahren sollen in den nächsten drei Abschnitten dargestellt werden.

5.2 Schätzverfahren mit Fahrzeugfolgenkorrelation

Das Schätzverfahren mit Fahrzeugfolgenkorrelationen soll als Referenzverfahren herangezogen werden. Dieses Verfahren wird in den letzten Jahren erfolgreich eingesetzt [62], [47], und ist in [9] und [61] ausführlich beschrieben. In dieser Arbeit sollen nur die wichtigsten Schritte des Schätzalgorithmus vorgestellt werden. Nach der Herleitung der theoretischen Ansätze erfolgt eine Erprobung des Verfahrens für wahre Meßdaten und simulierte Verkehrszustände.

5.2.1 Der Schätzalgorithmus

Zielsetzung des Verfahrens ist die Beschreibung des aktuellen Verkehrszustandes in einem Straßenabschnitt. Die zur Beschreibung des Zustands

eingesetzten streckenbezogenen Größen werden durch die Wiedererkennung von Fahrzeugkollektiven gewonnen, die an einem stromaufwärtigen Meßquerschnitt in den Streckenabschnitt eingefahren sind und den Streckenabschnitt an einem stromabwärtigen Meßquerschnitt wieder verlassen. Der Streckenabschnitt wird im folgenden als System betrachtet, die Fahrzeuge seien Folgen von Objekten, die das System durchlaufen. Ein Fahrzeug i wird durch seine Merkmalsvektoren $\mathbf{v}_{1,i}$ und $\mathbf{v}_{2,i}$ repräsentiert, die an den beiden Meßquerschnitten extrahiert werden.

Korrelationsanalyse

Zur Bestimmung der streckenbezogenen Verkehrsdaten werden Gruppen von Objekten mit einem Korrelationsverfahren zugeordnet. Hierzu werden Folgen von Objekten $f(s_1, t)$ und $f(s_2, t)$ definiert, die an den Meßquerschnitten zum Zeitpunkt t beobachtet werden. Zu einem festen Zeitpunkt wird am Meßquerschnitt 1 eine Gruppe von L Objekten herausgegriffen, und diese Folge wird fortlaufend mit den den Meßquerschnitt 2 passierenden Objekten verglichen. Die Reihenfolge der Objekte innerhalb der Gruppe sei vorerst unverändert (keine Vermischung der Fahrzeuge).

Die Folge $f(s_2, t)$ wird an der Folge $f(s_1, t)$ der Länge L vorbeigeschoben. Bei jedem Verschiebungsschritt n werden im Bereich der Gruppe Konjunktionen $q_{i,j}(n)$ ($i = 1, \dots, L$, $j = n, \dots, n + L - 1$) aufgestellt, wobei

$$q_{i,j}(n) = \begin{cases} 1 & \text{bei einer Übereinstimmung,} \\ 0 & \text{bei keiner Übereinstimmung} \end{cases} \quad (5.15)$$

der Fahrzeuge i und j ist. Aus der Summe der L Konjunktionen ergibt sich für jeden Verschiebungsschritt eine sogenannte Korrelationsfunktion von Objektfolgen

$$r(n) = \frac{1}{L} \sum_{i=1}^L q_{i,j}(n), \quad (5.16)$$

die proportional zur Anzahl der zugeordneten Fahrzeuge ist. Bild 5.7 zeigt schematisch die Vorgehensweise bei diesem Schätzverfahren.

Bild 5.8a zeigt ein Beispiel einer Kreuzkorrelationsfunktion $r(n)$. Hierzu werden die aus den Fahrzeugsignalen extrahierten Merkmalsvektoren

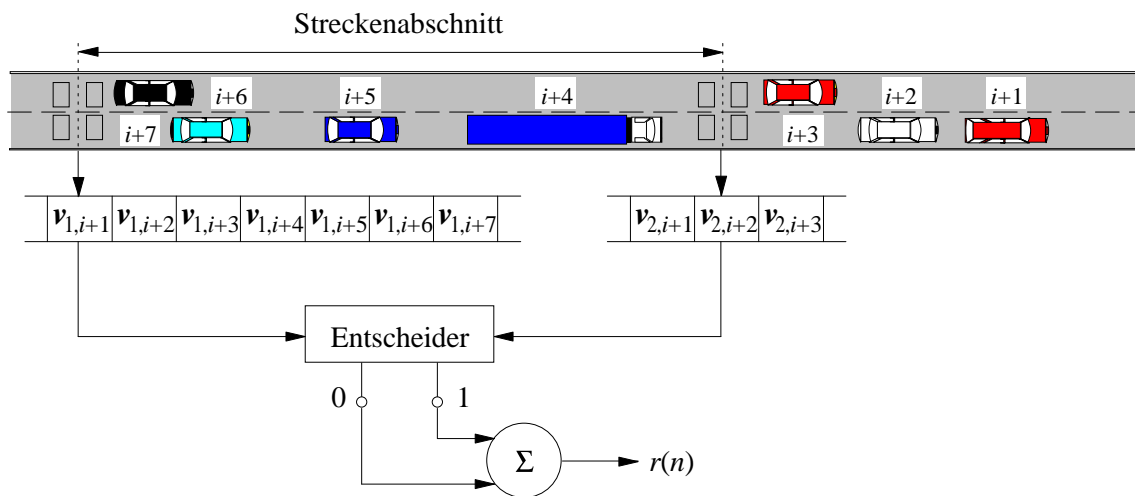


Bild 5.7: Beispiel der Fahrzeugfolgenkorrelation zur streckenbezogenen Verkehrsdatenerfassung

von zwei in geringem Abstand auseinander liegenden Meßstellen miteinander verglichen. Da keine Überholvorgänge in dem kurzen Streckenabschnitt stattfinden, erfolgt eine Korrelation ohne Vermischung. Deutlich erkennbar ist das Maximum der Korrelationsfunktion $r(n)$ bei $n = N$. Neben dem ausgeprägten Maximum ist ein Grundrauschen anhand zufälliger Übereinstimmungen von Fahrzeugen erkennbar.

In realen Verkehrssituationen führen Vermischungsvorgänge zu einer erheblichen Verminderung des theoretisch erreichbaren Korrelationsmaximums $r(n) = 1$. Als Beispiel zeigt Bild 5.8b die Fahrzeugfolgenkorrelationsfunktion $r(n)$ zu einem festen Zeitpunkt für einen simulierten Verkehrsablauf in einem 2,5 km langen Streckenabschnitt. Bedingt durch Zufallsübereinstimmungen und durch Vermischung ist ein eindeutiges Maximum der Funktion $r(n)$ nicht mehr erkennbar und die Position N für die Wiedererkennung einer Fahrzeuggruppe ist aus $r(n)$ nicht mehr bestimmbar. Deshalb wird eine Glättung der Korrelationsfunktion $r(n)$ durch Faltung mit einer Gaußfunktion

$$h(n) = \frac{1}{\sqrt{2\pi}\sigma} \cdot \exp\left(-\frac{n^2}{2\sigma^2}\right) \quad (5.17)$$

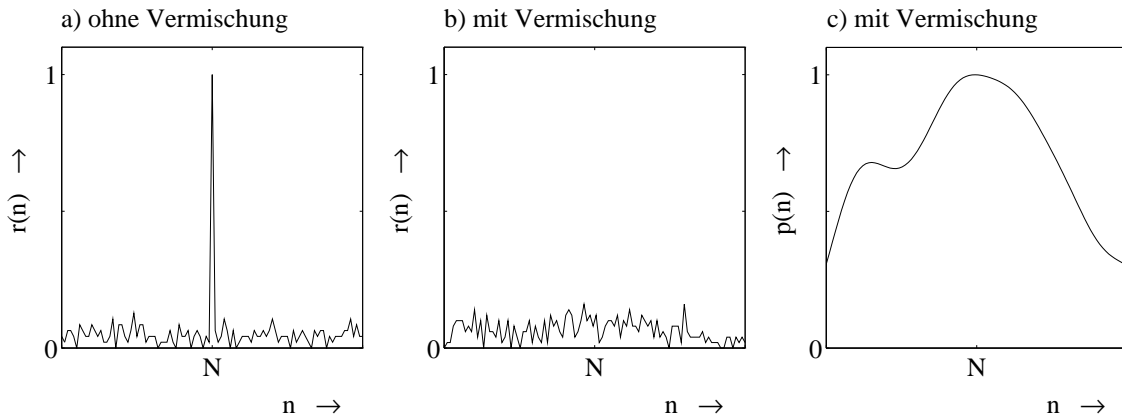


Bild 5.8: Beispiel der Fahrzeugfolgenkorrelationen: a) $r(n)$ ohne Vermischung, b) $r(n)$ mit Vermischung, c) $p(n)$ mit Vermischung

durchgeführt.

Damit ergibt sich eine tiefpaßgefilterte Korrelationsfunktion

$$p(n) = r(n) * h(n). \quad (5.18)$$

Bild 5.8c gibt die zu $r(n)$ gehörende geglättete und normierte Korrelationsfunktion $p(n)$ als Beispiel wieder. Die Wiedererkennung der Fahrzeuggruppe erfolgt nun näherungsweise durch Bestimmung des Maximums der Funktion $p(n)$:

$$N = n \Big|_{p(n)=\max}. \quad (5.19)$$

Streckenbezogene Verkehrsdaten

Die mittlere Reisezeit des Kollektivs ergibt sich aus der Differenz der gemittelten Detektionszeiten der Fahrzeuge eines Kollektivs an den beiden Meßquerschnitten zu

$$t_r = \frac{1}{L} \left(\sum_{i=N}^{L+N-1} t_{2,i} [f(s_2)] - \sum_{i=1}^L t_{1,i} [f(s_1)] \right). \quad (5.20)$$

Die Verkehrsdichte und die mittlere Geschwindigkeit des Kollektivs können mit der Streckenlänge s als

$$c = \frac{N}{s} \quad \text{und} \quad v = \frac{s}{t_r} \quad (5.21)$$

angegeben werden.

Entscheidungsregel

Zur Berechnung der Fahrzeugfolgenkorrelation ist die Bildung von Konjunktionen notwendig. Es muß deshalb eine Entscheidungsregel gefunden werden, die die Übereinstimmung bzw. Nicht-Übereinstimmung zweier Merkmalsvektoren $\mathbf{v}_{1,i}$ und $\mathbf{v}_{2,j}$ festlegt. In Abweichung zu [61] soll als Entscheidungsfunktion statt der City-Block-Metrik die euklidische Metrik benutzt werden, da sie nach Abschnitt 5.1.3 die besten Ergebnisse liefert. Die mit der Metrik (5.13) für zwei Merkmalsvektoren $\mathbf{v}_{1,i}$ und $\mathbf{v}_{2,j}$ ermittelten Abstände $d_{i,j}^2$ werden mit einer Entscheidungsschwelle d verglichen, so daß folgende Zuordnung erfolgen kann:

$$\begin{aligned} d_{i,j}^2 &\leq d && \text{Fahrzeuge sind identisch,} \\ d_{i,j}^2 &> d && \text{Fahrzeuge sind nicht identisch.} \end{aligned} \quad (5.22)$$

Die Bestimmung der optimalen Schwelle $d = d_{opt}$ erfolgt durch Minimierung des Zuordnungsfehlers. Unterscheidet man zwischen den Hypothesen H_1 (Fahrzeuge sind identisch) und H_2 (Fahrzeuge sind nicht identisch), so sind folgende von der Entscheidungsschwelle d abhängigen Fehlentscheidungen möglich:

- Fehler 1. Art – $f_1(d)$: Entscheidung für H_2 , wenn H_1 zutrifft.
- Fehler 2. Art – $f_2(d)$: Entscheidung für H_1 , wenn H_2 zutrifft.

Die Fehlerwahrscheinlichkeit $f(d)$ ergibt sich damit zu

$$f(d) = f_1(d) + f_2(d). \quad (5.23)$$

Durch Minimierung der Fehlerfunktion $f(d)$ ergibt sich die optimale Schwelle $d = d_{opt}$:

$$f(d)|_{d=d_{opt}} = \min. \quad (5.24)$$

Optimale Schwelle

Die Festlegung der optimalen Entscheidungsschwelle erfolgt zunächst für Meßreihen ohne Vermischung. Hierzu werden die Merkmalsvektoren aus den Fahrzeugsignalen einer Doppelschleife extrahiert. Allerdings kann dann die Fahrzeuglänge als Merkmal nicht berücksichtigt werden. Durch Variation des Parameters d können die Fehlerkurven $f_1(d)$ und $f_2(d)$ bestimmt werden. Bild 5.9a zeigt die Fehlerfunktionen $f_1(d)$, $f_2(d)$ und $f(d)$ für Signale von 1m-Schleifen. Hierbei wurden ca. 2000 Fahrzeuge miteinander verglichen, die Dimension der Merkmalsvektoren beträgt $N = 14$. Die optimale Schwelle ergibt sich zu $d_{opt} = 33$ bei einem Fehler von $f(d_{opt}) = 12\%$.

Man kann zeigen, daß die optimale Entscheidungsschwelle relativ unabhängig von der Vermischung des Verkehrs ist. In Bild 5.9b sind die entsprechenden Fehlerfunktionen für Signale von 1m-Schleifen bei einer relativ hohen Vermischung aufgetragen. Die Vermischung ergibt sich in diesem Fall aus der Simulation eines Verkehrsablaufs mit einem mikroskopischen Verkehrsmodell. Man erkennt, daß die relativ hohe Fehlerrate durch eine Glättung der Korrelationsfunktion verringert werden muß.

Variiert man die Anzahl der Merkmale, die zur Wiedererkennung eines Fahrzeugkollektivs notwendig sind, so ergibt sich $N \approx 6$ für Fahrzeugsignale von der 1m-Schleife und $N \approx 14$ für Fahrzeugsignale von 2,5m-Schleifen.

Zweidimensionale Fahrzeugfolgenkorrelation

Die Berechnung der Fahrzeugfolgenkorrelation erfolgt zur Bestimmung streckenbezogener Verkehrsdaten in einem festen Zeit- oder Fahrzeugraster. Somit ergeben sich zweidimensionale Korrelationsfunktionen $r(n, m)$ bzw. $p(n, m)$, die in Form eines Korrelationsgebirges aufgetragen werden können. Der Parameter n ist dabei im Fahrzeugraster diskretisiert. Die Diskretisierung von m erfolgt in festen Zeitabschnitten mT oder in Fahrzeuginkrementierungen Δm .

Zur Berechnung der Fahrzeugfolgenkorrelation

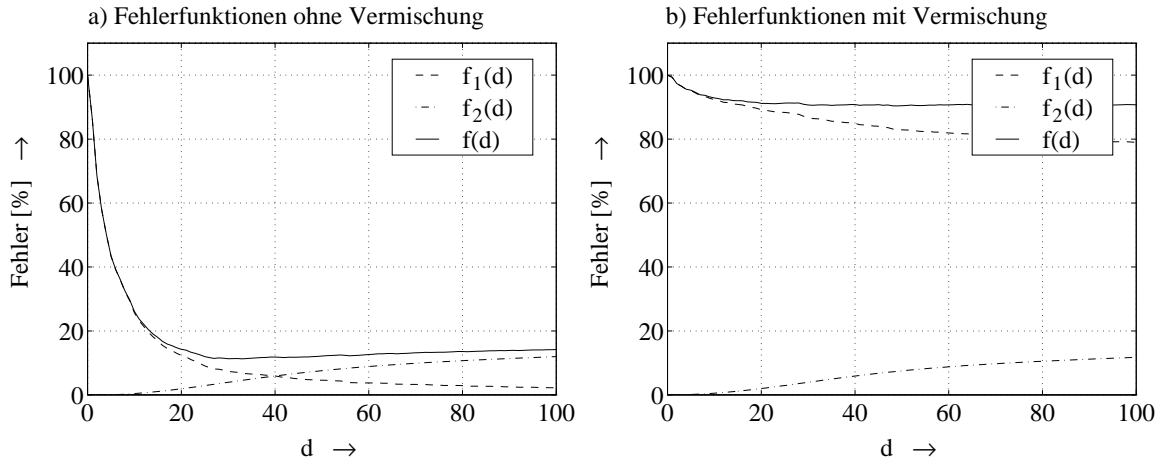


Bild 5.9: Fehlerfunktionen zur Bestimmung der optimalen Schwelle d_{opt} ohne und mit Vermischung

$$r(n, m) = \frac{1}{L} \sum_{i=1}^L q_{i,j}(n, m) \quad (5.25)$$

über eine Fahrzeuggruppe der Länge L wird die Konjunktion $q_{i,j}(n, m)$ beim Vergleich zweier Fahrzeuge folgendermaßen bestimmt:

$$\begin{aligned} q_{i,j}(n, m) &= 1 \quad \text{für } d_{i,j} \leq d_{opt} \rightarrow \text{Übereinstimmung,} \\ q_{i,j}(n, m) &= 0 \quad \text{für } d_{i,j} > d_{opt} \rightarrow \text{Keine Übereinstimmung.} \end{aligned} \quad (5.26)$$

Die Entscheidungsschwelle d_{opt} wird nach Gleichung (5.24) bestimmt. Bild 5.10a zeigt beispielhaft eine Korrelationsfunktion $r(n, m)$ und ihr Höhenprofil für die optimale Entscheidungsschwelle d_{opt} bei der Doppelschleifenmessung mit 1m-Schleifen. Es gibt eine geringe Anzahl an Zufallsübereinstimmungen und eine fast eindeutige Detektion von identischen Fahrzeugen.

Die Filterung der Korrelationsfunktion $r(n, m)$ bei Vermischung geschieht durch Faltung mit einem zweidimensionalen Filter $h(n, m)$:

$$p(n, m) = r(n, m) * h(n, m), \quad (5.27)$$

wobei

$$h(n, m) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{n^2}{2\sigma^2}\right) \cdot \exp(-\alpha m) \cdot \epsilon(m), \quad \alpha \geq 0, \quad (5.28)$$

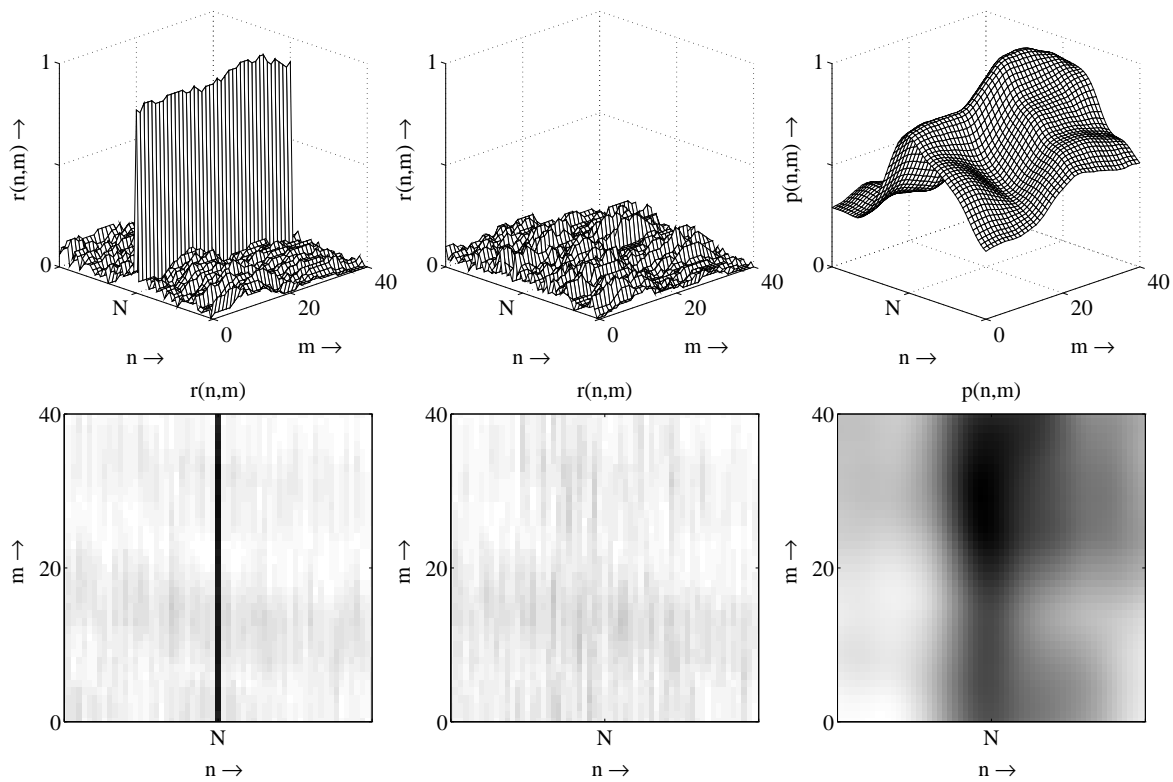


Bild 5.10: Fahrzeugfolgenkorrelation für Messungen mit der 1m-Schleife: a) $r(n, m)$ ohne Vermischung, b) $r(n, m)$ mit Vermischung, c) $p(n, m)$ mit Vermischung

gewählt wird. Damit erfolgt eine Tiefpaßfilterung der Korrelationsfunktion mit der Gaußfunktion zur Berücksichtigung der Vermischung sowie eine zeitliche gleitende Mittelwertbildung.

Die Verschiebungsposition $N(m)$ für die Wiederentdeckung der Fahrzeuggruppe ergibt sich aus der Bestimmung des Maximums der Korrelationsfunktion

$$N(m) = n(m) \Big|_{p(n,m)=\max_n} \quad (5.29)$$

Die geglättete Fahrzeugfolgenkorrelation $p(n, m)$ soll hierzu beispielhaft bei Vermischung dargestellt werden. Dazu wurden die dem Bild 5.10a zugrunde liegenden Daten wiederum anhand eines simulierten Verkehrsablaufs vermischt. Bild 5.10b zeigt die Fahrzeugfolgenkorrelation $r(n, m)$

und in Bild 5.10c ist die geglättete Korrelationsfunktion $p(n, m)$ aufgetragen. Im Höhenprofil zeigt sich, daß die Lage des Korrelationsmaximums für diesen Vermischungsgrad im Bereich $n = N$ noch sehr ausgeprägt ist und damit eine Wiedererkennung eines Fahrzeugkollektivs eindeutig durchgeführt werden könnte.

Die Anwendung des Verfahrens für simulierte Verkehrsabläufe soll im folgenden Abschnitt dargestellt werden.

5.2.2 Simulationsergebnisse

Der Verkehrsablauf wurde mit einem mikroskopischen Verkehrsmodell simuliert, wodurch die Schätzung von Reisezeiten in gestörten Verkehrsabläufen untersucht werden konnte. Die für die Simulationen gewählten Parameter finden sich in Tabelle 5.1.

Tabelle 5.1: Parameter zur Schätzung mit Fahrzeugfolgenkorrelation

	Schleife	1m	2,5m
Länge der Fahrzeugfolge	L	50	
Verschiebungsraster der Fahrzeuge	Δm	5, 10	
Filterkoeffizienten von $h(n, m)$	σ	8, 12, 16	
	α	0 ... 0, 1	
Entscheidungsschwelle	d_{opt}	33	22
Dimension der Merkmalsvektoren	N	14	15

Zur Demonstration der Schätzalgorithmen wird für dieses und die folgenden Schätzverfahren eine Simulation für einen 2,5 km langen zweispurigen Streckenabschnitt einer Autobahn betrachtet. Die Simulationsdauer beträgt 30 Minuten und die Störung des Verkehrsablaufs wird durch die Sperrung einer Fahrspur simuliert, die nach 10 Minuten für die Dauer von 10 Minuten einsetzt. Die Meßdaten dieser Simulation stammen zum einen von Fahrzeugsignalen der 1m-Doppelschleife und zum anderen von den 2,5m-Schleifen von verschiedenen Meßquerschnitten.

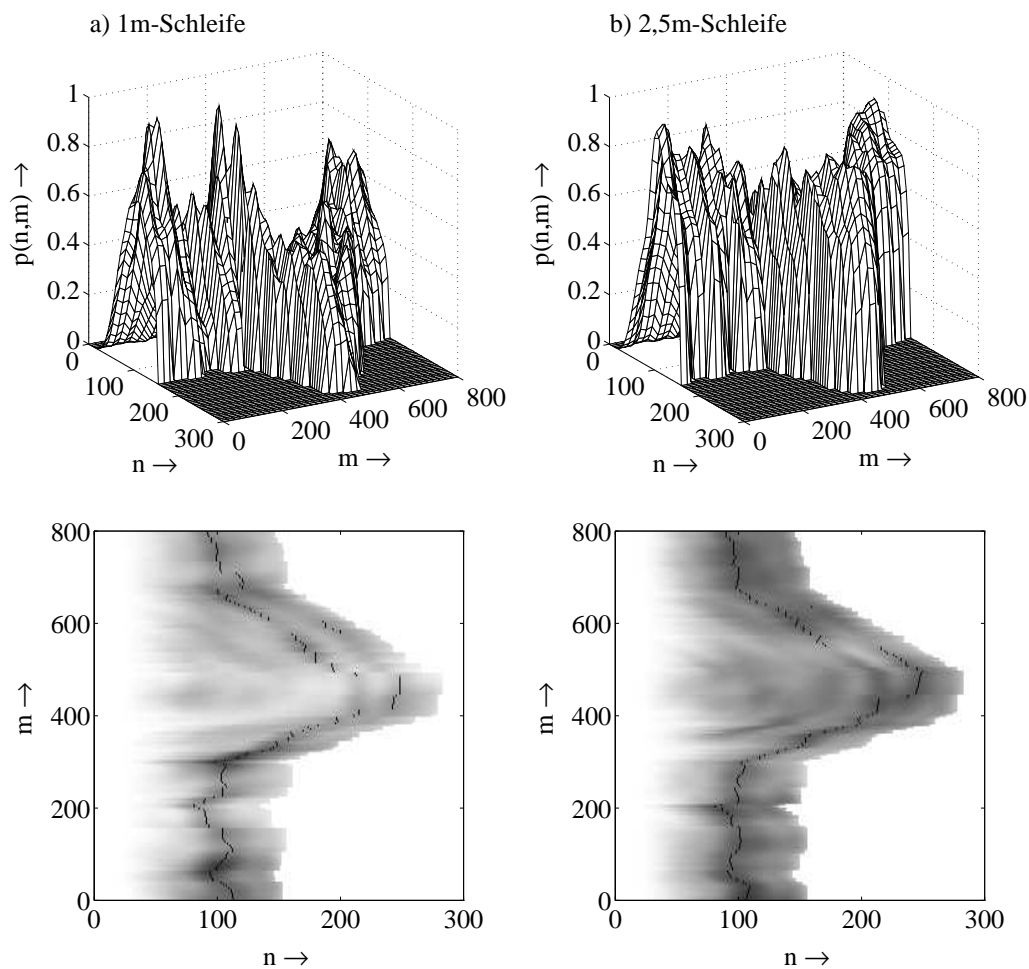


Bild 5.11: Fahrzeugfolgenkorrelation im simulierten Verkehrsablauf: a) 1m-Schleife, b) 2,5m-Schleife

Die Störung im Streckenabschnitt bedingt eine Erhöhung der Reisezeit sowie der Verkehrsdichte im Streckenabschnitt. Die Erhöhung der Verkehrsdichte zeigt sich deutlich in Bild 5.11, in dem die Fahrzeugfolgenkorrelation $p(n, m)$ dargestellt ist. Die Vergrößerung der Verkehrsdichte im mittleren Abschnitt ist durch eine deutliche Verbreiterung des Korrelationsgebirges sichtbar. Es wird ersichtlich, daß sich maximal bis zu 250 Fahrzeuge im Streckenabschnitt befinden. Die genaue Anzahl der Fahrzeuge ist durch die Tiefpaßfilterung der Korrelationsfunktionen $r(n, m)$ allerdings aus dem Bild nur näherungsweise bestimmbar.

Die ausgeprägten Maxima der Fahrzeugfolgenkorrelation zu Beginn der Simulation deuten auf eine sehr gute Schätzung der Reisezeit im freien Verkehrsablauf hin. Zur besseren Übersicht sind die Maximalwerte in den Höhenprofilen schwarz hervorgehoben. Aus Bild 5.11 wird deutlich, daß nach der Erhöhung der Fahrzeuganzahl durch die Störung zwei lokale Maxima auftreten. Diese werden dadurch erzeugt, daß verschiedene Fahrzeuggruppen am Streckenende spurselektiv detektiert werden. Die Ursache der Aufteilung in zwei Fahrzeuggruppen liegt in der Sperrung einer Fahrspur und dem Einfädeln auf eine gemeinsame Fahrspur. Diese Spurselektivität wurde auch in realen Messungen beobachtet [9].

Ein Problem ergibt sich in der Bestimmung der Reisezeit durch eine Mittelwertbildung, wie sie nach Gleichung (5.20) durchgeführt wird. Durch die Bestimmung der zugeordneten Fahrzeuggruppe aus dem globalen Maximum der Fahrzeugfolgenkorrelation geht die Information der Spurselektivität verloren. Im ungünstigsten Fall erfolgt je nach der Größe der Korrelationsfunktion in den lokalen Maxima alternierend eine Zuordnung der einen oder anderen Fahrzeuggruppe. Somit liegt der Schätzwert der Reisezeit entweder über oder unter dem wahren Mittelwert der Reisezeit.

Dieses verdeutlicht Bild 5.12, in dem für diese Simulationen die wahren und die aus den Fahrzeugfolgenkorrelationen geschätzten mittleren Reisezeiten aufgetragen sind. Es zeigt sich, daß die Mittelung im freien Verkehr problemlos durchgeführt werden kann, daß die oben angesprochenen Probleme allerdings keine eindeutige Bestimmung der Reisezeit im gestörten Verkehrsablauf ermöglichen.

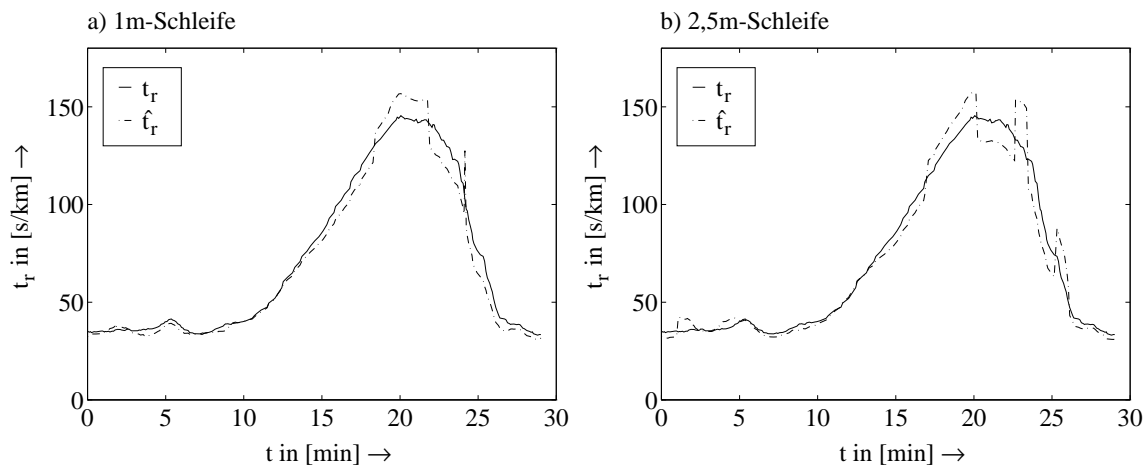


Bild 5.12: Wahre und geschätzte mittlere Reisezeiten aus Fahrzeugfolgenkorrelationen

5.3 Schätzverfahren mit Kostenfunktionen

Die Nachteile der Schätzung streckenbezogener Verkehrsdaten für kritische Belastungsprofile mit Fahrzeugfolgenkorrelationen könnten mit Verfahren, bei denen die individuelle Wiedererkennung von Fahrzeugen im Vordergrund steht, überwunden werden. Allerdings zeigt sich aus Abschnitt 5.1.4, daß die Zuordnung von Fahrzeugen mit ausschließlicher Verwendung der Merkmalsvektoren noch keine zufriedenstellenden Ergebnisse liefern kann.

In diesem Abschnitt soll ein Verfahren vorgestellt werden, das weitere charakteristische Fahrzeugdaten berücksichtigt, die zum einen aus lokalen Meßwerten an den Meßquerschnitten abgeleitet werden und zum anderen die Zusammensetzung des Verkehrs berücksichtigen. Für jede mögliche Einzelentscheidung wird eine Fehlerwahrscheinlichkeit ermittelt, deren Kehrwert als Kosten aufgefaßt werden kann. Jedes Fahrzeug wird dann im Verbund mit den übrigen im Streckenabschnitt befindlichen Fahrzeugen im Sinne einer minimalen Fehlerwahrscheinlichkeit erkannt.

Das Schätzverfahren soll im folgenden Abschnitt hergeleitet werden. Nach einer Untersuchung und Optimierung des Verfahrens werden abschließend Simulationen zur Reisezeitschätzung präsentiert.

5.3.1 Der Schätzalgorithmus

An den beiden Meßquerschnitten können für jedes Fahrzeug i neben der Extraktion der Merkmalsvektoren $\mathbf{v}_{1,i}$ und $\mathbf{v}_{2,i}$ sowie der Erfassung der Auftrittszeitpunkte $t_{1,i}$ und $t_{2,i}$ bei Doppelschleifen noch die Fahrzeuggeschwindigkeiten $v_{1,i}$ und $v_{2,i}$ ermittelt werden. Wird parallel zur streckenbezogenen Verkehrsdatenerfassung eine Fahrzeugklassifikation durchgeführt, ist auch eine Berücksichtigung der jeweiligen Fahrzeugklasse und der Entscheidungsvektoren $\mathbf{d}_{1,i}$ bzw. $\mathbf{d}_{2,i}$ möglich. Die Fahrzeugdaten können zu den neuen Fahrzeugmerkmalsvektoren $\mathbf{u}_{1,i}$ und $\mathbf{u}_{2,i}$ zusammengefaßt werden.

In der Bayes-Schätzung werden *Kostenfunktionen* zur Gewichtung von Schätzfehlern eingesetzt [40], [79]. Mit Hilfe derartiger Kostenfunktionen können die streckenbezogenen Verkehrsdaten eines Fahrzeugs nach der Detektion am stromabwärtigen Meßquerschnitt ermittelt werden. Aus den einzelnen Kostenfunktionen ergibt sich ein Kostenvektor

$$\mathbf{k}(\mathbf{u}_{1,i}, \mathbf{u}_{2,j}) = \begin{bmatrix} k_1(\mathbf{u}_{1,i}, \mathbf{u}_{2,j}) \\ \vdots \\ k_n(\mathbf{u}_{1,i}, \mathbf{u}_{2,j}) \end{bmatrix}. \quad (5.30)$$

Aus dem Kostenvektor können die Gesamtkosten für eine Fahrzeugzuordnung mit

$$K(\mathbf{u}_{1,i}, \mathbf{u}_{2,j}) = \mathbf{g}^T \mathbf{k}(\mathbf{u}_{1,i}, \mathbf{u}_{2,j}) \quad (5.31)$$

ermittelt werden. Der Vektor \mathbf{g} gewichtet hierbei die einzelnen Kostenanteile.

Kostenfunktionen

Die direkt aus den Merkmalsvektoren $\mathbf{v}_{l,i}$ resultierenden Kosten ergeben sich aus der Entscheidungsfunktion (5.11):

$$k_1(\mathbf{u}_{1,i}, \mathbf{u}_{2,j}) = f(d(\mathbf{v}_{1,i}, \mathbf{v}_{2,j})). \quad (5.32)$$

Ausgehend von der am ersten Meßquerschnitt ermittelten lokalen Geschwindigkeit $v_{1,i}$ eines Fahrzeugs kann eine *Prognose der Ankunftszeit*

$\hat{t}_{2,i}$ am zweiten Meßquerschnitt durchgeführt werden. Daraus ergibt sich eine Kostenfunktion

$$k_2(\mathbf{u}_{1,i}, \mathbf{u}_{2,j}) = f(\hat{t}_{2,i} - t_{2,j}). \quad (5.33)$$

Analog zu den oben angeführten Überlegungen kann eine *Prognose der Abfahrtszeit* $\hat{t}_{1,j}$ eines Fahrzeugs aus den Messungen am zweiten Meßquerschnitt durchgeführt werden. Die Kostenfunktion hierfür lautet

$$k_3(\mathbf{u}_{1,i}, \mathbf{u}_{2,j}) = f(\hat{t}_{1,j} - t_{1,i}). \quad (5.34)$$

Die Funktionen in (5.33) und (5.34) ergeben sich aus den statistischen Abweichungen der geschätzten und der wahren Ankunfts- bzw. Abfahrtszeiten der Fahrzeuge und müssen anhand einer Lernstichprobe ermittelt werden. In diesen beiden Prognosen könnten zusätzlich die aktuellen querschnittsbezogenen Verkehrsdaten berücksichtigt werden.

Eine *Reisezeitprognose* kann anhand der Reisezeitschätzungen von vorausgegangenen Fahrzeugen erfolgen. Dieser Ansatz beruht darauf, daß die Reisezeit eine quasistationäre Größe ist, die sich auch bei plötzlich auftretenden Störungen im Verkehrsablauf sehr langsam ändert.

Die gemittelte Reisezeit von vorher detektierten Fahrzeugen soll mit \hat{t}_r bezeichnet werden, wobei die Ermittlung dieser Größe später hergeleitet wird (siehe Abschnitt 5.3.2). Dann kann folgende Kostenfunktion aufgestellt werden:

$$k_4(\mathbf{u}_{1,i}, \mathbf{u}_{2,j}) = f(\hat{t}_r - [t_{2,j} - t_{1,i}]). \quad (5.35)$$

Die Funktion in Gleichung (5.35) ergibt sich ebenfalls aus den statistischen Abweichungen der geschätzten und der wahren Reisezeiten der betrachteten Fahrzeuge.

Eine weitere Kostenfunktion könnte aus dem Abstand der Entscheidungsvektoren $d(\mathbf{d}_{1,i}, \mathbf{d}_{2,j})$ der Klassifikation festgelegt werden. Obwohl die Vorverarbeitung und Merkmalsextraktion bei der Fahrzeugklassifikation anders als bei der Wiedererkennung ist, sind die Entscheidungsvektoren der Fahrzeuge gleicher Klasse einander trotzdem sehr ähnlich und

somit kommt es fast ausschließlich zu Falschzuordnungen innerhalb der gleichen Fahrzeugklasse. Deshalb ist es sinnvoll, die Entscheidungsvektoren nicht in die Kostenfunktionen zu integrieren, zumal der dann entstehende Merkmalsvektor um die Dimension der Klassenanzahl erweitert werden müßte.

Ablauf des Schätzverfahrens

Nach der Detektion eines Fahrzeugs j am stromabwärtigen Meßquerschnitt mit dem Fahrzeugvektor $\mathbf{u}_{2,j}$ erfolgt ein Vergleich des Fahrzeugs mit allen K auf der Strecke befindlichen Fahrzeugen mit den am stromaufwärtigen Meßquerschnitt erfaßten Fahrzeugvektoren $\mathbf{u}_{1,i}$, $i = 1, \dots, K$. Für jedes Fahrzeug k wird die Kostenfunktion $K(\mathbf{u}_{1,k}, \mathbf{u}_{2,j})$ bezüglich des Fahrzeugs j gebildet. Anschließend findet eine Vorauswahl statt, indem dasjenige Fahrzeug i von der ersten Meßstelle mit den geringsten Gesamtkosten zugeordnet wird,

$$K(\mathbf{u}_{1,i}, \mathbf{u}_{2,j}) = \min_{k=1, \dots, K} K(\mathbf{u}_{1,k}, \mathbf{u}_{2,j}). \quad (5.36)$$

Eine begrenzte Anzahl der Fahrzeuge vom ersten Meßquerschnitt mit den geringsten Gesamtkosten wird vorerst zusammen mit den dazugehörigen Kosten zwischengespeichert. Da es zu Doppelzuweisungen kommen kann, erfolgt anschließend eine Untersuchung, ob das bei der Vorauswahl zugeordnete Fahrzeug bereits zugeordnet worden ist. Ist das der Fall, so kommt es zu einer Korrektur für das Fahrzeug mit den größeren Kosten und für dieses Fahrzeug wird das nächstwahrscheinlichere zugeordnet. Gleichzeitig findet auch hier wieder eine Überprüfung bezüglich einer Doppelzuweisung statt, und die Korrekturphase wird erst beendet, wenn keine Doppelzuweisung mehr vorliegt.

Nach einer Wartezeit W erfolgt die endgültige Zuordnung eines Fahrzeugs. Die Wartezeit beruht auf einem festen Zeitabschnitt oder auf einer festen Anzahl von Fahrzeugen. Die Zuweisung soll allerdings nur stattfinden, wenn die Gesamtkosten des Fahrzeugs eine obere Schwelle nicht überschreiten. Ansonsten wird das Merkmalsvektorenpaar verworfen und trägt nicht zur Schätzung der streckenbezogenen Verkehrsdaten bei.

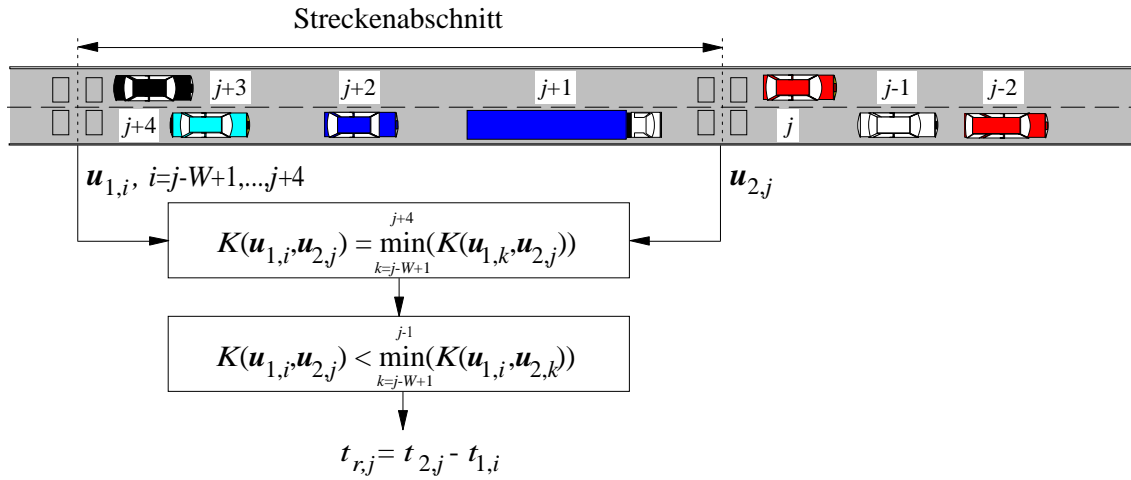


Bild 5.13: Beispiel zur streckenbezogenen Verkehrsdatenerfassung mit Kostenfunktionen

Nach der endgültigen Zuordnung eines Fahrzeugs werden die Fahrzeugvektoren der beiden Messstellen $\mathbf{u}_{1,i}$ und $\mathbf{u}_{2,j}$ eliminiert. Die Einzelfahrzeugreisezeit wird mit

$$t_{r,j} = t_{2,j} - t_{1,i} \quad (5.37)$$

festgelegt. Das Schätzverfahren ist vereinfacht in Bild 5.13 dargestellt.

Streckenbezogene Verkehrsdaten

Die mittlere Reisezeit einer Fahrzeuggruppe ergibt sich aus einer Mittelung der Einzelfahrzeugreisezeiten über einen festen Zeitraum T oder eine definierte Fahrzeuggruppe der Größe M . Die Art der Mittelung kann auf verschiedene Weisen erfolgen. Eine Möglichkeit besteht in der Mittelwertbildung, d.h.

$$t_r = \frac{1}{M} \sum_{m=1}^M t_{r,m}. \quad (5.38)$$

Dieses Verfahren berücksichtigt alle geschätzten Individualreisezeiten der Fahrzeuggruppe, wobei aber auch Falschzuordnungen einbezogen werden. Deshalb bietet es sich an, nach einer Häufung von nahezu gleichen Einzelreisezeiten zu suchen oder aber den Median der Einzelreisezeit des

Kollektivs als mittlere Reisezeit zu nehmen. Somit ist bei einer hohen Wiedererkennungsrates eine richtige Bestimmung der Reisezeit gewährleistet.

Die Verkehrsdichte und die mittlere Geschwindigkeit ergeben sich analog zur Gleichung (5.21).

5.3.2 Optimierung des Verfahrens

Zur optimalen Zuweisung zweier Fahrzeuge müssen die einzelnen Parameter der Kostenfunktionen bestimmt werden. Dieses geschieht mit einer Lernstichprobe von wahren Meßdaten und simulierten Verkehrsabläufen, wobei vornehmlich kritische durch Störungen verursachte Verkehrsverhältnisse berücksichtigt werden müssen.

Entscheidungsfunktionen

Nach den Untersuchungen in Abschnitt 5.1.3 werden die Kosten für den Abstand der Merkmalsvektoren $\mathbf{v}_{1,i}$ und $\mathbf{v}_{2,j}$ aus dem Quadrat der euklidischen Metrik nach Gleichung (5.13) bestimmt,

$$k_1(\mathbf{u}_{1,i}, \mathbf{u}_{2,j}) = d_{ij}^2(\mathbf{v}_{1,i}, \mathbf{v}_{2,j}). \quad (5.39)$$

Eine einfache Prognose der Ankunftszeit kann im freien Verkehrsablauf aus der Fahrzeuggeschwindigkeit und der Streckenlänge s mit

$$\hat{t}_{2,i} = t_{1,i} + \frac{s}{v_{1,i}} \quad (5.40)$$

gewonnen werden. Daraus ergibt sich die Kostenfunktion

$$k_2(\mathbf{u}_{1,i}, \mathbf{u}_{2,j}) = f\left(t_{1,i} + \frac{s}{v_{1,i}} - t_{2,j}\right). \quad (5.41)$$

Zur Bestimmung der Funktion (5.41) muß die Verteilungsdichte $p_n(\hat{t}_{2,i} - t_{2,j})$ ermittelt werden. Aus dieser kann auf den Verlauf der Kostenfunktion geschlossen werden. Hierzu wurden die Prognosen der Ankunftszeit und die wahren Ankunftszeiten von Fahrzeugen aus simulierten

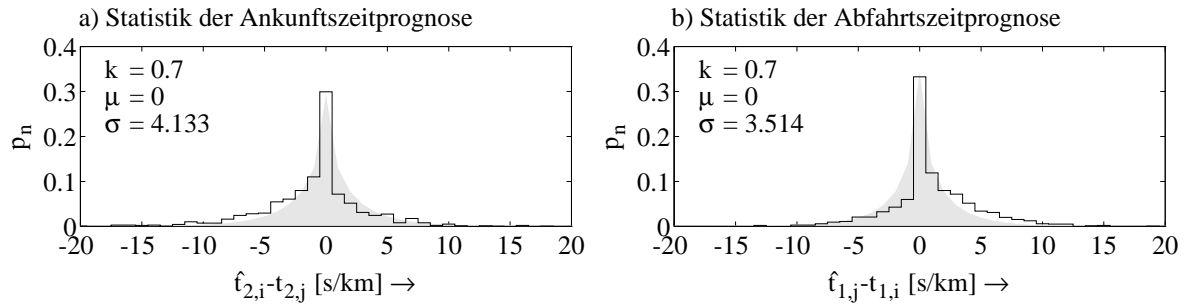


Bild 5.14: Beispiele zur Prognoseabweichung im freien Verkehrsfluß: a) Ankunftszeitprognose, b) Abfahrtszeitprognose

Verkehrsabläufen verglichen. Bild 5.14a zeigt die statistischen Abweichungen der prognostizierten und der wahren Werte für eine Simulation mit freiem Verkehrsfluß. Deutlich erkennbar ist eine symmetrische Streuung der Abweichungen im freien Verkehrsablauf.

Beschreibt man die Abweichungen zwischen wahrer und prognostizierter Ankunftszeit als mittelwertfreien Störprozeß mit einer generalisierten gaußschen Dichte

$$p_n(\hat{t}_{2,i} - t_{2,j}) = \frac{k}{2A(k)\Gamma(1/k)} \exp\left(-\left[\frac{|\hat{t}_{2,i} - t_{2,j}|}{A(k)}\right]^k\right), \quad (5.42)$$

wobei die Größe $A(k)$ sich mit Hilfe der Gamma-Funktionen

$$\Gamma(b) = \int_0^\infty x^{b-1} e^{-x} dx \quad (5.43)$$

zu

$$A(k) = \left(\sigma^2 \frac{\Gamma(1/k)}{\Gamma(3/k)}\right)^{1/2} \quad (5.44)$$

berechnet, so ergibt sich die Kostenfunktion

$$k_2(\mathbf{u}_{1,i}, \mathbf{u}_{2,j}) = \left|t_{1,i} + \frac{s}{v_{1,i}} - t_{2,j}\right|^k, \quad (5.45)$$

wobei die Varianz in die Gewichtung der Kostenfunktion eingeht.

Ähnliche Verhältnisse ergeben sich bei der Prognose der Abfahrtszeit der Fahrzeuge am Streckenanfang aus der lokalen Geschwindigkeit am Streckenende. Mit der Schätzgleichung

$$\hat{t}_{1,j} = t_{2,j} - \frac{s}{v_{2,j}} \quad (5.46)$$

ergibt sich die Kostenfunktion

$$k_3(\mathbf{u}_{1,i}, \mathbf{u}_{2,j}) = \left| t_{2,j} - \frac{s}{v_{2,j}} - t_{1,i} \right|^k. \quad (5.47)$$

Bild 5.14b gibt die statistischen Abweichungen zwischen der Prognose und der wahren Abfahrtszeit für ein Beispiel im freien Verkehrsablauf wieder. Grau eingetragen sind im Bild 5.14 die durch gaußsche Dichten angenäherten Verteilungskurven.

Da sich die Individualreisezeit der Fahrzeuge im Verkehrsablauf nur langsam ändert, kann eine Prognose der Reisezeit eines Fahrzeugs mit Hilfe aktueller geschätzter Reisezeiten vorausfahrender Fahrzeuge durchgeführt werden. In Folge von Schätzfehlern muß eine feste Anzahl von Fahrzeugen berücksichtigt werden, da das Schätzverfahren sonst nicht sehr robust ist. Bei Berücksichtigung einer Gruppe von m vorausfahrenden Fahrzeugen und einer Mittelung mit dem Median der Reisezeiten ergibt sich als Prognosewert der Fahrzeugreisezeit

$$\hat{t}_{r,j} = \text{Median}\{t_{r,j-m}, \dots, t_{r,j-1}\}, \quad (5.48)$$

und damit als Kostenfunktion

$$k_4(\mathbf{u}_{1,i}, \mathbf{u}_{2,j}) = f(\text{Median}\{t_{r,j-m}, \dots, t_{r,j-1}\} - [t_{2,j} - t_{1,i}]). \quad (5.49)$$

Die Funktion (5.49) ergibt sich aus den statistischen Abweichungen der wahren und der prognostizierten Reisezeiten. Bild 5.15 gibt diese Abweichungen für zwei simulierte Verkehrsabläufe wieder. Es zeigt sich, daß die statistischen Abweichungen sich im freien Verkehrsablauf und im gestörten Verkehrsablauf nur unwesentlich unterscheiden.

Bild 5.15 zeigt, daß die Reisezeitdifferenzen aufeinanderfolgender Fahrzeuge sich ebenfalls als mittelwertfreier Störprozeß mit einer generalisierten gaußschen Dichte nach den Gleichungen (5.42) bis (5.44) beschreiben

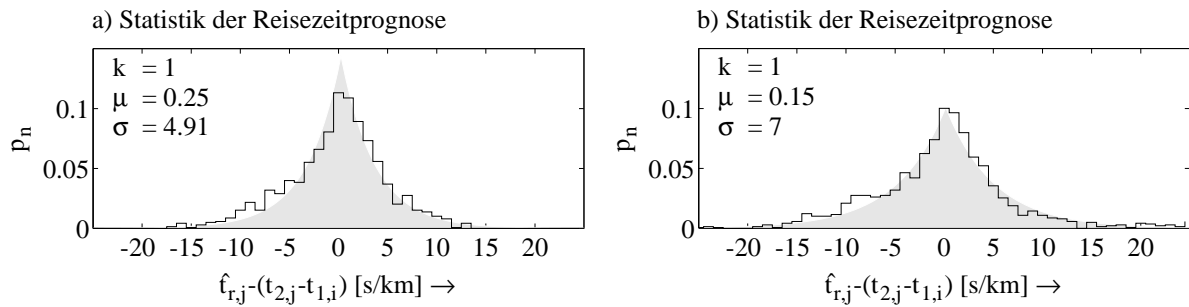


Bild 5.15: Beispiele zur Abweichung der Reisezeitprognose: a) freier Verkehrsfluß, b) gestörter Verkehrsfluß

lassen. Der Parameter k muß optimiert werden, die Standardabweichung σ ist nahezu vom Verkehrszustand unabhängig. Die Kostenfunktion hat dann folgende Form:

$$k_4(\mathbf{u}_{1,i}, \mathbf{u}_{2,j}) = |\text{Median}\{t_{r,j-m}, \dots, t_{r,j-1}\} - [t_{2,j} - t_{1,i}]|^k. \quad (5.50)$$

Gewichtsvektor

Nach der Festlegung der Kostenfunktionen müssen die optimalen Gewichtungsfaktoren der Kostenfunktionen ermittelt werden. Die Gewichtung der Kostenfunktion k_1 wird mit 1 vorgegeben, d.h. $\mathbf{g} = [1, g_2, g_3, g_4]^T$.

Grundlage zur Bestimmung des Gewichtsvektors sind verschiedene mikroskopische Verkehrssimulationen, aus denen die wahren Streckengrößen hervorgehen. Die Berechnung der Merkmalsvektoren $\mathbf{v}_{l,i}$ erfolgt anhand gemessener Fahrzeugsignale. Die weiteren Größen sind Resultate des Simulationsprogramms. Die Erzeugung des gesamten Fahrzeugvektors erfolgt anschließend durch die Zuordnung der Merkmalsvektoren eines detektierten Fahrzeugs zu den lokalen Verkehrsdaten eines simulierten Fahrzeugs.

Für die Kostenfunktionen (5.45) und (5.47) aus den Ankunfts- und Abfahrtszeitprognosen liegt der optimale Wert für den Parameter k bei $k = 0,7$. Bild 5.16a zeigt die Fehlerrate der Einzelfahrzeugerkennung für einen Gewichtsvektor $\mathbf{g} = [1, g_2, 0, 0]^T$ in einem 2,5 km langen Streckenabschnitt für verschiedene Verkehrsbelastungen und -zustände. Es zeigt sich,

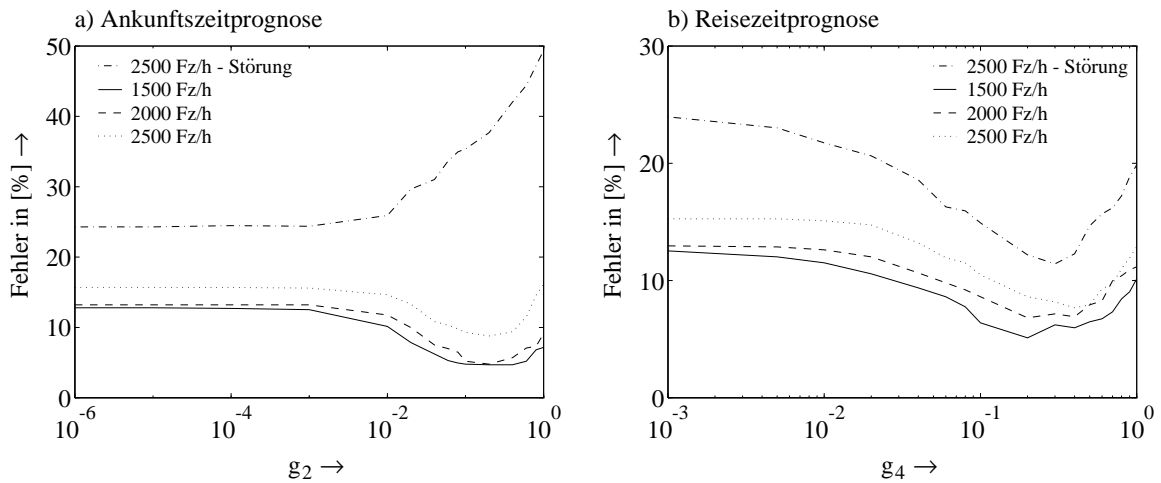


Bild 5.16: Fehlerraten der Wiedererkennung in Abhängigkeit der Gewichtungsfaktoren: a) Prognose der Ankunftszeit, b) Prognose der Reisezeit

daß ein Gewichtungsfaktor von $g_2 = 0,01$ die Fehlerrate für nicht gestörte Verkehrsverhältnisse nahezu halbiert. Bei einer Störung im Verkehrsablauf führt die Einbeziehung der Ankunftszeitprognose zu einer Erhöhung der Fehlerrate, so daß der Gewichtungsfaktor g_2 in diesem Fall zu Null gesetzt werden sollte. Für die Abfahrtszeitprognose ergeben sich die gleichen Ergebnisse.

Aus Simulationen für unterschiedliche Verkehrsverhältnisse ergibt sich, daß der Wert $k = 1$ in der Kostenfunktion (5.50) für die Reisezeitprognose optimal ist und die Kosten somit linear zum Abstand der Reisezeitprognose zunehmen. Der Gewichtungsfaktor g_4 liegt im Bereich $0,2 \leq g_4 \leq 0,4$. Dieses geht aus Bild 5.16b hervor, in dem die Fehlerrate der Wiedererkennung für einen Gewichtungsfaktor $\mathbf{g} = [1, 0, 0, g_4]$ aufgetragen ist. Dabei ist der optimale Gewichtungsfaktor vom Verkehrszustand unabhängig.

Die Gewichtungsfaktoren wurden für Merkmalsvektoren der Dimension $N = 14$ bzw. $N = 15$ festgelegt. Bei einer Verringerung der Dimension geht die Kostenfunktion k_1 mit einer anderen Gewichtung in das Verfahren ein und die restlichen Gewichtungsfaktoren müßten entsprechend angepaßt werden.

5.3.3 Simulationsergebnisse

Die den hier dargebrachten Ergebnissen zugrundeliegenden Fahrzeugsignale und mikroskopischen Verkehrssimulationen sind mit denen aus Abschnitt 5.2.2 identisch. Die für die Simulationen gewählten Parameter finden sich in Tabelle 5.2

Tabelle 5.2: Parameter zur Schätzung mit Kostenfunktionen

	Schleife	1m	2,5m
Wartelänge bis zur Zuordnung	W	1, ..., 70	
Gewichtsfaktoren	g_1	1	
	g_2	0 ... 0, 1	
	g_3	0 ... 0, 1	
	g_4	0, 0,2, 0,4	
Dimension der Merkmalsvektoren	N	14	15

Führt man die Einzelfahrzeugwiedererkennung ausschließlich mit den Merkmalsvektoren der Fahrzeuge ohne weitere Kostenfunktionen durch, so ergibt sich wie erwartet eine Vielzahl von Falschzuordnungen für diese Simulation. Bild 5.17a zeigt die aus der Simulation stammenden wahren und die geschätzten Reisezeiten der Fahrzeuge für Messungen mit der 2,5m-Schleife. In Bild 5.18a ist der über drei Zeitperioden der Länge $T = 10$ s gemittelte Median der Reisezeiten der Fahrzeuge aufgetragen. Die Intervalllänge wurde hierbei dem zur Verkehrszustandsschätzung in Kapitel 6 eingesetzten makroskopischen Verkehrsmodell angepaßt. Bei der Wahl des Medians statt des Ensemble-Mittelwerts hat man den Vorteil, daß Falschentscheidungen weniger gewichtet werden. Es zeigt sich, daß der Anstieg der mittleren Reisezeit im Störungsbereich trotz der Vielzahl von individuellen Fehlentscheidungen noch erkennbar ist. Die entsprechenden Simulationsergebnisse für die Daten der 1m-Schleife finden sich in den Bildern 5.17c und 5.18c.

Erweitert man die Fahrzeugzuordnungen durch die übrigen Kostenfunktionen, so reduziert sich die Anzahl der Fehlentscheidungen beträchtlich. Ein weiterer Vorteil liegt darin, daß die falsch zugewiesenen Fahrzeu-

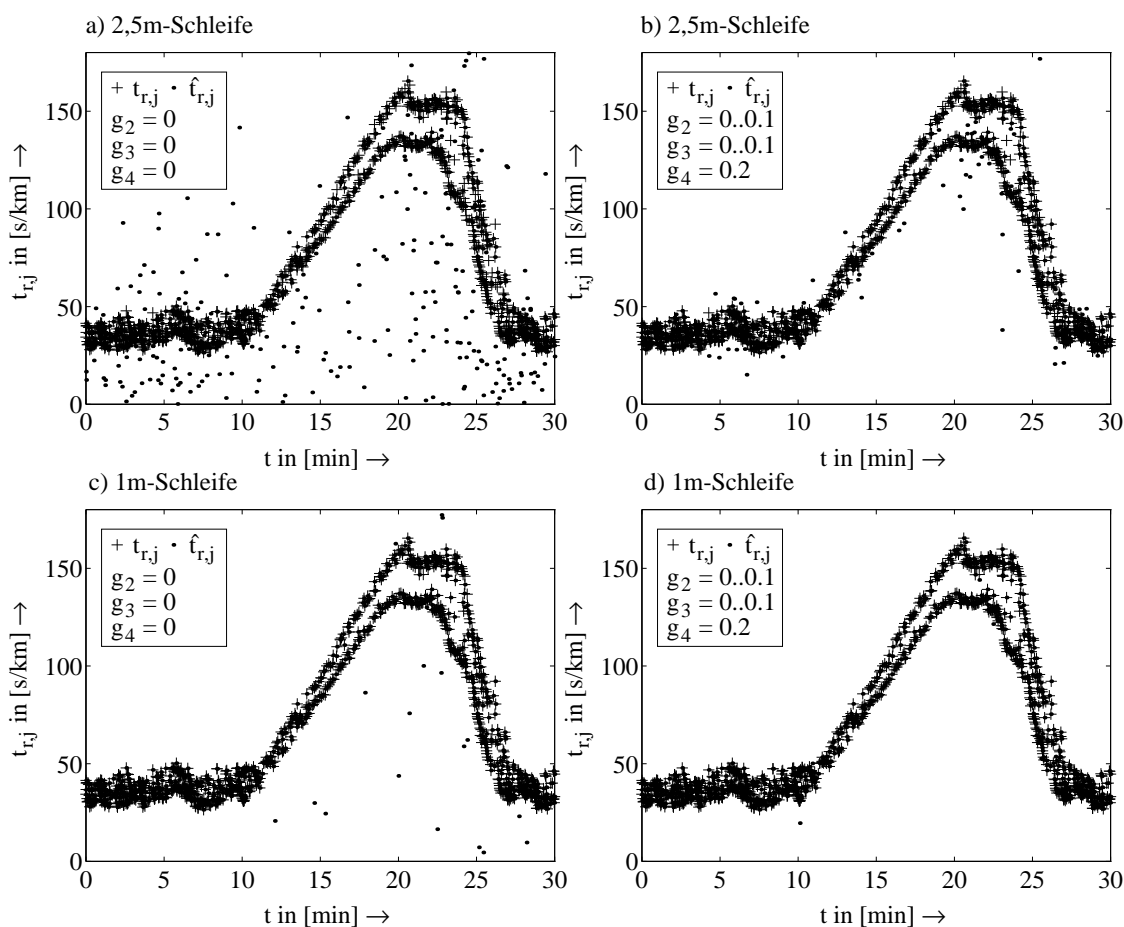


Bild 5.17: Wahre und geschätzte Fahrzeugreisezeiten

ge eine sehr geringe Reisezeitdifferenz zu den wahren Werten aufweisen. Hier wirkt sich der Abstand der Reisezeitprognose als Kostenfaktor aus. Damit ist auch eine erheblich bessere Schätzung der mittleren Reisezeit der Fahrzeuge möglich. In Bild 5.18b ist analog zum Bild 5.18a der wahre und der geschätzte Median der Reisezeiten für Signale der 2,5m-Schleife aufgetragen. Deutlich bessere Schätzergebnisse lassen sich auch hier mit Signalen der 1m-Schleife erzielen. In Bild 5.18d sind die Werte aus der Simulation dargestellt. Die Fehlerrate der Falschzuordnungen liegt bei 1 %, wobei aber wieder berücksichtigt werden muß, daß die Signale von Doppelschleifen stammen und damit hinsichtlich der Wiedererkennung optimal sind.

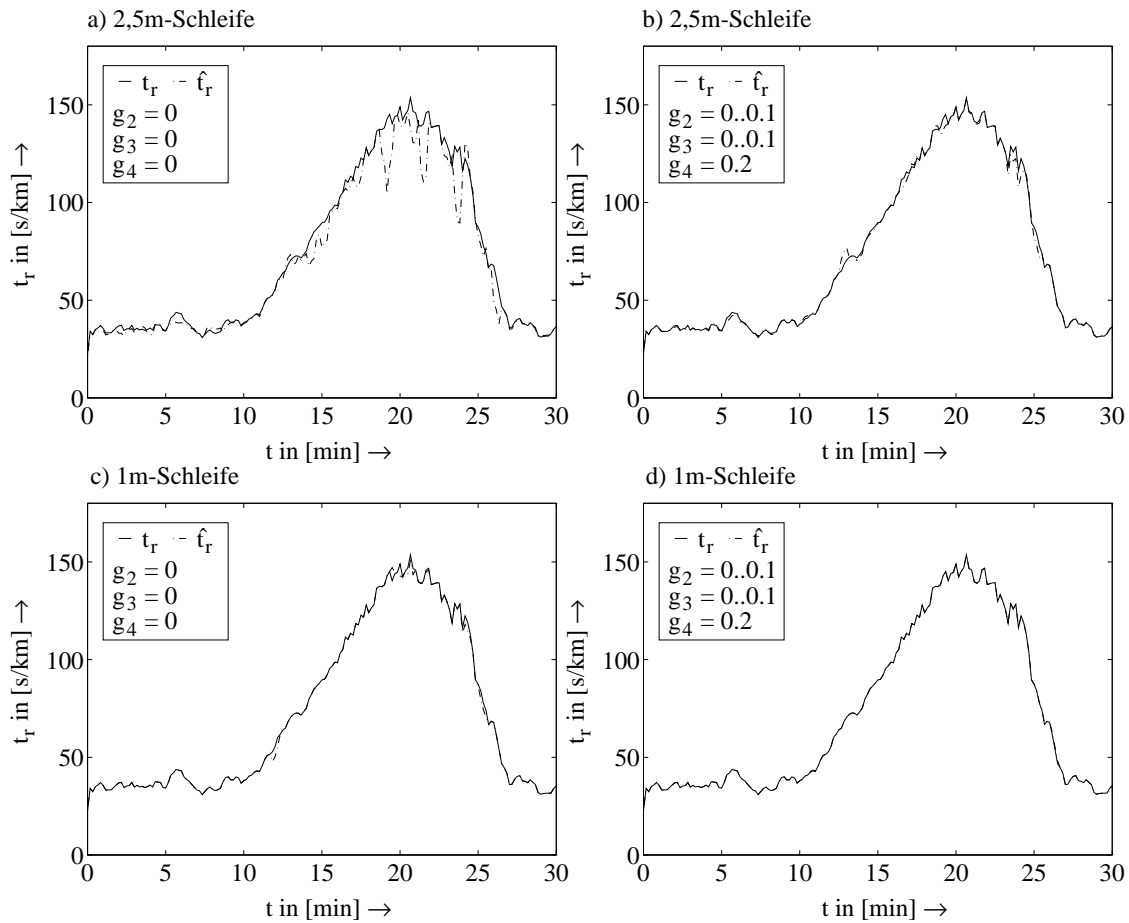


Bild 5.18: Wahrer und geschätzter Median der Reisezeiten für $T = 10$ s

Für die dargestellten Simulationen beträgt die Wartezeit bis zur endgültigen Zuordnung $W = 50$ Fahrzeuge. Als Anzahl der Merkmale wurde nach den Ergebnissen aus Abschnitt 5.1.3 $N = 14$ für die Signale der 1m-Schleife und $N = 15$ für die Signale der 2,5m-Schleife gewählt. Mit Hilfe weiterer Simulationen kann gezeigt werden, daß die Schätzungen auch bei einer Wahl von $N = 6$ noch gute Ergebnisse liefern und die Wartezeit einen Wert von $W = 30$ Fahrzeugen nicht überschreiten sollte. Bei einer Verkehrsstärke von 2500 Fz/h bedeutet das, daß das Schätzergebnis mit einer Verzögerung von maximal 40 s bereitgestellt wird. Für längere Streckenabschnitte erhöht sich diese Verzögerungszeit entsprechend.

5.4 Schätzverfahren mit Hidden-Markov-Modellen

Die Ergebnisse aus dem letzten Abschnitt zeigen, daß die Einbeziehung der zeitlichen statistischen Bindungen zwischen einzelnen Fahrzeugen im Streckenabschnitt die Bestimmung streckenbezogener Verkehrsdaten verbessert. Könnte man diese Bindungen durch einen mathematischen Prozeß beschreiben, so müßten sich verhältnismäßig einfache Algorithmen zur Schätzung der streckenbezogenen Verkehrsdaten ergeben. In [45] wird gezeigt, daß die Bewegung von Fahrzeugen in einem Streckenabschnitt als Punkt-Prozeß modelliert werden kann. Betrachtet man die extrahierten Merkmalsvektoren der Fahrzeuge als sogenannte Markierungen (*marks*) der Fahrzeuge, so kann man von einem markierten Punkt-Prozeß (*marked* oder *compound point-process*) sprechen [75]. Für derartige Prozesse können die Verzögerungen der Auftrittszeitpunkte einzelner Objekte zwischen zwei Orten mit Hidden-Markov-Modellen bestimmt werden [84].

In diesem Abschnitt soll ein Schätzverfahren mit Hidden-Markov-Modellen präsentiert werden. Die Theorie der Hidden-Markov-Modelle wurde Ende der sechziger Jahre von BAUM et.al. eingeführt [5], [6], [7] und in den siebziger Jahren in ersten Anwendungen in der Spracherkennung implementiert [42], [4]. In den letzten Jahren ist die Theorie der Hidden-Markov-Modelle intensiv behandelt und erweitert worden, woraus schließlich Anwendungen in zahlreichen technischen Applikationen resultierten, siehe hierzu z.B. [63]. In Abschnitt 5.4.1 erfolgt eine Übersicht über die Generierung von Hidden-Markov-Modellen, der die Entwicklung eines Modells zur Lösung des vorliegenden Schätzproblems folgt. Dieses Modell benutzt die extrahierten Merkmalsvektoren der Fahrzeuge als Beobachtungen und bestimmt diejenige unbekannte Fahrzeugfolge, die mit höchster Wahrscheinlichkeit zu dieser Beobachtung geführt hat. Aus der ermittelten Fahrzeugfolge können die streckenbezogenen Verkehrsdaten abgeleitet werden. Eine Optimierung des Verfahrens und Simulationsergebnisse schließen auch diesen Abschnitt ab.

5.4.1 Hidden-Markov-Modelle

Hidden-Markov-Modelle sind umfassend in [64], [65], [66] und [32] beschrieben. Im folgenden sollen die zur streckenbezogenen Verkehrsdatenerfassung relevanten Definitionen und Algorithmen aufgeführt werden.

Definition eines Hidden-Markov-Modells

Zur Definition eines Hidden-Markov-Modells werde ein System betrachtet, das sich zu jedem Zeitpunkt in einem von N möglichen Zuständen befindet. Dieses System kann in jedem Zeitschritt seinen Zustand mit einer bestimmten Wahrscheinlichkeit ändern. Die diskreten Zeitschritte werden mit $t = 1, 2, \dots, T$ bezeichnet, der aktuelle Zustand zum Zeitpunkt t wird mit q_t festgelegt. Im allgemeinen ist die Wahrscheinlichkeit, daß das System sich in einem bestimmten Zustand j zum Zeitpunkt t aufhält, abhängig von den Zuständen, die das System zu vergangenen Zeitpunkten $t - 1, t - 2, \dots$ durchlaufen hat. Für den speziellen Fall einer diskreten Markov-Kette 1. Ordnung ist die Wahrscheinlichkeit dafür, daß ein Zustand j zum Zeitpunkt t erreicht wird, allerdings lediglich vom Zustand i des vorhergehenden Zeitpunktes $t - 1$ abhängig:

$$P(q_t = j | q_{t-1} = i, q_{t-2} = k, \dots) = P(q_t = j | q_{t-1} = i). \quad (5.51)$$

Die Wahrscheinlichkeit in Gleichung (5.51) wird *Übergangswahrscheinlichkeit* genannt. Sind die Übergangswahrscheinlichkeiten $P(q_t = j | q_{t-1} = i)$ für alle Zustände i und j unabhängig vom Zeitpunkt t , so ist die diskrete Markov-Kette homogen und die Übergangswahrscheinlichkeiten können abgekürzt mit

$$a_{ij} = P(q_t = j | q_{t-1} = i), \quad 1 \leq i, j \leq N, \quad (5.52)$$

bezeichnet werden. Für die Wahrscheinlichkeiten gilt:

$$a_{ij} \geq 0 \quad \forall j, i, \quad (5.53)$$

$$\sum_{j=1}^N a_{ij} = 1 \quad \forall i. \quad (5.54)$$

Der hierdurch bestimmte stochastische Prozeß kann als beobachtbares Markov-Modell bezeichnet werden. Der Prozeß liefert eine Anzahl determinierter Zustände zu jedem Zeitpunkt, und jeder Zustand ist direkt beobachtbar. Anhand der beobachteten Zustände kann nun eine Wahrscheinlichkeitsaussage über folgende Zustände getroffen werden, falls die Zustandsübergangswahrscheinlichkeiten vorliegen.

In den meisten technischen Anwendungen sind die betrachteten Zustände jedoch nicht direkt beobachtbar. Das Markov-Modell wird dahingehend erweitert, daß die vorliegenden Beobachtungen stochastische Abhängigkeiten von den Zuständen besitzen. Das hieraus resultierende sogenannte Hidden-Markov-Modell besteht somit aus einem stochastischen Prozeß, dem ein weiterer nicht beobachtbarer (*hidden*) stochastischer Prozeß unterliegt.

Zur Beschreibung von Hidden-Markov-Modellen werden folgende Parameter definiert:

- Zustände $Q = \{q_1, q_2, \dots, q_N\}$
- Anzahl der Zustände im Modell N
- Aktueller Zustand q_t
- Aktueller Zeitpunkt t
- Beobachtungssequenz $\mathbf{O} = (\mathbf{o}_1, \mathbf{o}_2, \dots, \mathbf{o}_T)$
- Anzahl der Objekte T
- Beobachtete Symbole $\mathbf{V} = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_M\}$
- Anzahl der Symbolklassen M
- Die Übergangswahrscheinlichkeiten von einem Zustand i zum Zeitpunkt t in einen Zustand j zum Zeitpunkt $t + 1$ werden mit der Transitionsmatrix $\mathbf{A} = \{a_{ij}\}$ mit a_{ij} nach (5.52) angegeben.

- Die *Erzeugungswahrscheinlichkeiten* eines Symbols werden mit der Matrix $\mathbf{B} = \{b_j(\mathbf{o}_t)\}$ beschrieben, wobei

$$b_j(\mathbf{o}_t) = P(\mathbf{o}_t = \mathbf{v}_k | q_t = j), \quad 1 \leq k \leq M, \quad (5.55)$$

die Wahrscheinlichkeit für die Beobachtung des Symbols $\mathbf{o}_t = \mathbf{v}_k$ im Zustand j mit $j = 1, 2, \dots, N$ angibt.

- Die Wahrscheinlichkeiten für die Anfangszustände zum Zeitpunkt $t = 1$ werden mit $\boldsymbol{\pi} = \{\pi_i\}$ bezeichnet, wobei

$$\pi_i = P(q_1 = i), \quad 1 \leq i \leq N, \quad (5.56)$$

ist.

Aus diesen Festlegungen wird ersichtlich, daß zur eindeutigen Angabe eines Hidden-Markov-Modells die Modellparameter N und M sowie die Symbole \mathbf{v}_k festgelegt und die Wahrscheinlichkeiten \mathbf{A} , \mathbf{B} und $\boldsymbol{\pi}$ bestimmt werden müssen. Hierzu wird die kompakte Schreibweise

$$\lambda = (\mathbf{A}, \mathbf{B}, \boldsymbol{\pi}) \quad (5.57)$$

zur Indizierung des Modells eingeführt.

Berechnung der Beobachtungswahrscheinlichkeit

Bei gegebener Beobachtungssequenz $\mathbf{O} = (\mathbf{o}_1, \mathbf{o}_2, \dots, \mathbf{o}_T)$ und gegebenen Modellparametern $\lambda = (\mathbf{A}, \mathbf{B}, \boldsymbol{\pi})$ ergibt sich die Frage nach einer effizienten Berechnung von $P(\mathbf{O}|\lambda)$, d.h. der Wahrscheinlichkeit für die Erzeugung der vorliegenden Beobachtungssequenz unter der Bedingung, daß das Modell mit λ vorgegeben ist. Für ein bestimmtes Modell und festem λ soll die Wahrscheinlichkeit im folgenden mit $P(\mathbf{O})$ bezeichnet werden.

Bei einem Modell mit N Zuständen und einer Beobachtungssequenz von T Objekten existieren insgesamt N^T mögliche Zustandssequenzen. Betrachtet man eine feste Zustandssequenz

$$\mathbf{q} = (q_1, q_2, \dots, q_T), \quad (5.58)$$

wobei der Zustand q_1 der Initialisierungszustand ist, so kann die Wahrscheinlichkeit für die Beobachtungssequenz \mathbf{O} bei gegebener Zustandssequenz \mathbf{q} mit (5.55) als

$$P(\mathbf{O}|\mathbf{q}) = \prod_{t=1}^T P(\mathbf{o}_t|q_t) = b_{q_1}(\mathbf{o}_1) \cdot b_{q_2}(\mathbf{o}_2) \cdot \dots \cdot b_{q_T}(\mathbf{o}_T) = \prod_{t=1}^T b_{q_t}(\mathbf{o}_t) \quad (5.59)$$

geschrieben werden, wobei eine statistische Unabhängigkeit der einzelnen Beobachtungen vorausgesetzt ist.

Die Wahrscheinlichkeit für eine Zustandssequenz \mathbf{q} kann mit (5.52) und (5.56) als

$$P(\mathbf{q}) = \pi_{q_1} \cdot a_{q_1 q_2} \cdot a_{q_2 q_3} \cdot \dots \cdot a_{q_{T-1} q_T} \quad (5.60)$$

geschrieben werden. Die Verbundwahrscheinlichkeit von \mathbf{O} und \mathbf{q} , d.h. die Wahrscheinlichkeit, daß die Beobachtungssequenz \mathbf{O} und die Zustandssequenz \mathbf{q} gleichzeitig auftreten, ergibt sich zu

$$P(\mathbf{O}, \mathbf{q}) = P(\mathbf{O}|\mathbf{q}) \cdot P(\mathbf{q}). \quad (5.61)$$

Die gesuchte Wahrscheinlichkeit für das Auftreten der gegebenen Beobachtungssequenz \mathbf{O} ergibt sich nun aus der Summe der Verbundwahrscheinlichkeiten von \mathbf{O} und allen möglichen Zustandssequenzen \mathbf{q} :

$$P(\mathbf{O}) = \sum_{\text{alle } \mathbf{q}} P(\mathbf{O}|\mathbf{q}) P(\mathbf{q}) \quad (5.62)$$

$$= \sum_{q_1, \dots, q_T} \pi_{q_1} b_{q_1}(\mathbf{o}_1) a_{q_1 q_2} b_{q_2}(\mathbf{o}_2) \cdot \dots \cdot a_{q_{T-1} q_T} b_{q_T}(\mathbf{o}_T). \quad (5.63)$$

Der Aufwand zur direkten Bestimmung von $P(\mathbf{O})$ nach Gleichung (5.63) ist allerdings enorm, da zu jedem Zeitpunkt N mögliche Zustände erreicht werden können und für jede Zustandssequenz $2T$ Berechnungen notwendig sind. Zur effizienten Berechnung der Wahrscheinlichkeit einer Beobachtungssequenz bedient man sich deshalb eines Trellis-Diagramms, in dem alle Pfade enthalten sind, auf denen das Hidden-Markov-Modell durchlaufen werden kann. Bild 5.19 zeigt ein Trellis-Diagramm für $N = 6$ Zustände und $T = 4$ Beobachtungen.

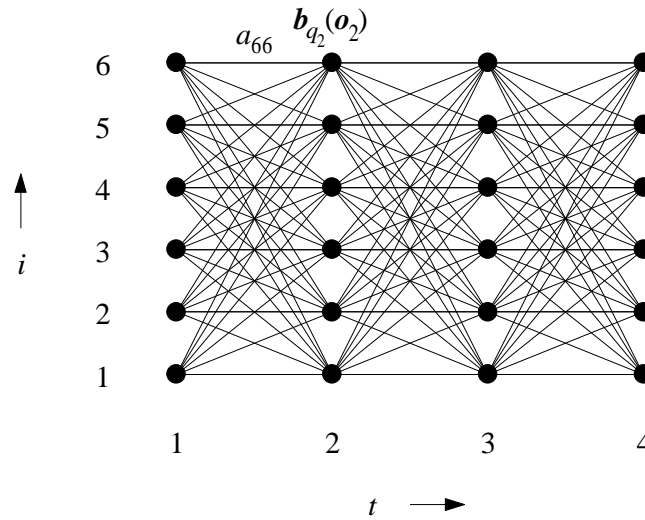


Bild 5.19: Beispiel eines Trellis-Diagramms für die Beobachtungszeitpunkte t und die Zustände i mit $T = 4$ und $N = 6$

Die vereinfachte Berechnung im Trellis erfolgt mit einer sogenannten *Vorwärts-Prozedur* [66]. Betrachtet wird dabei eine Vorwärts-Variable $\alpha_t(i)$, die die Wahrscheinlichkeit angibt, daß bis zum Zeitpunkt t unter Erzeugung der Teil-Beobachtungssequenz $\mathbf{o}_1, \mathbf{o}_2, \dots, \mathbf{o}_t$ der Zustand i erreicht worden ist:

$$\alpha_t(i) = P(\mathbf{o}_1, \mathbf{o}_2, \dots, \mathbf{o}_t, q_t = i). \quad (5.64)$$

Die Berechnung der Größe $\alpha_t(i)$ erfolgt iterativ:

- Initialisierung

$$\alpha_1(i) = \pi_i b_i(\mathbf{o}_1), \quad 1 \leq i \leq N. \quad (5.65)$$

- Induktion

$$\alpha_{t+1}(j) = \left[\sum_{i=1}^N \alpha_t(i) a_{ij} \right] b_j(\mathbf{o}_{t+1}), \quad \begin{array}{l} 1 \leq t \leq T-1, \\ 1 \leq j \leq N. \end{array} \quad (5.66)$$

- Ende

$$P(\mathbf{O}) = \sum_{i=1}^N \alpha_T(i). \quad (5.67)$$

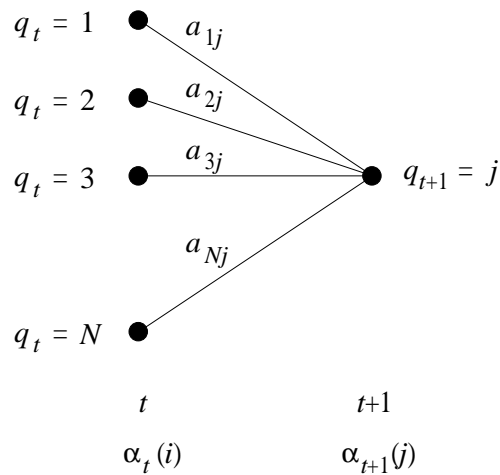


Bild 5.20: Berechnung der Vorwärtswahrscheinlichkeit $\alpha_{t+1}(j)$

Der erste Schritt initialisiert die Vorwärts-Wahrscheinlichkeit als Verbundwahrscheinlichkeit des Zustandes i und der Beobachtung \mathbf{o}_1 . Der Induktionsschritt wird anhand von Bild 5.20 verdeutlicht. Beim Übergang $t \rightarrow t + 1$ kann der Zustand j von N möglichen Zuständen i , $1 \leq i \leq N$, aus erreicht werden. Da $\alpha_t(i)$ als Wahrscheinlichkeit dafür, daß der Zustand i bei gegebener Sequenz $\mathbf{o}_1, \dots, \mathbf{o}_t$ erreicht worden ist, bereits bekannt ist, gibt das Produkt $\alpha_t(i) a_{ij}$ die Wahrscheinlichkeit dafür an, daß der Zustand j zum Zeitpunkt $t + 1$ vom Zustand i aus erreicht worden ist. Die Summe dieser Produkte über alle N möglichen Zustände ergibt die Wahrscheinlichkeit dafür, daß der Zustand j bei gegebener Beobachtungssequenz $\mathbf{o}_1, \dots, \mathbf{o}_t$ erreicht worden ist. Die Multiplikation mit $b_j(\mathbf{o}_{t+1})$ schließt jetzt die Beobachtung \mathbf{o}_{t+1} zum aktuellen Zeitpunkt mit ein, so daß sich hieraus die neue Vorwärts-Variable $\alpha_{t+1}(j)$ ergibt.

Nach $T - 1$ Iterationsschritten erhält man die Variable

$$\alpha_T(i) = P(\mathbf{o}_1 \mathbf{o}_2 \dots \mathbf{o}_T, q_T = i). \quad (5.68)$$

Die Summation über alle Zustände ergibt dann nach Gleichung (5.67) die gesuchte Wahrscheinlichkeit $P(\mathbf{O})$.

Die Anzahl der Rechenoperationen reduziert sich durch diese iterative Berechnung drastisch, da die Wahrscheinlichkeiten für alle Wege durch

einen Zustand i in Form von $\alpha_t(i)$ zusammengefaßt werden.

In gleicher Weise wie die Vorwärts-Variable $\alpha_t(i)$ kann auch eine Rückwärts-Variable $\beta_t(i)$ definiert werden [66].

Der Viterbi-Algorithmus

Aus den unbekanntem Zustandssequenzen $\mathbf{q} = (q_1, \dots, q_T)$ muß die *optimale* Sequenz ermittelt werden, die am wahrscheinlichsten zur gegebenen Beobachtungssequenz $\mathbf{O} = (\mathbf{o}_1, \dots, \mathbf{o}_T)$ gehört, d.h. deren Wahrscheinlichkeit $P(\mathbf{q}|\mathbf{O})$ maximal ist. Diese Maximierung kann effizient mit der Methode der *dynamischen Programmierung* durchgeführt werden, was in stochastischen Anwendungen gleichbedeutend mit dem Viterbi-Algorithmus ist [80], [35].

Der Viterbi-Algorithmus ist der Vorwärts-Prozedur sehr ähnlich. Der wesentliche Unterschied liegt in der Maximierung anstelle der Summierung über alle Zustände nach Gleichung (5.66). Hierzu wird die Größe

$$\delta_t(i) = \max_{q_1, \dots, q_t} P(q_1, \dots, q_{t-1}, q_t = i, \mathbf{o}_1, \dots, \mathbf{o}_t) \quad (5.69)$$

eingeführt. $\delta_t(i)$ gibt für die ersten t Beobachtungen die maximale Wahrscheinlichkeit entlang eines Pfades an, der zum Zeitpunkt t im Zustand i endet. Durch Induktion ergibt sich für den nächsten Zeitschritt der Wert

$$\delta_{t+1}(j) = [\max_i \delta_t(i)] \cdot b_j(\mathbf{o}_{t+1}). \quad (5.70)$$

Zur Bestimmung der Zustandssequenz müssen noch die Argumente $\psi_t(i)$ gespeichert werden, die Gleichung (5.70) maximieren. Der Viterbi-Algorithmus hat damit folgenden Ablauf:

- Initialisierung

$$\delta_1(i) = \pi_i b_i(\mathbf{o}_1), \quad (5.71)$$

$$\psi_1(i) = 0, \quad 1 \leq i \leq N. \quad (5.72)$$

- Rekursion

$$\delta_t(j) = \max_{1 \leq i \leq N} [\delta_{t-1} a_{ij}] b_j(\mathbf{o}_t), \quad (5.73)$$

$$\psi_t(j) = \arg \max_{1 \leq i \leq N} [\delta_{t-1}(i) a_{ij}], \quad \begin{array}{l} 2 \leq t \leq T, \\ 1 \leq j \leq N. \end{array} \quad (5.74)$$

- Ende

$$P^* = \max_{1 \leq i \leq N} [\delta_T(i)] \quad (5.75)$$

$$q_T^* = \arg \max_{1 \leq i \leq N} [\delta_T(i)]. \quad (5.76)$$

- Rückrechnung des optimalen Pfades der Zustandssequenzen

$$q_t^* = \psi_{t+1}(q_{t+1}^*), \quad t = T - 1, T - 2, \dots, 1. \quad (5.77)$$

Die Berechnung der Summenpfadkosten zur Bestimmung des optimalen Pfades nach den Gleichungen (5.71) bis (5.77) geschieht vorwiegend durch Multiplikationen. In der Praxis werden die vorgeschriebenen Berechnungen mit logarithmierten Werten durchgeführt, so daß anstelle der Multiplikationen nur noch Additionen benötigt werden. Effektive Algorithmen zur Implementierung des Viterbi-Algorithmus finden sich in [53].

Ein Problem ergibt sich aus der Länge der abzuspeichernden Zustandssequenz, die sich in jedem Iterationsschritt um eins erhöht, so daß sich bei fortlaufender Berücksichtigung neuer Beobachtungssymbole ein akkumulierender Speicherplatzbedarf ergeben würde. In der Praxis zeigt es sich allerdings, daß die von einem Zeitpunkt t aus zurückverfolgten sogenannten überlebenden Pfade sich nach einer endlichen Anzahl t_{max} an Zeitschritten vereinigen und die den N Zuständen zugeordneten Beobachtungssymbole ab der Speicherzelle t_{max} identische Werte aufweisen. Somit kann zum Zeitschritt t eine Entscheidung über den Zustand $q_{t-t_{max}}$ getroffen werden [44].

Klassifizierung von Hidden-Markov-Modellen

Hidden-Markov-Modelle können anhand der Struktur der Transitionsmatrix \mathbf{A} klassifiziert werden. Können von einem Zustand i zu einem Zeitpunkt $t = 1, \dots, T-1$ im nächsten Zeitschritt alle Zustände $j = 1, \dots, N$ erreicht werden, so ist die Transitionsmatrix \mathbf{A} voll besetzt. Das zugehörige Hidden-Markov-Modell heißt in diesem Fall ergodisch. Ein Beispiel für ein derartiges Modell ist in Bild 5.21a für $N = 4$ dargestellt.

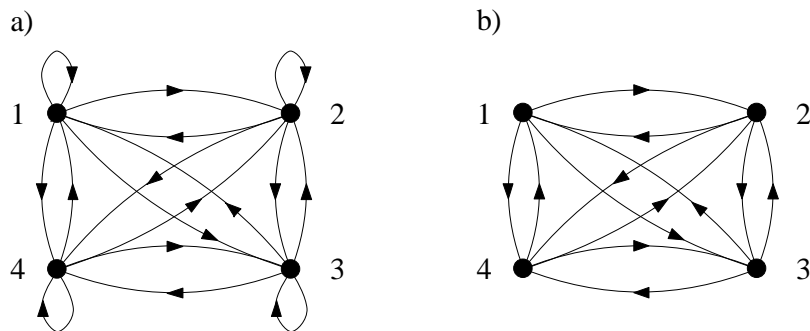


Bild 5.21: Beispiele für Hidden-Markov-Modelle mit jeweils 4 Zuständen: a) Ergodisches Modell, b) Nicht-ergodisches Modell

In vielen Anwendungen treten nicht-ergodische Modelle auf, in denen bestimmte Zustandsübergänge nicht erlaubt sind und die Transitionsmatrix nicht mehr voll besetzt ist. Ein Beispiel hierfür ist in Bild 5.21b dargestellt. Die Koeffizienten der Transitionsmatrix des zugehörigen Hidden-Markov-Modells haben die Eigenschaft, daß

$$a_{ij} = 0 \quad \text{für} \quad i = j \quad (5.78)$$

ist. In diesem Fall muß sich beim Übergang von einem Zeitschritt t zum nächsten Zeitschritt $t+1$ der Zustand des Systems ändern. Ein derartiges Modell wird zur streckenbezogenen Verkehrsdatenerfassung eingesetzt.

Beobachtungssymbole mit kontinuierlichen Wahrscheinlichkeitsdichtefunktionen

Die bislang durchgeführten Betrachtungen bezogen sich ausschließlich auf Beobachtungssequenzen mit diskreten Beobachtungssymbolen aus einem begrenzten Wertevorrat der Größe M . Für diesen Fall konnte für jeden Zustand j zu einem beobachteten Symbol \mathbf{v}_k eine diskrete Wahrscheinlichkeitsdichte angegeben werden. In vielen Anwendungsfällen bestehen die Beobachtungen allerdings aus kontinuierlichen Signalen oder Vektoren. In diesen Fällen besteht zum einen die Möglichkeit, solche Signale mittels Vektorquantisierung einem diskreten Prototypen zuzuordnen. Die andere Möglichkeit liegt in der Modellierung der Wahrscheinlichkeitsdichtefunktionen der Vektoren durch parametrische Funktionen.

Der einfachste Fall liegt vor, wenn die Wahrscheinlichkeitsdichtefunktion der kontinuierlichen Signale oder Vektoren durch eine Normalverteilung angenähert werden kann. Die Wahrscheinlichkeitsdichte im Zustand j kann dann mit

$$b_j(\mathbf{o}) = \frac{1}{[2\pi]^{n/2} |\mathbf{R}_j|^{1/2}} \exp\left(-\frac{1}{2}[\mathbf{o} - \boldsymbol{\mu}_j]^T \mathbf{R}_j^{-1} [\mathbf{o} - \boldsymbol{\mu}_j]\right) \quad (5.79)$$

beschrieben werden. \mathbf{R}_j ist die Kovarianzmatrix und $\boldsymbol{\mu}_j$ der Mittelwertvektor der Beobachtungsvektoren \mathbf{o} im Zustand j .

Besitzt der Prozeß eine Einheitsmatrix \mathbf{I} als Kovarianzmatrix, so vereinfacht sich Gleichung (5.79) zu

$$b_j(\mathbf{o}) = \frac{1}{[2\pi]^{n/2}} \exp\left(-\frac{1}{2}[\mathbf{o} - \boldsymbol{\mu}_j]^T [\mathbf{o} - \boldsymbol{\mu}_j]\right). \quad (5.80)$$

Die Konstante $([2\pi]^{n/2})^{-1}$ und der Faktor $-1/2$ sind unabhängig vom Zustand j . Die Erzeugungswahrscheinlichkeitsdichte eines Beobachtungsvektors kann somit auch durch das Quadrat des euklidischen Abstands

$$d^2(\mathbf{o}, \boldsymbol{\mu}_j) = (\mathbf{o} - \boldsymbol{\mu}_j)^T (\mathbf{o} - \boldsymbol{\mu}_j) \quad (5.81)$$

ausgedrückt werden. Diese Vereinfachung kann zur streckenbezogenen Verkehrsdatenerfassung durchgeführt werden.

5.4.2 Der Schätzalgorithmus

Die Idee dieses Schätzverfahrens liegt darin, die am stromabwärtigen Meßquerschnitt erfaßten Merkmalsvektoren einer Fahrzeuggruppe als Beobachtungsvektoren eines Hidden-Markov-Modells anzusehen. Die aufgrund der Vermischung des Verkehrs im Streckenabschnitt unbekannt Positionen dieser Fahrzeuggruppe am stromaufwärtigen Meßquerschnitt entsprechen der unbekannt Zustandssequenz des Modells. Die Anzahl der Fahrzeuge im Streckenabschnitt bestimmt die Anzahl der Zustände. Mit dem Viterbi-Algorithmus kann ausgehend von der Beobachtungssequenz eine optimale Zustandssequenz und damit die wahrscheinlichste Fahrzeuggruppe geschätzt werden. Aus dieser Schätzung können dann die streckenbezogenen Verkehrsdaten der einzelnen Fahrzeuge bestimmt werden.

Die im letzten Abschnitt definierten Parameter eines Hidden-Markov-Modells sind in diesem Verfahren folgendermaßen belegt:

Zustände $Q = \{q_1, q_2, \dots, q_T\}$

Die Position der Fahrzeuge am stromaufwärtigen Meßquerschnitt wird durch die Zustände beschrieben.

Anzahl der Zustände N

Die Anzahl der Zustände entspricht der Fahrzeuganzahl im Streckenabschnitt.

Aktueller Zustand q_t

Die Position des am zweiten Meßquerschnitt detektierten Fahrzeugs beim Eintritt in den Streckenabschnitt entspricht dem aktuellen Zustand.

Aktueller Beobachtungszeitpunkt t

Die Position des aktuell detektierten Fahrzeugs am stromabwärtigen Meßquerschnitt innerhalb der betrachteten Fahrzeuggruppe ist gleichbedeutend mit dem aktuellen Beobachtungszeitpunkt t , der bei diesem Modell somit als diskreter Wert verstanden wird.

Beobachtungssequenz $\mathbf{O} = (\mathbf{o}_1, \mathbf{o}_2, \dots, \mathbf{o}_T)$

Die Beobachtungssequenz setzt sich aus den Merkmalsvektoren $\mathbf{v}_{2,t}$ der betrachteten Fahrzeuggruppe vom zweiten Meßquerschnitt zusammen.

Anzahl der Objekte T

Die Größe der untersuchten Fahrzeuggruppe am zweiten Meßquerschnitt entspricht der Anzahl der Objekte des Modells.

Übergangswahrscheinlichkeiten A

Die Übergangswahrscheinlichkeiten von einem Zustand i zum Zeitpunkt t in einen Zustand j zum Zeitpunkt $t + 1$

$$a_{ij} = P(q_{t+1} = j | q_t = i), \quad 1 \leq i, j \leq N, \quad (5.82)$$

ergeben sich aus der Vermischung der Fahrzeuge im Streckenabschnitt.

Die Zustandsübergänge sollen an einem Beispiel mit Bild 5.22 verdeutlicht werden. Dargestellt sind zwei Möglichkeiten für einen Zustandsübergang anhand der Positionen von drei Fahrzeugen am ersten und am zweiten Meßquerschnitt. Zum Zeitpunkt t wird in Bild 5.22a am zweiten Meßquerschnitt ein LKW detektiert, dessen Position am ersten Meßquerschnitt mit dem Zustand i bezeichnet werden soll. Anschließend erfolgt zum Zeitpunkt $t+1$ die Detektion des weißen PKWs, dessen Position bzw. Zustand $j = i + 1$ ist, da er am ersten Meßquerschnitt unmittelbar hinter dem LKW in den Streckenabschnitt eingefahren ist. Im zweiten Fall wird am zweiten Meßquerschnitt zum Zeitpunkt $t + 1$ ein grauer PKW detektiert, dessen Zustand $j = i + 2$ ist, da er am ersten Meßquerschnitt zwei Fahrzeuge hinter dem LKW detektiert wurde.

Aus diesen Überlegungen ergibt sich ein nicht-ergodisches Modell nach Bild 5.21b. Da Mehrfachzuweisungen von Fahrzeugen unterdrückt werden sollen, ist ein Verbleiben des Systems in einem Zustand nicht erlaubt und nach Gleichung (5.78) ist $a_{ij} = 0$ für $i = j$. Die Beschränkung auf ein Markov-Modell 1. Ordnung führt allerdings dazu, daß ein Übergang in einen schon einmal belegten Zustand nach mehr als einem Zeitschritt nicht unterbunden werden kann.

Als Transitionsmatrix ergibt sich eine Bandmatrix mit Nullelementen

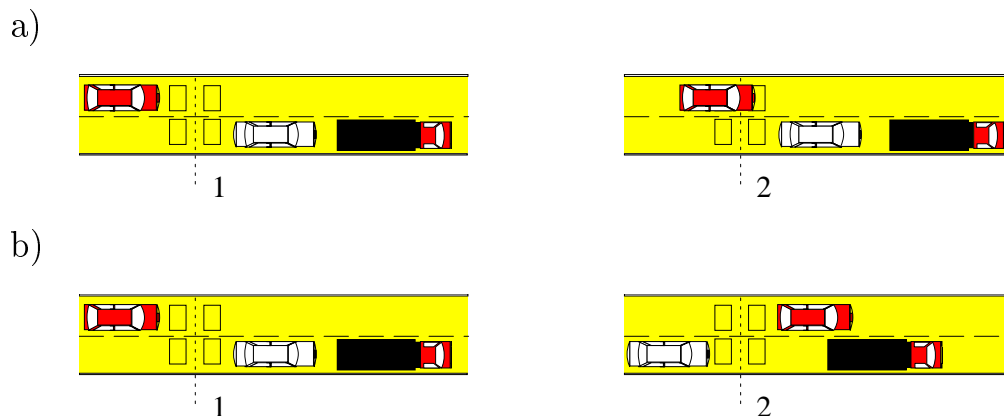


Bild 5.22: Beispiel für den Übergang vom Zustand i in den Zustand j :
a) $j = i + 1$, b) $j = i + 2$

auf der Hauptdiagonalen:

$$\mathbf{A} = \begin{bmatrix} 0 & a_1 & \cdots & a_{N-1} \\ a_{-1} & 0 & \ddots & \vdots \\ \vdots & \ddots & 0 & a_1 \\ a_{-N+1} & \cdots & a_{-1} & 0 \end{bmatrix}. \quad (5.83)$$

Die Übergangswahrscheinlichkeiten müssen aus einer Lernstichprobe bestimmt werden. Hierauf wird in Abschnitt 5.4.3 noch näher eingegangen.

Erzeugungswahrscheinlichkeiten \mathbf{B}

Zum Vergleich eines Merkmalsvektors $\mathbf{v}_{2,j}$ mit den Merkmalsvektoren vom ersten Meßquerschnitt $\mathbf{v}_{1,i}$ müssen kontinuierliche Wahrscheinlichkeitsdichtefunktionen verwendet werden. Nach Abschnitt 5.1.3 handelt es sich bei dem Prozeß der Merkmalsvektorendifferenzen um einen weißen Prozeß, so daß die Erzeugungswahrscheinlichkeitsdichte mit Gleichung (5.80) angegeben werden kann. Hierbei sind der Beobachtungsvektor mit $\mathbf{o} = \mathbf{v}_{2,j}$ und die Mittelwertsvektoren mit $\boldsymbol{\mu}_i = \mathbf{v}_{1,i}$ gegeben und die

Wahrscheinlichkeitsdichte lautet:

$$b_i(\mathbf{o}) = \frac{1}{[2\pi]^{n/2}} \exp\left(-\frac{1}{2}[\mathbf{v}_{2,j} - \mathbf{v}_{1,i}]^T [\mathbf{v}_{2,j} - \mathbf{v}_{1,i}]\right). \quad (5.84)$$

Bei der Berechnung der Summenpfadkosten zur Ermittlung der optimalen Zustandssequenz mit dem Viterbi-Algorithmus in logarithmierter Form ergibt sich die Funktion

$$\begin{aligned} B_i(\mathbf{o}) = \ln[b_i(\mathbf{o})] &= -\ln([2\pi]^{n/2}) - \frac{1}{2}[\mathbf{v}_{2,j} - \mathbf{v}_{1,i}]^T [\mathbf{v}_{2,j} - \mathbf{v}_{1,i}] \\ &= -\ln([2\pi]^{n/2}) - \frac{1}{2}d^2(\mathbf{v}_{2,j}, \mathbf{v}_{1,i}). \end{aligned} \quad (5.85)$$

Die Konstante $\ln([2\pi]^{n/2})$ ist ebenso wie der Vorfaktor $1/2$ unabhängig vom Zustand i und braucht deshalb beim Viterbi-Algorithmus nicht berücksichtigt zu werden.

Wahrscheinlichkeiten für die Anfangszustände π

Die Wahrscheinlichkeiten für die Anfangszustände $\pi = \{\pi_i\}$ zum Zeitpunkt $t = 1$ werden mit

$$\pi_i = P(q_1 = i) = 1/N, \quad 1 \leq i \leq N, \quad (5.86)$$

als gleich angenommen. Damit ist auch diese Größe unabhängig vom Zustand i und bleibt ebenso unberücksichtigt.

Das Schätzverfahren ist vereinfacht in Bild 5.23 dargestellt. Das vorliegende Beispiel wird in Abschnitt 5.4.4 noch näher erläutert.

5.4.3 Optimierung des Verfahrens

Für verschiedenartige Streckenabschnitte und Verkehrszusammensetzungen muß ein entsprechendes Modell mit den Parametern $\lambda = (\mathbf{A}, \mathbf{B}, \pi)$ festgelegt werden. Die in der Literatur beschriebenen Verfahren zur Parameteroptimierung ([6], [29], [50], [43]) brauchen bei diesem Modell nicht

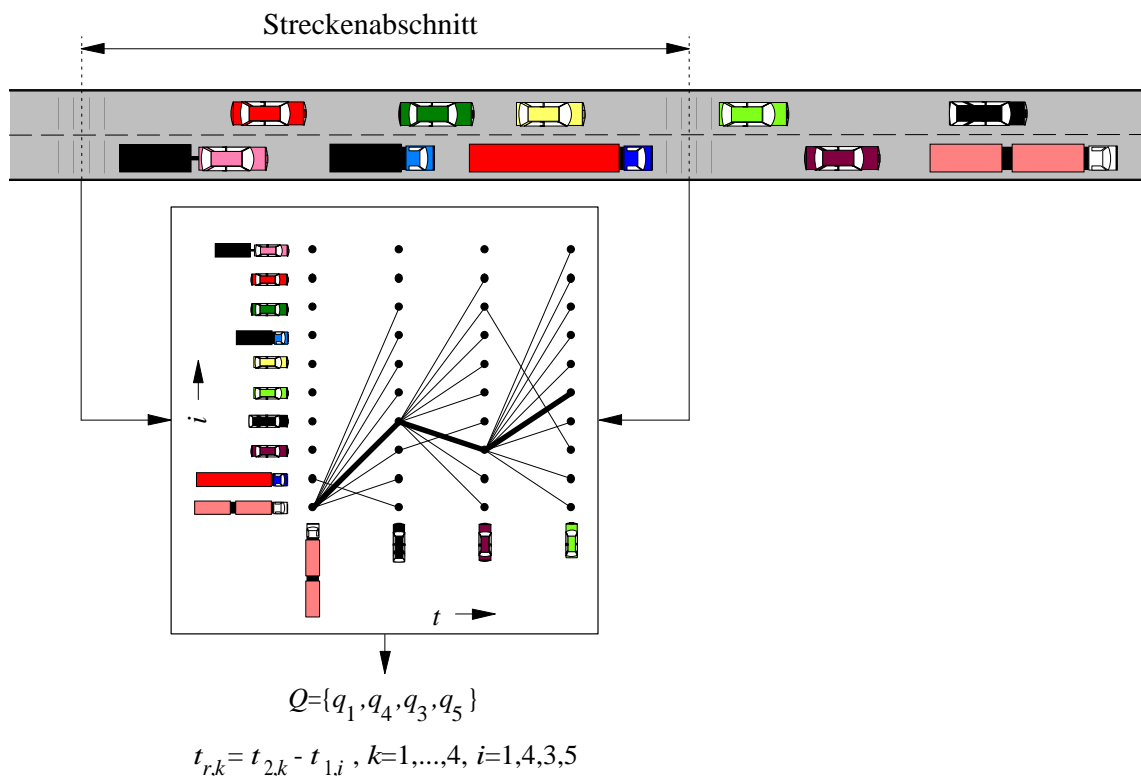


Bild 5.23: Beispiel zur streckenbezogenen Verkehrsdatenerfassung mit Hidden-Markov-Modellen

angewendet werden, da die Parameter der Wahrscheinlichkeitsdichtefunktion (5.77) mit den Merkmalsvektoren vom ersten Meßquerschnitt gegeben sind bzw. durch die Whitening-Transformation festgelegt sind. Die Anfangswahrscheinlichkeiten werden a priori als gleich angenommen. Deshalb wird bei diesem Modell die Transitionsmatrix direkt aus einer umfassenden Lernstichprobe geschätzt, in der Zustandsübergänge aus verschiedenen Verkehrsabläufen beinhaltet sind.

Schätzung der Übergangswahrscheinlichkeiten

Für die folgenden Betrachtungen wird ein zweispuriger Streckenabschnitt mit einer Länge von 2,5 km betrachtet. Als Grundlage zur Bestimmung der Übergangswahrscheinlichkeiten dienen mikroskopische Verkehrssimulationen, mit denen verschiedene Verkehrszustände berücksichtigt werden

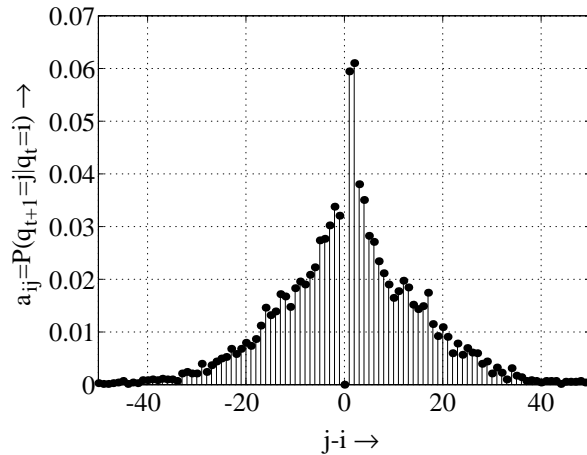


Bild 5.24: Aus einer Stichprobe ermittelte Wahrscheinlichkeitsverteilung für die Übergänge vom Zustand i zu einem Zeitpunkt t in einen Zustand j zu einem Zeitpunkt $t + 1$

können. Bild 5.24 zeigt die aus einer Stichprobe von ca. 8000 Fahrzeugen ermittelten Übergangswahrscheinlichkeiten a_{ij} von einem Zustand i zu einem Zeitpunkt t in einen Zustand j zu einem Zeitpunkt $t + 1$ für den Verkehrsablauf in dem betrachteten Streckenabschnitt.

Die wahrscheinlichsten Übergänge vom Zustand i in den Zustand j liegen für $j = i + 1$ und $j = i + 2$ vor. Dieses entspricht den Beispielen in Bild 5.22. Somit werden zwei am ersten Meßquerschnitt kurz hintereinander detektierte Fahrzeuge mit einer hohen Wahrscheinlichkeit auch am zweiten Meßquerschnitt wieder zusammen detektiert.

Bestimmung des optimalen Pfades mit Viterbi-Detektion

Für eine vorliegende Beobachtungssequenz $\mathbf{o}_1, \dots, \mathbf{o}_T$ wird die optimale zugrundeliegende Zustandssequenz q_1, \dots, q_T gesucht. Im Fall der Fahrzeugwiedererkennung bedeutet dieses, daß ausgehend von der Beobachtungssequenz der Merkmalsvektoren $\mathbf{v}_{2,1}, \dots, \mathbf{v}_{2,T}$ vom zweiten Meßquerschnitt bei Kenntnis der Merkmalsvektoren $\mathbf{v}_{1,1}, \dots, \mathbf{v}_{1,N}$ vom ersten Meßquerschnitt aller sich zum Zeitpunkt T im Streckenabschnitt befindlichen

Fahrzeuge diejenige Fahrzeuggruppe aus den N Fahrzeugen ausgewählt werden soll, deren Wahrscheinlichkeit $P(\mathbf{q}|\mathbf{O})$ am größten ist.

Die optimale Zustandssequenz wird nach Abschnitt 5.4.1 mit dem Viterbi-Algorithmus ermittelt. Bei der Initialisierung des Verfahrens werden die euklidischen Distanzen zwischen dem Merkmalsvektor $\mathbf{v}_{2,1}$ des ersten am stromabwärtigen Meßquerschnitt detektierten Fahrzeugs mit allen Merkmalsvektoren $\mathbf{v}_{1,i}$ der bis dahin am ersten Meßquerschnitt erfaßten Fahrzeuge ermittelt. Die Initialisierung erfolgt in logarithmierter Form mit

$$\Delta_1(i) = B_i(\mathbf{o}_1), \quad (5.87)$$

$$\Psi_1(i) = 0, \quad 1 \leq i \leq N. \quad (5.88)$$

Anschließend folgt für die nachfolgenden Fahrzeugdetektionen am zweiten Meßquerschnitt mit

$$A_{ij} = \ln[a_{ij}], \quad 1 \leq i, j \leq N, \quad (5.89)$$

der Rekursionsschritt

$$\Delta_t(j) = \max_{1 \leq i \leq N} [\Delta_{t-1}(i) + A_{ij}] + B_j(\mathbf{o}_t), \quad (5.90)$$

$$\Psi_t(j) = \arg \max_{1 \leq i \leq N} [\Delta_{t-1}(i) + A_{ij}], \quad \begin{array}{l} 2 \leq t < T, \\ 1 \leq j \leq N. \end{array} \quad (5.91)$$

Am Ende der Rekursion ergeben sich der Logarithmus der maximalen Wahrscheinlichkeit zu

$$P^* = \max_{1 \leq i \leq N} [\Delta_T(i)] \quad (5.92)$$

und der wahrscheinlichste Zustand zu

$$q_T^* = \arg \max_{1 \leq i \leq N} [\Delta_T(i)]. \quad (5.93)$$

Der optimale Pfad der Zustandssequenzen ergibt sich wieder zu

$$q_t^* = \psi_{t+1}(q_{t+1}^*), \quad t = T-1, T-2, \dots, 1. \quad (5.94)$$

Praktisch erfolgt die Viterbi-Detektion der Fahrzeuge wie in Abschnitt 5.4.1 beschrieben mit unbegrenzten Beobachtungssequenzen. In diesem Fall muß die Anzahl an Zeitschritten, t_{max} , festgelegt werden, bei der sich die überlebenden Pfade vereinigen. Günstigerweise wählt man die Länge der Beobachtungssequenz zu $T = t_{max} + 1$, so daß zum Zeitpunkt $t = 1$ alle Pfade vereinigt sind.

Zur Bestimmung des Fahrzeugs mit der Beobachtung \mathbf{o}_1 wird der mit Gleichung (5.94) festgelegte optimale Pfad der Zustandssequenzen bestimmt und das Fahrzeug dem Zustand q_1^* zugeordnet. Anschließend wird die Beobachtung \mathbf{o}_1 aus dem Speicher sofort oder nach einer festen Wartezeit eliminiert. Die Beobachtungssequenz $\mathbf{o}_2, \dots, \mathbf{o}_T$ wird dann der neuen Beobachtungssequenz zum Zeitpunkt $T + 1$ zugeordnet, die um die Beobachtung eines neuen Fahrzeugs \mathbf{o}_{T+1} ergänzt wird. Gleichzeitig erfolgt eine Korrektur der Zustände der verbleibenden Fahrzeuge, da der Zustand q_1 eliminiert wurde. Außerdem werden die Zustände um diejenigen Fahrzeuge erweitert, die im neuen Zeitintervall den ersten Meßquerschnitt passiert haben. Nach dem Übergang $T + 1 \rightarrow T$ und $\mathbf{o}_2, \dots, \mathbf{o}_{T+1} \rightarrow \mathbf{o}_1, \dots, \mathbf{o}_T$ wird wieder mit der Bestimmung des optimalen Zustands q_1^* fortgefahren. Die streckenbezogenen Verkehrsdaten können nach der Wiedererkennung eines Fahrzeugs gemäß den Gleichungen (5.21) und (5.38) bestimmt werden.

Die Zeitpunkte der Pfadvereinigung können allgemein nicht bestimmt werden, sie sind vielmehr abhängig von der betrachteten Verkehrslage sowie der Zusammensetzung des Verkehrs. Für die Untersuchungen mit einem 2,5 km langen Streckenabschnitt kann die minimale Entscheidungsverzögerung mit $t_{max} = 10$ festgelegt werden.

Die Zusammenhänge zwischen der Bestimmung der optimalen Zustandssequenz und der Pfadvereinigung zeigt beispielhaft Bild 5.25. Dargestellt sind bei einer Anzahl von ca. 110 Fahrzeugen im Streckenabschnitt und fortlaufender Fahrzeugdetektion zu den Zeitpunkten $t_0, t_0 + 1, \dots, t_0 + 5$ die jeweils überlebenden Pfade. In den ersten 5 Zeitschritten sind deutlich zwei divergierende Pfade zu erkennen ($t_{max} = 7$ für $t_0 + 4$). Im Zeitschritt $t_0 + 5$ entfällt ein Pfad, da die in ihm enthaltenen Zustände nicht mehr optimal sind.

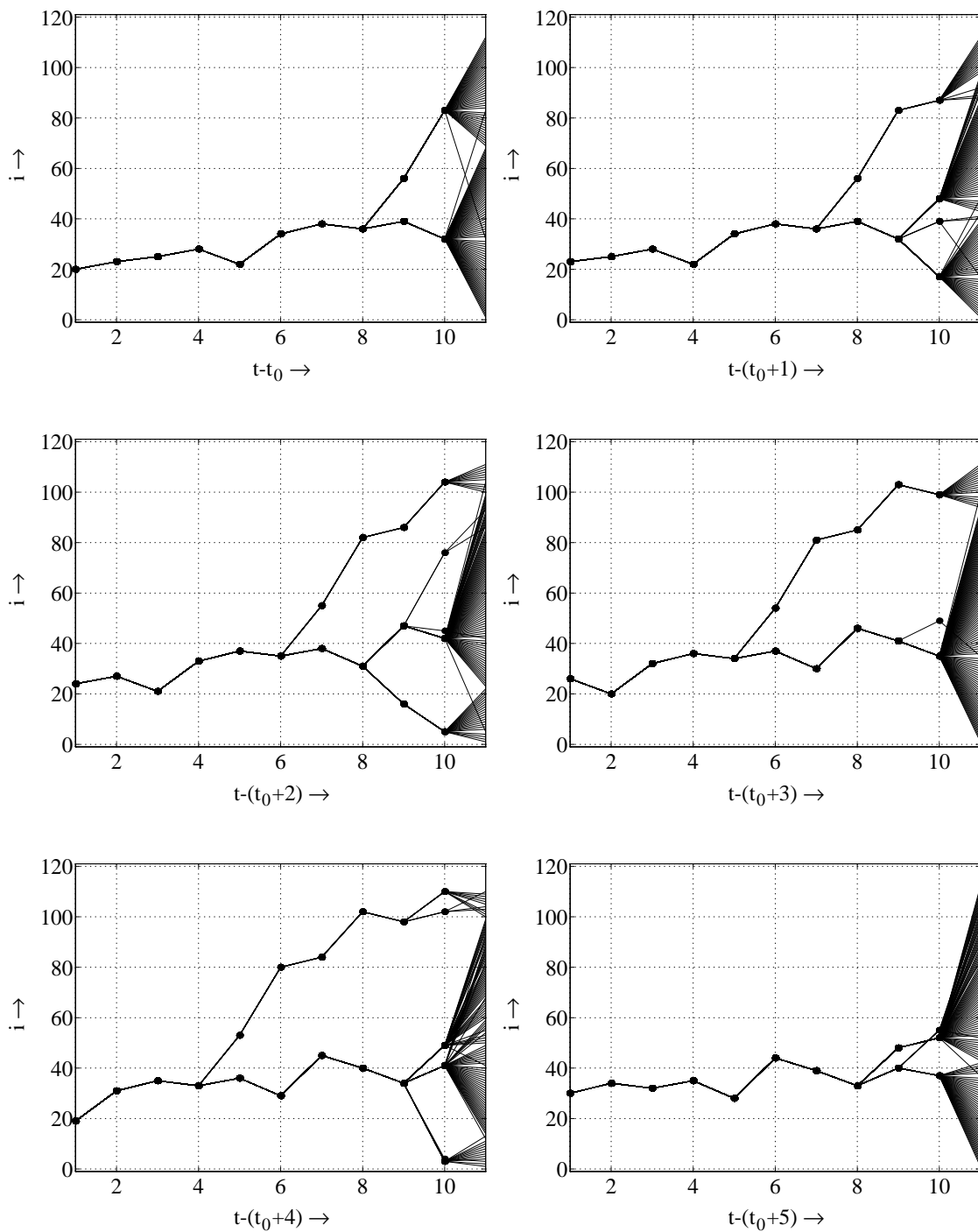


Bild 5.25: Beispiel zur Bestimmung des optimalen Pfades mit einem Gedächtnis von $t_{max} = 10$ Fahrzeugen

5.4.4 Beispiel zur Fahrzeugwiedererkennung

Die Fahrzeugwiedererkennung zur streckenbezogenen Verkehrsdatenerfassung mit Hilfe eines Hidden-Markov-Modells soll nachfolgend anhand eines Beispiels verdeutlicht werden. Betrachtet wird der in Bild 5.26 dargestellte Streckenabschnitt mit den zu 5 verschiedenen Zeitpunkten vorliegenden Verkehrsverhältnissen. Typische Überfahrkurven aller in diesem Beispiel betrachteten Fahrzeuge sind in Bild 5.27 dargestellt.

Zum Zeitpunkt $t = 0$ befinden sich fünf Fahrzeuge im Streckenabschnitt. Ihre Positionen sind anhand der Meßzeitpunkte am stromaufwärtigen Meßquerschnitt festgelegt und entsprechen den Zuständen 1 bis 5. Im Zeitintervall $]0, 1]$ passieren drei Fahrzeuge den ersten Meßquerschnitt.

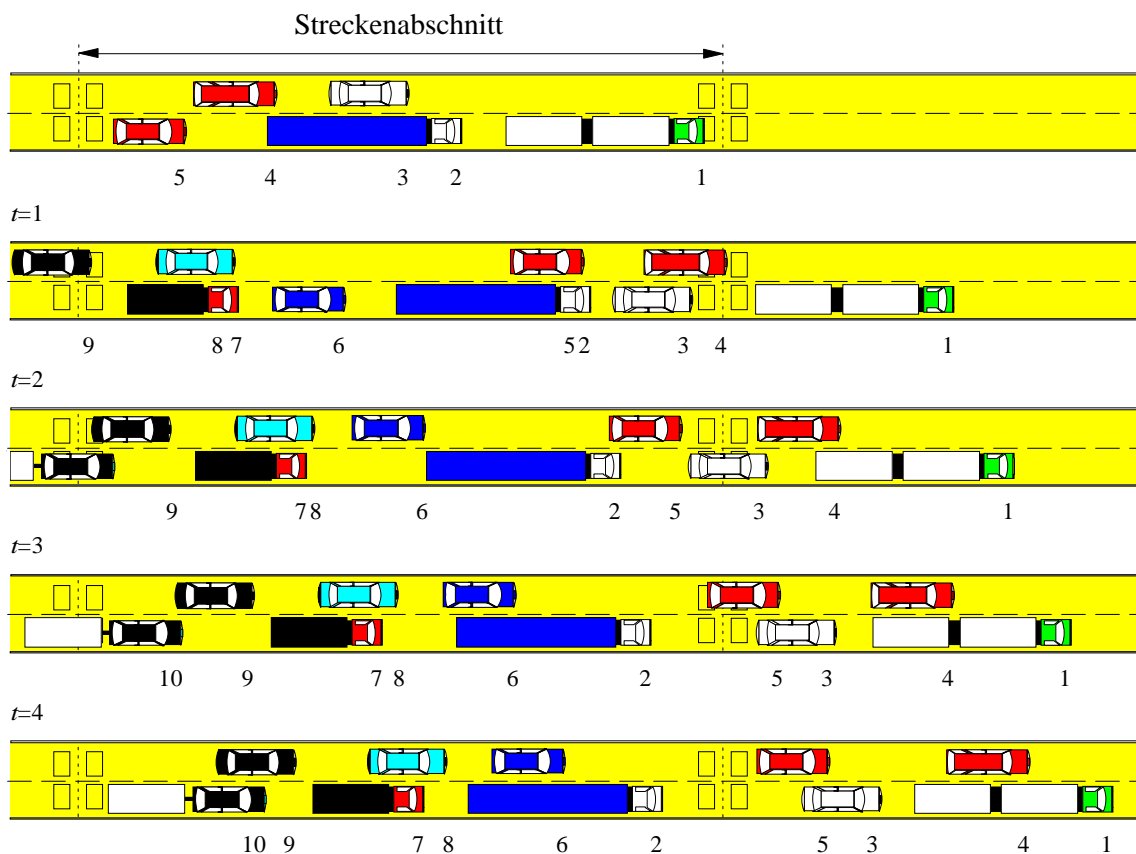


Bild 5.26: Beispiel zur Fahrzeugwiedererkennung mit Hidden-Markov-Modellen

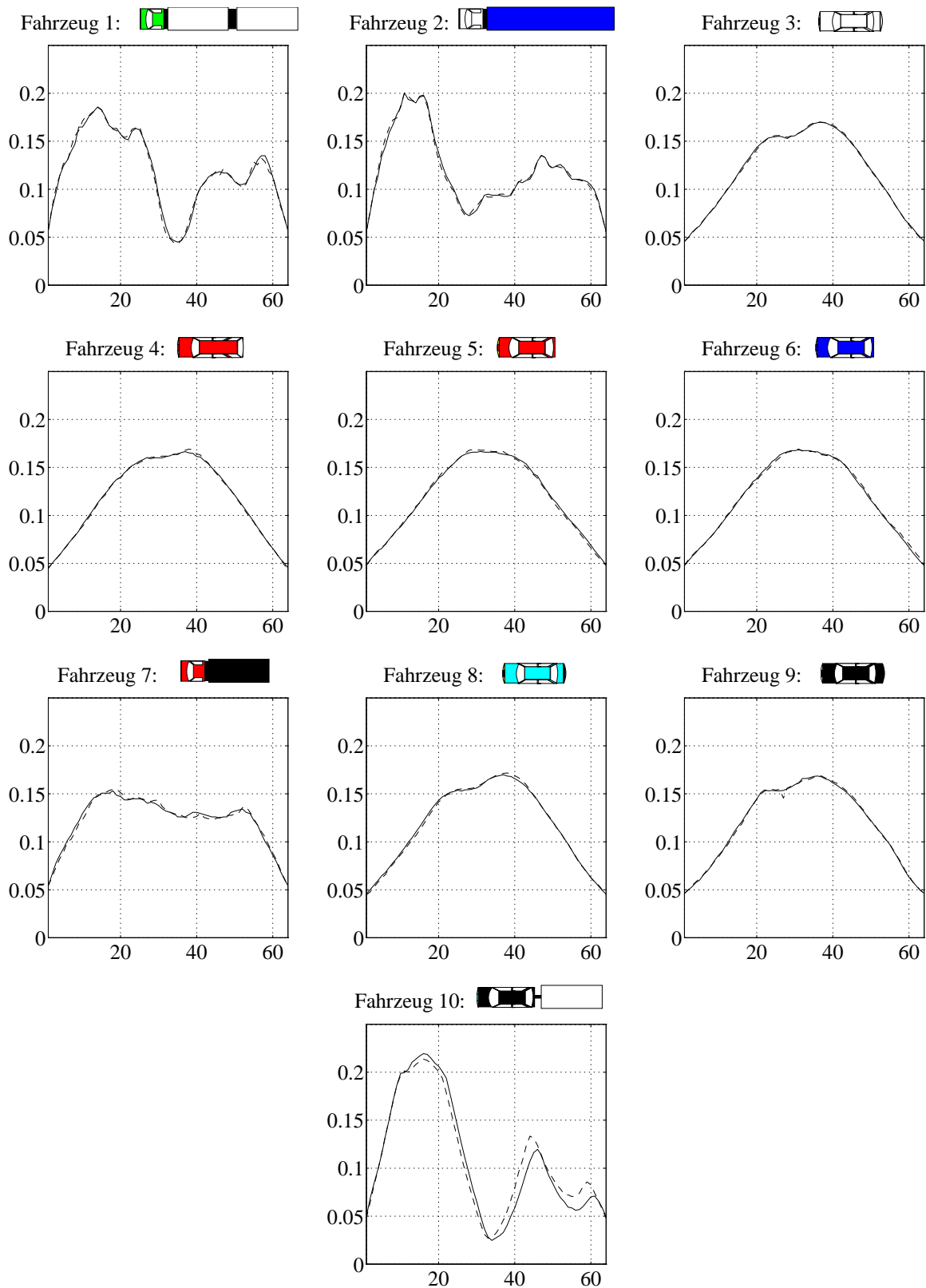


Bild 5.27: Beispielhafte Überfahrkurven an zwei Meßstellen

Ihnen werden die Positionen 6, 7 und 8 zugeordnet. Zum Zeitpunkt $t = 1$ wird am stromabwärtigen Meßquerschnitt der LKW mit Anhänger mit dem noch unbekanntem Zustand $q_1 = 1$ erfaßt. Der von diesem Fahrzeug extrahierte Merkmalsvektor sei $\mathbf{v}_{2,1}$, die Anzahl der Zustände $N = 8$ entspricht der Anzahl der Fahrzeuge im Streckenabschnitt (einschließlich des LKWs mit Anhänger).

Die Erzeugungswahrscheinlichkeiten der einzelnen Zustände können nach Gleichung (5.85) mit den euklidischen Abständen der Merkmalsvektoren $\mathbf{v}_{1,i}$ zum Merkmalsvektor $\mathbf{v}_{2,1}$ ermittelt werden.

Im nächsten Zeitschritt $t = 2$ erfolgt die Detektion des Fahrzeugs mit der Position 4, das im Streckenabschnitt die Fahrzeuge mit den Positionen 2 und 3 überholt hat. Der Merkmalsvektor dieses Fahrzeugs ist $\mathbf{v}_{2,2}$. Nachfolgend werden die Fahrzeuge mit den Positionen 3 zum Zeitpunkt $t = 3$ und 4 zum Zeitpunkt $t = 4$ detektiert. Mittlerweile sind die Fahrzeuge mit der Position 9 im Zeitintervall $]2, 3]$ und mit der Position 10 im Zeitintervall $]3, 4]$ in den Streckenabschnitt eingefahren und vervollständigen die betrachtete Fahrzeuggruppe. Es ergibt sich die Zustandssequenz $q = \{1, 4, 3, 5\}$, die mit dem Viterbi-Algorithmus bestimmt werden muß.

Die logarithmierten Werte der Erzeugungswahrscheinlichkeiten sind

$$\mathbf{B}(\mathbf{o}) = -\frac{1}{2} \begin{bmatrix} d(\mathbf{v}_{2,1}, \mathbf{v}_{1,1}) & d(\mathbf{v}_{2,2}, \mathbf{v}_{1,1}) & \dots & d(\mathbf{v}_{2,4}, \mathbf{v}_{1,1}) \\ \vdots & \vdots & \ddots & \vdots \\ d(\mathbf{v}_{2,1}, \mathbf{v}_{1,8}) & d(\mathbf{v}_{2,2}, \mathbf{v}_{1,8}) & \dots & d(\mathbf{v}_{2,4}, \mathbf{v}_{1,8}) \\ -\infty & -\infty & d(\mathbf{v}_{2,3}, \mathbf{v}_{1,9}) & d(\mathbf{v}_{2,4}, \mathbf{v}_{1,9}) \\ -\infty & -\infty & -\infty & d(\mathbf{v}_{2,4}, \mathbf{v}_{1,10}) \end{bmatrix} \quad (5.95)$$

$$= \begin{bmatrix} -25 & -9846 & -10037 & -10660 \\ -1601 & -7942 & -8030 & -8737 \\ -9531 & -21 & -3 & -63 \\ -9455 & -4 & -15 & -37 \\ -10157 & -39 & -55 & -2 \\ -10195 & -37 & -52 & -2 \\ -5458 & -935 & -1025 & -1180 \\ -9588 & -20 & -4 & -59 \\ -\infty & -\infty & -15 & -63 \\ -\infty & -\infty & -\infty & -8902 \end{bmatrix} \cdot \quad (5.96)$$

Die Erzeugungswahrscheinlichkeiten $b_9(\mathbf{o}_1)$ und $b_{10}(\mathbf{o}_1)$, $b_{10}(\mathbf{o}_2)$ und $b_{10}(\mathbf{o}_3)$ sind Null, da die entsprechenden Fahrzeuge bei den gegebenen Zuordnungen noch nicht im Streckenabschnitt sind und somit unberücksichtigt bleiben müssen.

Mit den logarithmierten Werten der Zustandsübergangswahrscheinlichkeiten ergibt sich die folgende Transitionsmatrix:

$$\mathbf{A} = \begin{bmatrix} -\infty & -2.82 & -2.80 & -3.27 & -3.35 & -3.57 & -3.61 & -3.75 & -3.86 & -3.96 \\ -3.44 & -\infty & -2.82 & -2.80 & -3.27 & -3.35 & -3.57 & -3.61 & -3.75 & -3.86 \\ -3.39 & -3.44 & -\infty & -2.82 & -2.80 & -3.27 & -3.35 & -3.57 & -3.61 & -3.75 \\ -3.50 & -3.39 & -3.44 & -\infty & -2.82 & -2.80 & -3.27 & -3.35 & -3.57 & -3.61 \\ -3.59 & -3.50 & -3.39 & -3.44 & -\infty & -2.82 & -2.80 & -3.27 & -3.35 & -3.57 \\ -3.60 & -3.59 & -3.50 & -3.39 & -3.44 & -\infty & -2.82 & -2.80 & -3.27 & -3.35 \\ -3.80 & -3.60 & -3.59 & -3.50 & -3.39 & -3.44 & -\infty & -2.82 & -2.80 & -3.27 \\ -3.87 & -3.80 & -3.60 & -3.59 & -3.50 & -3.39 & -3.44 & -\infty & -2.82 & -2.80 \\ -3.96 & -3.87 & -3.80 & -3.60 & -3.59 & -3.50 & -3.39 & -3.44 & -\infty & -2.82 \\ -3.93 & -3.96 & -3.87 & -3.80 & -3.60 & -3.59 & -3.50 & -3.39 & -3.44 & -\infty \end{bmatrix}. \quad (5.97)$$

Die Wahrscheinlichkeiten für die Anfangszustände $\boldsymbol{\pi}$ werden als gleich angenommen und deshalb nicht berücksichtigt.

Wird die Erzeugungswahrscheinlichkeit in jedem Zustand maximiert, so ergibt sich mit Gleichung (5.96) allein die korrekte Fahrzeugfolge $q = \{1, 4, 3\}$ für die Zeitpunkte $t = 1, \dots, 3$. Bei der Bestimmung des Zustandes q_4 sind die Erzeugungswahrscheinlichkeiten $b_5(\mathbf{o}_4)$ und $b_6(\mathbf{o}_4)$ beide maximal. Der Grund hierfür liegt darin, daß es sich bei beiden Fahrzeugen um den gleichen Fahrzeugtyp handelt und somit auch ihre extrahierten Merkmalsvektoren ähnlich sind [†].

Bei der Ermittlung der optimalen Zustandssequenz bei der vorliegenden Beobachtungssequenz mit dem Viterbi-Algorithmus ergeben sich folgende Größen:

[†]Dieses verdeutlicht auch Bild 5.27

$$\Delta = \begin{bmatrix} -25 & -11450 & -10073 & -10702 \\ -1601 & -7970 & -8066 & -8779 \\ -9531 & -49 & -39 & -107 \\ -9455 & -33 & -66 & -78 \\ -10157 & -67 & -90 & -43 \\ -10195 & -66 & -87 & -44 \\ -5458 & -964 & -1061 & -1222 \\ -9588 & -49 & -40 & -101 \\ -\infty & -\infty & -51 & -105 \\ -\infty & -\infty & -\infty & -8944 \end{bmatrix}, \Psi = \begin{bmatrix} 2 & 4 & 3 \\ 1 & 4 & 3 \\ 1 & 4 & 8 \\ 1 & 3 & 3 \\ 1 & 4 & 3 \\ 1 & 4 & 3 \\ 1 & 4 & 3 \\ 1 & 4 & 3 \\ & 4 & 3 \\ & & 3 \end{bmatrix}. \quad (5.98)$$

Die maximale Wahrscheinlichkeit $P^* = \max_{1 \leq i \leq N} [\Delta_4(i)] = -43$ führt auf die Zuordnung des Zustandes $q_4^* = \arg \max_{1 \leq i \leq N} [\Delta_4(i)] = 5$. Die Rückrechnung des optimalen Pfades mit $q_k^* = \psi_{k+1}(q_{k+1})$ liefert die gesuchte Zustandssequenz $q^* = \{1, 4, 3, 5\}$. Bild 5.28 zeigt die möglichen Trellispfade sowie den optimalen Pfad für das gegebene Beispiel.

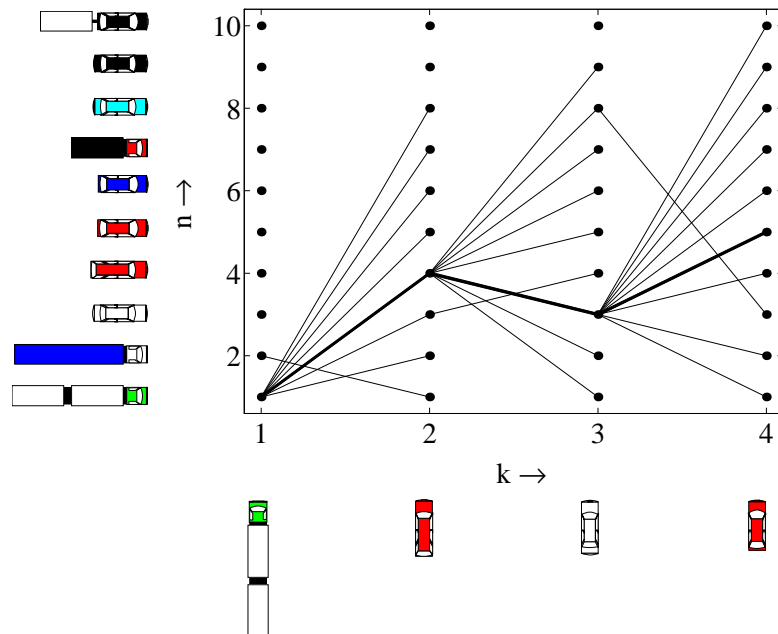


Bild 5.28: Trellis-Diagramm und optimaler Pfad bei der Fahrzeugzuordnung

5.4.5 Simulationsergebnisse

Die den hier dargebrachten Ergebnissen zugrunde liegenden Fahrzeugsignale und mikroskopischen Verkehrssimulationen sind wiederum mit denen aus Abschnitt 5.2.2 identisch. Die für die Simulationen gewählten Parameter finden sich in Tabelle 5.3

Tabelle 5.3: Parameter zur Schätzung mit Hidden-Markov-Modellen

	Schleife	1m	2,5m
Verzögerung bei der Viterbi-Detektion	t_{max}		10
Anzahl der Beobachtungsobjekte	T		11
Transitionsmatrix	\mathbf{A}	siehe Bild 5.24	
Dimension der Merkmalsvektoren	N	14	15

Bild 5.29 zeigt die Ergebnisse der Einzelfahrzeugwiedererkennung für diese Simulation für die Signale von den 1m-Doppelschleifen und den 2,5m-Schleifen. Die Signale der 1m-Schleife können wiederum als nahezu ideal betrachtet werden und dementsprechend ist auch bei dieser Simulation die Fehlerrate der Wiedererkennung sehr gering. Sie liegt bei diesem Beispiel bei ungefähr 2 %. Realistischer sind die Bedingungen für die Simulationen mit den Merkmalsvektoren von den weiter auseinander liegenden 2,5m-Schleifen.

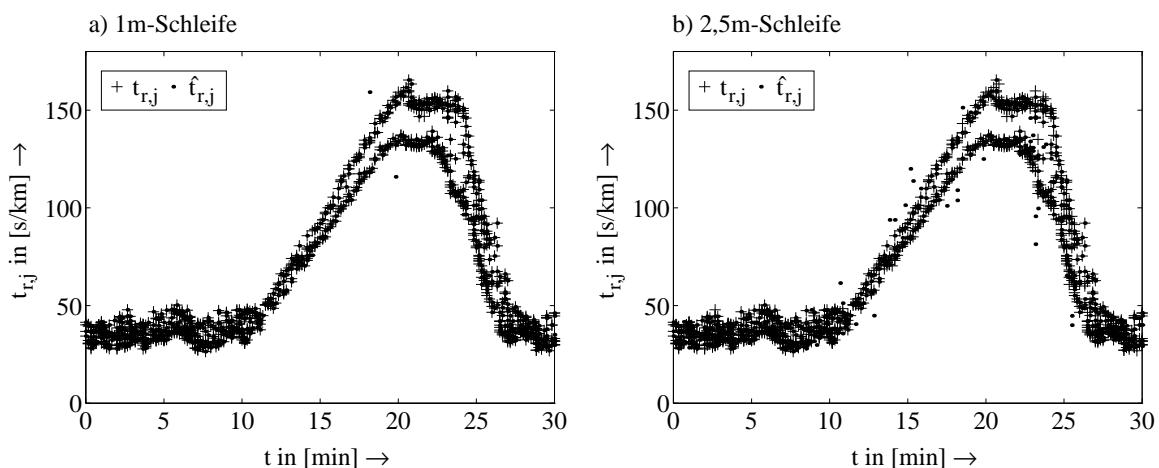


Bild 5.29: Wahre und geschätzte Fahrzeugreisezeiten

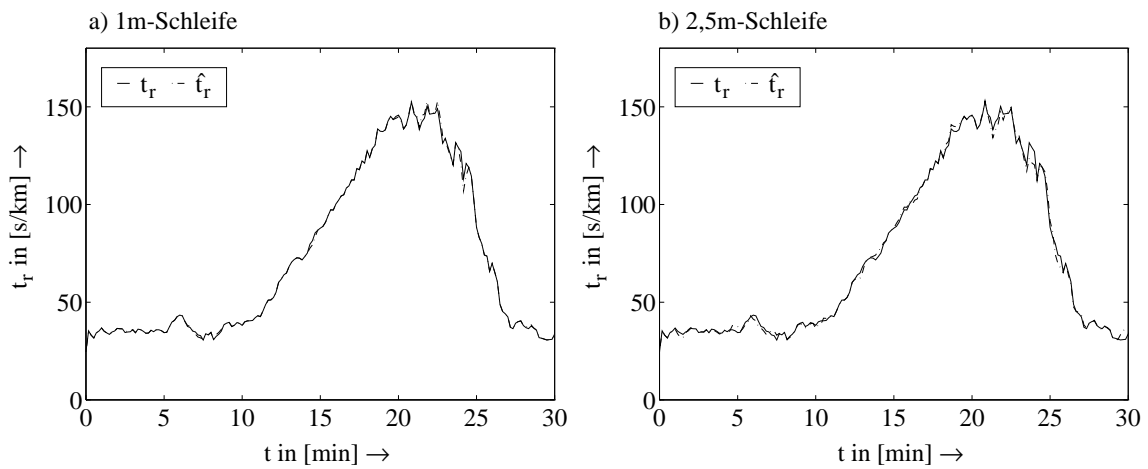


Bild 5.30: Wahrer und geschätzter Median der Reisezeiten für $T = 10$ s

Es zeigt sich allerdings, daß die Schätzergebnisse trotz der relativ schlechten Schleifeneigenschaften als gut angesehen werden können. Ähnlich wie bei dem letzten Verfahren mit Kostenfunktionen ergibt sich auch bei diesem Verfahren, daß Falschzuordnungen von Fahrzeugen nur einen kleinen Fehler in der Reisezeitschätzung bewirken, da die durch die Transitionsmatrix festgelegten Abhängigkeiten der Fahrzeugpositionen „Ausreißer“ unterbinden.

In Bild 5.30 sind die dazugehörigen gemittelten Reisezeiten aufgetragen. Es werden der wahre und der geschätzte Median der Fahrzeugreisezeiten in Zeitintervallen von $T = 10$ s einander gegenübergestellt. Zur Glättung der Werte wurde auch für diese Darstellung über drei Perioden gemittelt.

5.5 Verfahrensvergleich

Abschließend sollen die in den letzten drei Abschnitten vorgestellten Verfahren miteinander verglichen werden. Hierbei müssen die Schätzergebnisse dem jeweiligen Realisierungsaufwand gegenübergestellt werden.

Aus den Simulationsergebnissen wird anhand der Bilder 5.12, 5.18 und 5.30 deutlich, daß alle Verfahren eine zuverlässige Schätzung für die mitt-

leren Reisezeiten liefern. Auf den Nachteil der Mittelung der Reisezeit im gestörten Verkehrsablauf bei dem Schätzverfahren mit Fahrzeugfolgenkorrelationen wurde hingewiesen. Aus diesem Grund müssen die Verfahren mit Kostenfunktionen und mit Hidden-Markov-Modellen als besser bewertet werden.

Der Rechenaufwand wird in allen drei Verfahren vorwiegend von der Berechnung der Metriken zwischen dem am stromaufwärtigen Meßquerschnitt detektierten Fahrzeug und den Fahrzeugen im Streckenabschnitt geprägt. Im Vergleich schneidet das Verfahren mit den Kostenfunktionen am besten ab, da hier nach der Zuordnung eines Fahrzeugs nur dann weitere Berechnungen notwendig sind, wenn das bei der Vorauswahl zugeordnete Fahrzeug bereits mit geringeren Kosten zugeordnet wurde. Bei dem Verfahren mit Fahrzeugfolgenkorrelation folgt nach der Metrikberechnung eine zweidimensionale Faltung. Am aufwendigsten ist das Verfahren mit Hidden-Markov-Modellen, in dem nach der Zuordnung eines Fahrzeugs eine aufwendige Neuordnung der Modellzustände und der Zeiger für die Rückrechnung des optimalen Pfades der Zustandssequenzen für den nächsten Zeitschritt erfolgen muß. Insgesamt sind für dieses Verfahren ungefähr doppelt so viele Rechenoperationen erforderlich wie für das Verfahren mit Kostenfunktionen. Das Verfahren mit Fahrzeugfolgenkorrelation ist bezüglich des Rechenaufwands zwischen den beiden anderen Verfahren angesiedelt.

Der Vorteil des Verfahrens mit Hidden-Markov-Modellen ist in der schnellen Entscheidung zu sehen. Nach einer Verzögerung von 10 Fahrzeugen konnte bei der dargestellten Simulation bereits eine eindeutige Zuordnung eines Fahrzeugs stattfinden, das Schätzergebnis wird demnach mit einer Verzögerung von ca. 20 s bereitgestellt. Bei dem Verfahren mit Kostenfunktionen ergab sich, daß die Wartezeit einen Wert von 30 Fahrzeugen nicht unterschreiten sollte. Legt man zum Vergleich eine Anzahl von 10 fest, so ist dieses Verfahren demjenigen mit Hidden-Markov-Modellen deutlich unterlegen. Für das Verfahren mit Fahrzeugfolgenkorrelationen muß eine Fahrzeuggruppe von 50 Fahrzeugen detektiert werden, ehe das erste Schätzergebnis vorliegt. Somit ist dieses Verfahren auch unter diesem Aspekt als schlechter zu bewerten.

6 Verkehrszustandsschätzung mit streckenbezogenen Verkehrsdaten

Auf dem Gebiet der Verkehrszustandsschätzung und der Störungserkennung sind in den letzten beiden Jahrzehnten zahlreiche Verfahren entstanden, z.B. [2], [18], [19], [23], [39], [49], [56], [59], [82]. Man unterteilt diese Verfahren in sogenannte *Querschnittsverfahren* und *Abschnittsverfahren*. Bei den Querschnittsverfahren werden lediglich Verkehrsdaten von einem einzigen Meßquerschnitt ausgewertet, während die Abschnittsverfahren die Verkehrsdaten von mindestens zwei benachbarten Meßstellen miteinbeziehen. Die Verkehrsdaten können hierbei *mikroskopisch* oder *makroskopisch* vorliegen. Mikroskopische Verkehrsdaten stammen von Individualfahrzeugen während makroskopische Verkehrsdaten kollektive Daten sind, die über räumliche oder zeitliche Intervalle gemittelt werden.

Aufbauend auf den Ergebnissen des letzten Kapitels steht im Mittelpunkt dieses Kapitels die Erweiterung eines makroskopischen Abschnittsverfahrens, das die Fahrzeugdaten durch besondere Schätzverfahren und Filtertechniken verarbeitet. Bislang wurden allerdings als Ein- und Ausgangsgrößen vorwiegend querschnittsbezogene Verkehrsdaten wie die Verkehrsstärken oder lokale zeitlich gemittelte Geschwindigkeiten an den Meßquerschnitten benutzt. Durch Hinzunahme der mittleren Reisezeit der Fahrzeuge als Meßgröße soll dieses Verfahren im folgenden erweitert werden. Ziel ist es, hierdurch eine Verbesserung der Verkehrszustandsschätzung sowie der Detektion von Verkehrsstörungen zu erreichen.

Auf der Basis eines Modells von PAYNE [59] wurde von CREMER ein Schätzverfahren entwickelt [21] und später validiert und laufend verbessert [23], [24], [25], [26], [27], [74]. Dieses Schätzverfahren zeichnet sich dadurch aus, daß es als einziges auch bei einem großen Abstand der Meßquerschnitte (3 bis 5 km) noch akzeptable Schätzwerte liefert. Dieses geht aus einem Verfahrensvergleich von BUSCH eindrucksvoll hervor [16], [17].

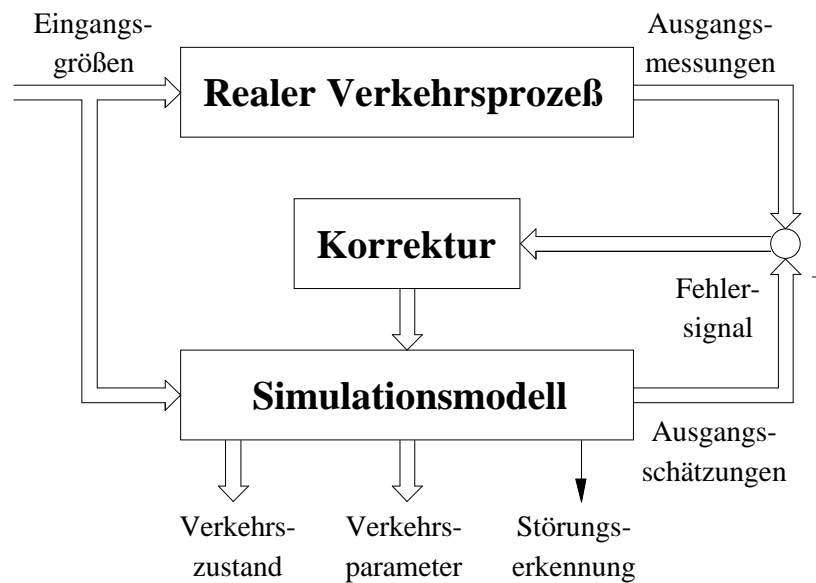


Bild 6.1: Blockdiagramm des modellgestützten Schätzverfahrens

Kern des Verfahrens ist ein aus nichtlinearen Differenzgleichungen bestehendes dynamisches Modell für den Straßenverkehrsfluß. Auf diesem makroskopischen Verkehrsmodell aufbauend wird eine Prädiktion aktueller Messungen von den Meßquerschnitten durchgeführt. Ein Vergleich der vorausgesagten und der realen Messungen führt zu einer Korrektur des geschätzten Verkehrszustandes (Bild 6.1).

Zu Beginn des Kapitels wird das dem Schätzverfahren zugrundeliegende makroskopische Verkehrsmodell vorgestellt. Anschließend wird die Erweiterung des Modells um streckenbezogene Verkehrsdaten aufgeführt. Einer Darstellung der stochastischen Zustandsvektoren des Verkehrsmodells schließt sich die Lösung des modellgestützten Schätzproblems unter Hinzunahme der mittleren Reisezeit mit einem erweiterten Kalman-Filter an. Anschließend wird der Einfluß der zusätzlichen Kenntnis von streckenbezogenen Verkehrsdaten auf das Schätzproblem mit einer Beobachtbarkeitsanalyse bewertet. Nach der Analyse bildet eine Präsentation von Simulationsergebnissen den Abschluß dieses Kapitels.

6.1 Das makroskopische Verkehrsmodell

Das makroskopische Verkehrsmodell ist umfassend in [22] beschrieben. In diesem Abschnitt soll lediglich eine Übersicht über die Variablen und die Gleichungen des Modells gegeben und die Erweiterungen des Modells dargestellt werden.

6.1.1 Variablen des Verkehrsflußmodells

Zur Beschreibung der Modellgleichungen wird eine mehrspurige Richtungsfahrbahn eines Autobahnabschnitts herangezogen, die eine Länge von wenigen Kilometern hat. Dieser Straßenabschnitt wird in n Segmente mit einer Länge Δ_j von ungefähr 500 m unterteilt, in denen die Verkehrsverhältnisse in erster Näherung als gleichförmig angesehen werden können. Falls in diesem Autobahnabschnitt Zu- und Abfahrten bestehen, werden diese am Ein- bzw. am Ausgang des Straßenstücks angenommen (Bild 6.2).

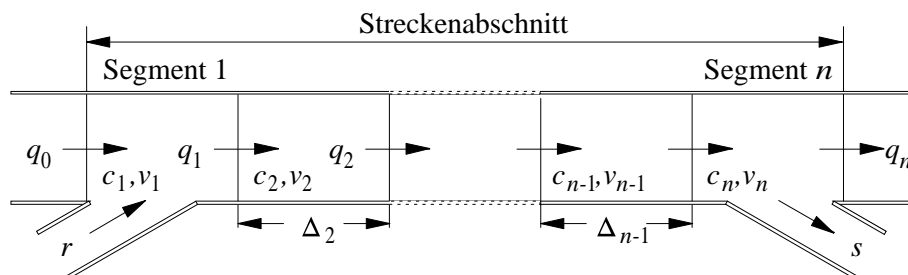


Bild 6.2: Unterteilung eines Schnellstraßen-Streckenabschnitts

Der Verkehrsfluß in einem Segment j kann mit folgenden Parametern beschrieben werden:

- **Verkehrsdichte $c_j(k)$ in [Fz/km]**
Anzahl der Fahrzeuge im Segment j zum Zeitpunkt kT bezogen auf die Segmentlänge Δ_j .

- **Mittlere Geschwindigkeit $v_j(k)$ in [km/h]**
Mittelwert der Geschwindigkeiten der Fahrzeuge im Segment j zum Zeitpunkt kT .
- **Verkehrsstärke $q_j(k)$ in [Fz/h]**
Anzahl der Fahrzeuge, die im Zeitintervall $kT \leq t < (k+1)T$ den Fahrbahnquerschnitt zwischen den Segmenten j und $j+1$ passieren, bezogen auf die Intervalldauer T .
- **Rampenverkehrsstärken $r_j(k)$ und $s_j(k)$ in [Fz/h]**
Anzahl der Fahrzeuge, die im Zeitintervall $kT \leq t < (k+1)T$ am Anfang bzw. Ende des Segments j den Straßenabschnitt befahren bzw. verlassen, bezogen auf die Intervalllänge T .
- **Lokale Geschwindigkeit $w_j(k)$ in [km/h]**
Mittelwert der Geschwindigkeiten der Fahrzeuge, die im Zeitintervall $kT \leq t < (k+1)T$ den Fahrbahnquerschnitt zwischen den Segmenten j und $j+1$ überfahren.

Mit diesen Parametern ergibt sich ein zeitdiskretes Modell mit der Zeitschrittweite T und dem laufenden Zeitindex $k \in \mathbb{N}$. Der Systemzustandsübergang vom Zeitpunkt kT zum Zeitpunkt $(k+1)T$ kann durch Differenzgleichungen für die Verkehrsdichte $c_j(k)$ und für die mittlere Geschwindigkeit $v_j(k)$ im Segment j beschrieben werden.

Für die Verkehrsdichte ergibt sich aus einer einfachen Fahrzeugbilanz

$$c_j(k+1) = c_j(k) + \frac{T}{\Delta_j} [q_{j-1}(k) - q_j(k)]. \quad (6.1)$$

Die Rampenverkehrsstärken $r_j(k)$ und $s_j(k)$ eventuell vorhandener Zu- und Abfahrten werden im Klammersausdruck für $j = 1$ addiert bzw. für $j = n$ subtrahiert.

Die mittlere Geschwindigkeit wird von drei Termen beeinflusst und kann durch folgende Differenzgleichung beschrieben werden:

$$\begin{aligned} v_j(k+1) = & v_j(k) + \frac{T}{\tau} [V(c_j(k)) - v_j(k)] + \frac{T}{\Delta_j} v_j(k) [v_{j-1}(k) - v_j(k)] + \\ & + \frac{\nu T}{\Delta_j \tau} \frac{c_j(k) - c_{j+1}(k)}{c_j(k) + \kappa}. \end{aligned} \quad (6.2)$$

Der erste Term berücksichtigt die dynamische Annäherung der stationären Geschwindigkeits-Dichte-Charakteristik $V(c)$. Der zweite Term berechnet die Konvektion von einem Geschwindigkeitsgradienten, und der letzte Term beinhaltet eine stromabwärtige Vorausschau des Fahrers auf den Dichtegradienten. Die Parameter in dieser Gleichung werden später noch näher spezifiziert.

Während die Verkehrsstärken $q_0(k)$ und $q_n(k)$ am Übergang verschiedener Autobahnabschnitte gemessen werden, werden die Verkehrsstärken $q_j(k)$ zwischen den Segmenten als gewichteter örtlicher Mittelwert über die Produkte aus Geschwindigkeit und Dichte berechnet:

$$q_j(k) = \alpha c_j(k) v_j(k) + (1 - \alpha) c_{j+1}(k) v_{j+1}(k), \quad (6.3)$$

α ist ein konstanter Gewichtungsfaktor aus dem Intervall $[0.5, 1]$.

Die lokalen Geschwindigkeiten $w_j(k)$ errechnen sich analog zu den Verkehrsstärken zu

$$w_j(k) = \alpha v_j(k) + (1 - \alpha) v_{j+1}(k). \quad (6.4)$$

Die stationäre Geschwindigkeits-Dichte-Charakteristik $V(c)$ in Gleichung (6.2) wird mit

$$V(c_j) = v_f \left[1 - \left(\frac{c_j}{c_{\max}} \right)^l \right]^m \quad (6.5)$$

berechnet. v_f ist die mittlere Geschwindigkeit der Fahrzeuge bei freiem Verkehr und c_{\max} die maximale Verkehrsdichte. Die Größen l und m sind reelle positive Exponenten.

Die lokalen Geschwindigkeiten und Verkehrsstärken am Ein- und Ausgang des Straßenabschnitts werden als Eingangs- oder Ausgangsgrößen des Systems betrachtet. Die Eingangsverkehrsstärke q_0 wird dem System als Eingangsgröße zur Verfügung gestellt, während die Meßwerte w_0 , q_n und w_n als Systemreaktion (Ausgangsgrößen) behandelt werden.

Diese Größen werden wie folgt aus den Zustandsgrößen berechnet:

$$w_0(k) = (1 + \alpha \epsilon) v_1(k) - \alpha \epsilon v_2(k), \quad (6.6)$$

$$q_n(k) = (1 + (1 - \alpha) \epsilon) c_n(k) v_n(k) - (1 - \alpha) \epsilon c_{n-1}(k) v_{n-1}(k), \quad (6.7)$$

$$w_n(k) = (1 + (1 - \alpha) \epsilon) v_n(k) - (1 - \alpha) \epsilon v_{n-1}(k). \quad (6.8)$$

Tabelle 6.1: Modellparameter

T	Zeitschrittweite des diskreten Modells	10	s
Δ_j	Länge des Segments j	500	m
α	Gewichts- oder Mittelungsfaktor	0.8	
ϵ	Extrapolationsfaktor	0.2	
κ	minimale Dichte	20	Fz/km
ν	Empfindlichkeitsfaktor	21.4	km ² /h
τ	Anpassungszeit	34	s
v_f	freie Geschwindigkeit	122.4	km/h
c_{\max}	maximale Verkehrsdichte	200	Fz/km
m		4.0	
l	dimensionslose Konstanten	1.4	

ϵ ist ein Extrapolationsfaktor.

6.1.2 Modellvalidierung

Die Gleichungen (6.1) bis (6.3) und (6.5) beschreiben ein nichtlineares, zeitdiskretes Modell der Ordnung $2n$ zur Simulation des Verkehrsflusses in einem Autobahnabschnitt.

Die freien Systemparameter sind in [23] mit einer Sensitivitätsanalyse identifiziert und anhand realer Meßdaten kalibriert worden. In Tabelle 6.1 sind die optimalen Modellparameter in ihrer Größe und Bedeutung aufgetragen.

Die Verhältnisse aus der Geschwindigkeits-Dichte-Charakteristik (6.5) lassen sich unter der Voraussetzung gleichförmiger, stationärer Verkehrsbedingungen durch die Beziehung

$$q = c \cdot v = c \cdot V(c) \quad (6.9)$$

in das Fundamentaldiagramm übertragen, das den stationären Zusammenhang zwischen der Verkehrsstärke q und der Verkehrsdichte c wiedergibt. Bild 6.3 zeigt die V-C-Charakteristik und das Fundamentaldiagramm für die in Tabelle 6.1 beschriebenen Parameter.

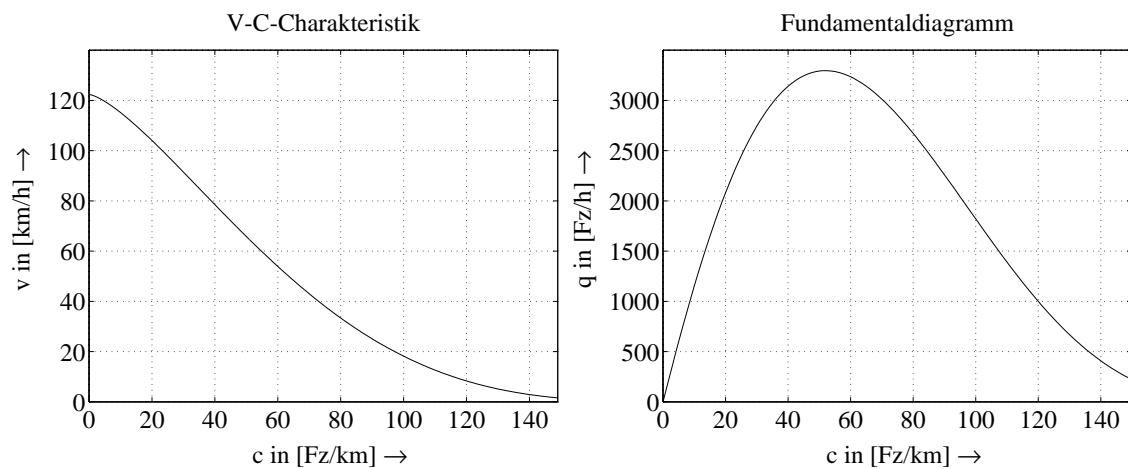


Bild 6.3: V-C-Charakteristik und Fundamentaldiagramm

6.1.3 Modellerweiterung zur Störfallerkennung

Zur Nachbildung von Störungseinflüssen auf die Meßdaten am Rand eines untersuchten Straßenabschnitts wurde in [22] und [73] ein hypothetischer Zufluß q_d eingeführt, der in das Segment eines Störungsbereiches hineinfließt und diesen am nächsten Segmentanfang wieder verläßt (Bild 6.4). Dadurch können Auswirkungen von Störungen untersucht werden, die durch ein plötzlich verringertes Fassungsvermögen in dem entsprechenden Segment durch Sperrung einer oder mehrerer Fahrspuren verursacht werden.

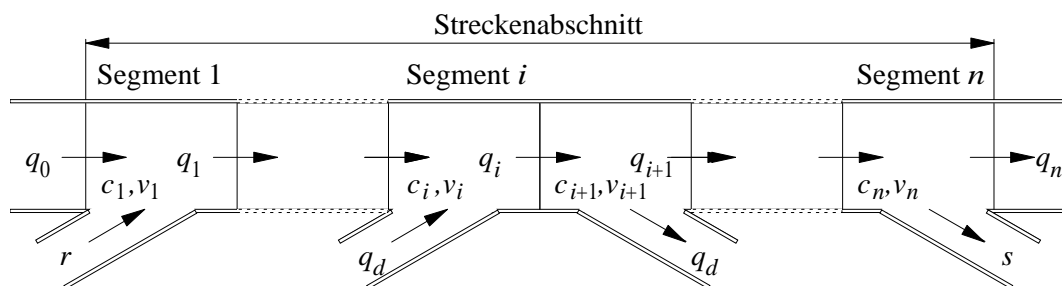


Bild 6.4: Schnellstraßenabschnitt mit hypothetischem Zufluß

Der hypothetische Zufluß wird in den zwei Segmenten i und $i+1$ jeweils als linearer Term hinzugefügt bzw. abgezogen. Damit ergeben sich für die

Verkehrsdichten in diesen Segmenten folgende veränderten Gleichungen:

$$c_i(k+1) = c_i(k) + \frac{T}{\Delta_i} [q_{i-1}(k) - q_i(k) + q_d(k)], \quad (6.10)$$

$$c_{i+1}(k+1) = c_{i+1}(k) + \frac{T}{\Delta_{i+1}} [q_i(k) - q_{i+1}(k) - q_d(k)]. \quad (6.11)$$

Der hypothetische Zufluß

$$q_d(k+1) = q_d(k) + n_q(k) \quad (6.12)$$

erweitert das makroskopische Modell um eine weitere Differenzgleichung. Hier ist $n_q(k)$ ein mittelwertfreier, weißer Rauschterm mit entsprechender Varianz.

6.1.4 Modellerweiterung mit streckenbezogenen Verkehrsdaten

Für eine Erweiterung des makroskopischen Verkehrsmodells um streckenbezogene Verkehrsdaten muß eine Meßgleichung aufgestellt werden, in der die mittlere Reisezeit der Fahrzeuge durch die makroskopischen Größen der Verkehrsdichte und der mittleren Geschwindigkeit ausgedrückt werden kann:

$$t_r(k) = f[c_j(k), v_j(k)], \quad j = 1, \dots, n, \quad k \in \mathbb{N}. \quad (6.13)$$

Nicht festgelegt ist die Form der Mittelung der individuellen Reisezeiten. Unter der Annahme einer eindeutigen Wiedererkennung von Fahrzeugen am stromabwärtigen Meßquerschnitt (siehe Kapitel 5) kann die Mittelung über alle Fahrzeuge, die den Querschnitt im betrachteten Zeitintervall passieren, erfolgen. Somit sind folgende Definitionen sinnvoll:

- **Reisezeit $t_{r_{fz}}$ in [s]**
Reisezeit eines Fahrzeugs vom Eingang bis zum Ausgang des Straßenabschnitts.

- **Mittlere Reisezeit im Segment $t_{r_j}(k)$ in [s]**
Mittlere Reisezeit der Fahrzeuge vom Eingang bis zum Ausgang des Segments j . Die Mittelung erfolgt über alle Fahrzeuge, die im Zeitintervall $kT \leq t < (k+1)T$ das Segment j verlassen.
- **Mittlere Reisezeit $t_r(k)$ in [s]**
Mittlere Reisezeit der Fahrzeuge vom Eingang bis zum Ausgang des Streckenabschnitts. Die Mittelung erfolgt über alle Fahrzeuge, die im Zeitintervall $kT \leq t < (k+1)T$ den Streckenabschnitt verlassen.

Die Reisezeit eines Fahrzeugs vom Eingang des Streckenabschnitts am Ort s_1 zum Ausgang am Ort s_2 ergibt sich als Trajektorie entlang des Geschwindigkeitsprofils $v[s, t(s)]$ zu

$$t_{r_{fz}} = \int_{s_1}^{s_2} \frac{1}{v[s, t(s)]} ds. \quad (6.14)$$

Im zeitdiskreten Fall geht das Integral in eine Summe über die Wegelemente Δs mit dem Zeitindex $k \in \mathbb{N}$ über:

$$t_{r_{fz}} = \sum_{s_1}^{s_2} \frac{\Delta s}{v[s, k(s)]}. \quad (6.15)$$

Aus Untersuchungen von Geschwindigkeitsprofilen und Reisezeiten für verschiedene Szenarien geht hervor, daß die mittlere Reisezeit der Fahrzeuge im Falle eines quasistationären Geschwindigkeitsprofils mit

$$t_r(k) = \sum_{j=1}^n t_{r_j}(k) = \begin{cases} \sum_{j=1}^n \frac{\Delta_j}{v_j(k)} & \text{für } v_j(k) > v_{\min} \\ \sum_{j=1}^n \frac{2 v_{\min} \Delta_j}{v_j^2(k) + v_{\min}^2} & \text{für } 0 \leq v_j(k) < v_{\min} \end{cases} \quad (6.16)$$

aus den Zustandsgrößen berechnet werden kann [83], [25]. Der Parameter v_{\min} berücksichtigt eine obere Beschränkung der mittleren Reisezeit im Streckenabschnitt. Die mittlere Reisezeit der Fahrzeuge im betrachteten Streckenabschnitt erweitert das bisher beschriebene Systemmodell um eine weitere Ausgangsgröße.

6.2 Modellgestützte Verkehrszustandsschätzung

Das Zeitverhalten eines dynamischen Systems ist bei einem gegebenen mathematischen Modell und Kenntnis der Eingangsgrößen sowie des Anfangszustands berechenbar. Dafür wird ein Zustandsvektor zur Ermittlung des Zeitverhaltens benötigt. Dieser ist allerdings häufig meßtechnisch nicht zugänglich, sondern muß aus den meßbaren Ausgangsgrößen des Systems berechnet werden. Wegen auftretender Meßfehler liefert die Berechnung allerdings nur einen Schätzwert für den Zustandsvektor. Bei Kenntnis der stochastischen Verteilungen der Meßwerte kann die Schätzung des Zustandsvektors mit einem Kalman-Filter erfolgen.

In diesem Abschnitt werden die stochastischen Zustandsvektoren des makroskopischen Verkehrsmodells aufgestellt. Nach einer kurzen Darstellung der Gleichungen des erweiterten Kalman-Filters schließt sich die Lösung des modellgestützten Verkehrszustandsschätzproblems unter Hinzunahme von streckenbezogenen Verkehrsdaten an.

6.2.1 Das stochastische Verkehrsflußmodell

Zur Anwendung stochastischer Filterverfahren wird ein stochastisches Verkehrsflußmodell benötigt. Dieses besteht aus einer Erweiterung des in Abschnitt 6.1.1 beschriebenen makroskopischen Verkehrsmodells. Die Differenzgleichungen des stochastischen Modells ergeben sich nach [22] für die Verkehrsdichte c_j und für die mittlere Geschwindigkeit v_j wie folgt:

$$c_j(k+1) = c_j(k) + \frac{T}{\Delta_j} [q_{j-1}(k) - q_j(k) + \xi_{j-1}(k) - \xi_j(k)], \quad (6.17)$$

$$v_j(k+1) = v_j(k) + \frac{T}{\tau} [V(c_j(k)) - v_j(k)] + \frac{T}{\Delta_j} v_j(k) [v_{j-1}(k) - v_j(k)] + \frac{\nu T}{\Delta_j \tau} \frac{c_j(k) - c_{j+1}(k)}{c_j(k) + \kappa} + \eta_j(k). \quad (6.18)$$

Die Verkehrsstärken und die lokalen Geschwindigkeiten werden als lo-

kale Referenzgrößen an den Meßquerschnitten n betrachtet:

$$q_n(k) = \alpha c_n(k) v_n(k) + (1 - \alpha) c_{n+1}(k) v_{n+1}(k) + \xi_n(k) + \zeta_{q_n}(k), \quad (6.19)$$

$$w_n(k) = \alpha v_n(k) + (1 - \alpha) v_{n+1}(k) + \zeta_{w_n}(k). \quad (6.20)$$

$\xi_n(k)$, $\eta_n(k)$, $\zeta_{q_n}(k)$ und $\zeta_{w_n}(k)$ sind stochastische Korrekturterme, die als weiße Rauschprozesse mit gaußscher Amplitudenverteilung angenommen werden.

$\xi_n(k)$ steht stellvertretend für die Abweichungen der tatsächlichen Verkehrsstärkewerte von den aus Gleichung (6.1) errechneten Werten. $\zeta_{q_n}(k)$ enthält die Meßfehler der aufnehmenden Sensoren und ist mit $\xi_n(k)$ unkorreliert. Da die lokale Geschwindigkeit $w_n(k)$ nicht in die Aufstellung der Systemgleichungen eingeht, erübrigt sich eine derartige Trennung für den stochastischen Korrekturterm $\zeta_{w_n}(k)$.

Ähnlich muß nun noch eine stochastische Fehlerbehandlung der mittleren Reisezeit $t_r(k)$ am Meßquerschnitt n erfolgen. Dahingehend wird (6.16) zu

$$t_r(k) = \begin{cases} \sum_{j=1}^n \frac{\Delta_j}{v_j(k)} + \zeta_{t_r}(k) & \text{für } v_j(k) \geq v_{\min} \\ \sum_{j=1}^n \frac{2 v_{\min} \Delta_j}{v_j^2(k) + v_{\min}^2} + \zeta_{t_r}(k) & \text{für } 0 \leq v_j(k) < v_{\min} \end{cases} \quad (6.21)$$

erweitert, wobei ζ_{t_r} sowohl die statistisch schwankenden Abweichungen der Größe $t_r(k)$ vom nach (6.16) errechneten Wert als auch Meßfehler berücksichtigt.

Mit diesen Gleichungen kann der dynamische Ablauf des stochastischen Verkehrsflusses makroskopisch beschrieben werden. Die vollständige Information über den Systemzustand liegt in den Verkehrsdichten c_j und den mittleren Geschwindigkeiten v_j für $j = 1, \dots, n$, die bei bekannter Eingangsgröße q_0 zum Zeitpunkt $(k + 1)T$ in eindeutiger Weise aus den Werten $c_j(k)$ und $v_j(k)$ hervorgehen.

6.2.2 Zustandsraumdarstellung

Die Vektoren des stochastischen Zustandsraummodells sollen im folgenden zusammenfassend dargestellt werden:

- **Zustandsvektor**

- Modell ohne Störfallerkennung

$$\mathbf{x}(k) = [c_1(k), \dots, c_n(k), v_1(k), \dots, v_n(k)]^T \quad (6.22)$$

- Modellerweiterung zur Störfallerkennung

$$\mathbf{x}(k) = [c_1(k), \dots, c_n(k), v_1(k), \dots, v_n(k), q_d(k)]^T \quad (6.23)$$

- **Eingangsvektor**

$$\mathbf{u}(k) = [q_0(k)] \quad (6.24)$$

- **Ausgangsvektor**

- Ausschließliche Verwendung querschnittsbezogener Verkehrsdaten

$$\mathbf{y}(k) = [w_0(k), q_n(k), w_n(k)]^T \quad (6.25)$$

- Modellerweiterung mit streckenbezogenen Verkehrsdaten

$$\mathbf{y}(k) = [w_0(k), q_n(k), w_n(k), t_r(k)]^T \quad (6.26)$$

- **Störprozeß am Systemeingang**

- Modell ohne Störfallerkennung

$$\mathbf{w}(k) = [\xi_0(k), \dots, \xi_n(k), \eta_1(k), \dots, \eta_n(k)]^T \quad (6.27)$$

- Modellerweiterung zur Störfallerkennung

$$\mathbf{w}(k) = [\xi_0(k), \dots, \xi_n(k), \eta_1(k), \dots, \eta_n(k), n_q(k)]^T \quad (6.28)$$

- **Störprozeß am Systemausgang**

- Ausschließliche Verwendung querschnittsbezogener Verkehrsdaten

$$\mathbf{n}(k) = [\zeta_{w_0}(k), \zeta_{q_n}(k) + \xi_n(k), \zeta_{w_n}(k)]^T \quad (6.29)$$

- Modellerweiterung mit streckenbezogenen Verkehrsdaten

$$\mathbf{n}(k) = [\zeta_{w_0}(k), \zeta_{q_n}(k) + \xi_n(k), \zeta_{w_n}(k), \zeta_{t_r}(k)]^T \quad (6.30)$$

Die Zustandsgleichungen und Gleichungen der Ausgangsgrößen für das stochastische Verkehrsflußmodell

$$\mathbf{x}(k+1) = \mathbf{f}(\mathbf{x}(k), \mathbf{u}(k), \mathbf{w}(k)) \quad (6.31)$$

$$\mathbf{y}(k) = \mathbf{g}(\mathbf{x}(k)) + \mathbf{n}(k) \quad (6.32)$$

können zur Parameterschätzung mit dem erweiterten Kalman-Filter herangezogen werden.

6.2.3 Zustandsschätzung mit dem erweiterten Kalman-Filter

Der Einsatz des erweiterten Kalman-Filters zur modellgestützten Zustandsschätzung des Verkehrsflusses ist ausführlich in [22] beschrieben. An dieser Stelle soll nur ein kurzer Überblick über die Arbeitsweise und die Gleichungen des Filters gegeben werden, womit eine Darstellung des Schätzproblems bei zusätzlicher Kenntnis der Fahrzeugreisezeiten im folgenden Abschnitt durchgeführt werden kann. Herleitungen der Filtergleichungen finden sich z.B. in [3], [12], [13] und [46].

Gegeben sei ein durch die Zustandsgleichungen (6.31) und (6.32) beschriebenes nichtlineares, zeitdiskretes System, in dem $\mathbf{x}(k)$ den Signal- oder Zustandsvektor, $\mathbf{u}(k)$ den Eingangsvektor, $\mathbf{y}(k)$ den Empfangs- oder Ausgangsvektor und $\mathbf{w}(k)$ und $\mathbf{n}(k)$ weiße, gaußverteilte Störprozesse am Eingang und am Empfänger darstellen.

Für die Erwartungswerte der Störprozesse gelte:

$$E\{\mathbf{w}(k)\} = E\{\mathbf{n}(k)\} = \mathbf{0}, \quad (6.33)$$

$$E\{\mathbf{w}(k)\mathbf{w}^T(\ell)\} = \mathbf{R}_{ww}(k) \cdot \delta_{k\ell}, \quad (6.34)$$

$$E\{\mathbf{n}(k)\mathbf{n}^T(\ell)\} = \mathbf{R}_{nn}(k) \cdot \delta_{k\ell}, \quad (6.35)$$

$$E\{\mathbf{w}(k)\mathbf{n}^T(\ell)\} = \mathbf{R}_{wn}(k) \cdot \delta_{k\ell}. \quad (6.36)$$

Der Anfangszustand des Systems sei gaußverteilt mit der Kovarianzmatrix

$$E\{[\mathbf{x}(0) - E\{\mathbf{x}(0)\}][\mathbf{x}(0) - E\{\mathbf{x}(0)\}]^T\} = \mathbf{R}_{xx}(0). \quad (6.37)$$

Zudem seien der Anfangszustand des Systems und die Störprozesse unkorreliert:

$$E\{\mathbf{x}(0)\mathbf{n}^T(k)\} = E\{\mathbf{x}(0)\mathbf{w}^T(k)\} = \mathbf{0}. \quad (6.38)$$

Die Komponenten der nichtlinearen Vektorfunktionen $\mathbf{f}(\cdot)$ und $\mathbf{g}(\cdot)$ seien nach allen Argumenten stetig differenzierbar. Durch Linearisierung in den interessierenden Arbeitspunkten ergibt sich

$$\mathbf{F}(k) = \left. \frac{\partial \mathbf{f}}{\partial \mathbf{x}} \right|_{\substack{\mathbf{x} = \hat{\mathbf{x}}(k) \\ \mathbf{u} = \mathbf{u}(k) \\ \mathbf{w} = \mathbf{0}}}, \quad \mathbf{\Gamma}(k) = \left. \frac{\partial \mathbf{f}}{\partial \mathbf{w}} \right|_{\substack{\mathbf{x} = \hat{\mathbf{x}}(k) \\ \mathbf{u} = \mathbf{u}(k) \\ \mathbf{w} = \mathbf{0}}}, \quad \mathbf{H}(k) = \left. \frac{\partial \mathbf{g}}{\partial \mathbf{x}} \right|_{\mathbf{x} = \hat{\mathbf{x}}(k)}. \quad (6.39)$$

Gesucht wird eine lineare, erwartungstreue Schätzung des Signalvektors $\mathbf{x}(k)$ auf der Basis der Empfangssignale $\mathbf{y}(0), \dots, \mathbf{y}(k)$. Für diesen Fall liefert das erweiterte Kalman-Filter Schätzwerte $\hat{\mathbf{x}}(k)$ kleinster Fehlervarianz für die unbekanntenen Werte $\mathbf{x}(k)$.

Das erweiterte Kalman-Filter setzt sich aus folgenden Gleichungen zusammen:

Filtermodell

$$\hat{\mathbf{x}}(k+1|k) = \mathbf{f}(\hat{\mathbf{x}}(k), \mathbf{u}(k), \mathbf{0}), \quad \hat{\mathbf{x}}(0|-1) = E\{\mathbf{x}(0)\} \quad (6.40)$$

$$\hat{\mathbf{x}}(k) = \hat{\mathbf{x}}(k|k-1) + \mathbf{K}(k)[\mathbf{y}(k) - \mathbf{g}(\hat{\mathbf{x}}(k|k-1))] \quad (6.41)$$

Verstärkungsmatrix

$$\mathbf{K}(k) = \mathbf{R}_{ee}(k|k-1)\mathbf{H}^T(k)[\mathbf{H}(k)\mathbf{R}_{ee}(k|k-1)\mathbf{H}^T(k) + \mathbf{R}_{nn}(k)]^{-1} \quad (6.42)$$

Fehlerkovarianzmatrizen

$$\mathbf{R}_{ee}(k) = \mathbf{R}_{ee}(k|k-1) - \mathbf{K}(k)\mathbf{H}(k)\mathbf{R}_{ee}(k|k-1) \quad (6.43)$$

$$\mathbf{R}_{ee}(k+1|k) = \mathbf{F}(k)\mathbf{R}_{ee}(k)\mathbf{F}^T(k) + \mathbf{\Gamma}(k)\mathbf{R}_{ww}(k)\mathbf{\Gamma}^T(k) \quad (6.44)$$

Anfangsbedingung

$$\mathbf{R}_{ee}(0|-1) = \mathbf{R}_{xx}(0) \quad (6.45)$$

Für lineare, zeitinvariante Prozeßgleichungen und bei weißen Störprozessen mit gaußscher Amplitudenverteilung ist der Schätzwert optimal in dem Sinne, daß die Komponenten des Schätzfehlers minimale Varianzen aufweisen. Im nichtlinearen Fall ist die Schätzung allerdings nur suboptimal. Die Schätzung des Verkehrszustandes bei Erweiterung der Ausgangsgleichungen durch die Meßgröße Reisezeit soll im folgenden erläutert werden.

6.2.4 Verkehrszustandsschätzung mit Reisezeitinformation

Mit dem Einsatz des erweiterten Kalman-Filters kann das Schätzproblem für den Verkehrsfluß auf Schnellstraßen gelöst werden. Der Verkehrsablauf werde durch das stochastische Verkehrsflußmodell (6.31), (6.32) repräsentiert, wobei der Ausgangsvektor die in Gleichung (6.26) beschriebene Form besitzt. Die Modellparameter sind nach Tabelle 6.1 festgelegt. Das erweiterte Kalman-Filter bildet den Verkehrsfluß nach dem deterministischen Modell ($\mathbf{w}(k) = \mathbf{n}(k) = \mathbf{0}$) nach und korrigiert die Schätzwerte in jedem Zeitschritt durch den Korrekturterm.

Die für diesen Fall einzusetzenden Werte für die Matrizen $\mathbf{F}(k)$, $\mathbf{H}(k)$, $\mathbf{\Gamma}(k)$, $\mathbf{R}_{ww}(k)$ und $\mathbf{R}_{nn}(k)$ finden sich in [22]. Dabei müssen die Kovarianzmatrix $\mathbf{R}_{nn}(k)$ und die Linearisierung $\mathbf{H}(k)$ jeweils in jeder Komponente um eine Dimension vergrößert werden, um die Meßgleichung für die Reisezeit in das Filterproblem einzubeziehen.

Mit den stochastischen Korrekturtermen ξ_j und ζ_j hat die Kovarianzmatrix $\mathbf{R}_{mn}(k)$ für die Meßabgriffe w_0 , q_n , w_n und t_r folgende Struktur:

$$\mathbf{R}_{mn}(k) = \begin{bmatrix} E\{\zeta_{w_0}^2(k)\} & 0 & 0 & 0 \\ 0 & E\{\zeta_{q_n}^2(k)\} + E\{\xi_n^2(k)\} & 0 & 0 \\ 0 & 0 & E\{\zeta_{w_n}^2(k)\} & 0 \\ 0 & 0 & 0 & E\{\zeta_{t_r}^2(k)\} \end{bmatrix}. \quad (6.46)$$

Im stationären Geschwindigkeitsprofil ergab sich Gleichung (6.21) als kombinierte Meßgleichung für die mittlere Reisezeit. Deshalb muß die Jacobische Matrix $\mathbf{H}(k)$ aus (6.39) um die Werte

$$\frac{\partial t_{r_j}(k)}{\partial v_j(k)} = \begin{cases} -\frac{\Delta_j}{v_j^2(k)} & \text{für } v_j(k) \geq v_{\min} \\ -\frac{4 v_{\min} v_j(k) \Delta_j}{(v_j^2(k) + v_{\min}^2)^2} & \text{für } 0 \leq v_j(k) < v_{\min} \end{cases} \quad (6.47)$$

erweitert werden.

Die Vorgehensweise der modellgestützten Verkehrszustandsschätzung mit dem erweiterten Kalman-Filter wird im Blockdiagramm in Bild 6.5 verdeutlicht.

Anhand von realistischen Situationen soll im folgenden das Filter für die Modellerweiterung durch die Fahrzeugreisezeit erprobt werden. Im Vordergrund steht dabei die Frage der Verbesserung der Schätzung durch die zusätzliche Reisezeitinformation. Diese Frage soll zunächst durch die Berechnung eines Beobachtbarkeitsmaßes theoretisch beantwortet werden.

6.3 Beobachtbarkeitsanalyse

Die aus den Meßgrößen gewonnene Information über den Verkehrszustand in einem Straßenabschnitt kann quantitativ bewertet werden. Hierzu wurden in den letzten Jahrzehnten Beobachtbarkeitsmaße eingeführt, die die Beobachtbarkeit der einzelnen Zustandsvariablen eines Systems in einer gegebenen Konfiguration bewerten.

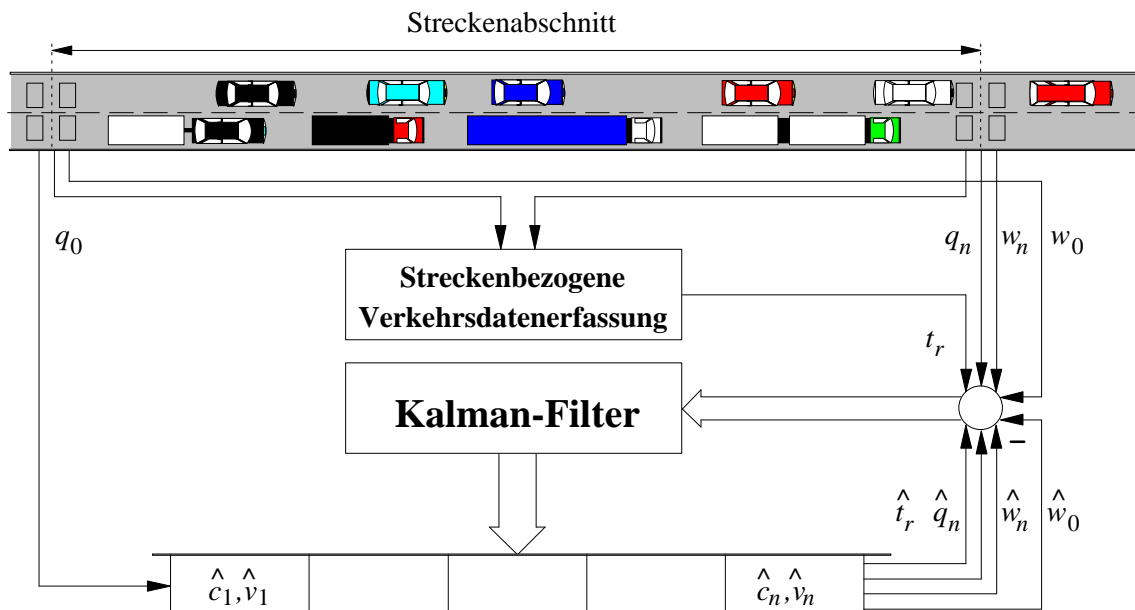


Bild 6.5: Zustandsschätzung für ein Verkehrsflußmodell mit dem erweiterten Kalman-Filter

Ein in [22] eingeführtes und in [15] in vielen Anwendungsfällen eingesetztes Beobachtbarkeitsmaß soll in diesem Kapitel nach einer kurzen Herleitung zur Abschätzung der Beobachtbarkeitsgüte für das um streckenbezogene Verkehrsdaten erweiterte Verkehrsmodell benutzt werden.

6.3.1 Theoretische Grundlagen

Definition des Beobachtbarkeitsmaßes

Nach [22] und [14] wird ein lineares, zeitinvariantes, zeitdiskretes System ohne Eingangsgrößen betrachtet:

$$\mathbf{x}(k+1) = \mathbf{F}\mathbf{x}(k), \quad \mathbf{x} \in \mathbb{R}^n, \quad (6.48)$$

$$\mathbf{y}(k) = \mathbf{H}\mathbf{x}(k), \quad \mathbf{y} \in \mathbb{R}^m, \quad (6.49)$$

wobei $\mathbf{x}(k)$ den Zustandsvektor und $\mathbf{y}(k)$ den Meßvektor zum Zeitpunkt kT repräsentieren. Die Matrizen \mathbf{F} und \mathbf{H} sind die Systemmatrizen entsprechender Dimension.

Der Meßinformationsvektor \mathbf{y} zum Zeitpunkt $K T$ läßt sich ausgehend von einem Anfangszustand $\mathbf{x}(0)$ durch wiederholte Anwendungen von (6.48) und (6.49) zu

$$\mathbf{y} = \begin{bmatrix} \mathbf{y}(0) \\ \mathbf{y}(1) \\ \vdots \\ \mathbf{y}(K) \end{bmatrix} = \begin{bmatrix} \mathbf{H} \\ \mathbf{H} \mathbf{F} \\ \vdots \\ \mathbf{H} \mathbf{F}^K \end{bmatrix} \mathbf{x}(0) = \mathbf{Q} \mathbf{x}(0) \quad (6.50)$$

berechnen, wobei \mathbf{Q} für $K = n - 1$ Beobachtbarkeitsmatrix genannt wird. Ist der Rang der Matrix \mathbf{Q} gleich der Ordnung n des Systems, so heißt das System (6.48), (6.49) vollständig beobachtbar.

Das Maß der Beobachtbarkeit einer Zustandsvariablen x_i geht aus der Betrachtung des Informationsvektors \mathbf{y}_i hervor, der sich errechnen läßt, wenn zum Zeitpunkt $t = 0$ nur die i -te Zustandsvariable um eine Einheit aus der Ruhelage ausgelenkt wird.

$$\mathbf{y}_i = \mathbf{y}|_{\mathbf{x}(0) = \mathbf{e}_i} = \mathbf{Q} \mathbf{e}_i = \mathbf{q}_i \quad (6.51)$$

\mathbf{e}_i ist der i -te Einheitsvektor, \mathbf{q}_i die i -te Spalte der Matrix \mathbf{Q} . Gesucht wird ein Maß, das ausdrückt, wie gut sich der Meßinformationsvektor \mathbf{y}_i von den Meßinformationsvektoren \mathbf{y}_i^* unterscheidet, die sich aus beliebigen Linearkombinationen der übrigen Zustandsvektoren berechnen lassen.

$$\mathbf{y}_i^* = \sum_{\substack{j=1 \\ j \neq i}}^n \alpha_j \mathbf{q}_j \quad (6.52)$$

Insbesondere interessiert die Unterscheidbarkeit von demjenigen Meßfolgenvektor \mathbf{y}_i^* , der dem Vektor \mathbf{y}_i im Sinne der euklidischen Norm am ähnlichsten ist. Das Beobachtbarkeitsmaß $\beta(x_i)$ für die i -te Zustandsvariable x_i wird damit durch die Norm

$$\beta(x_i) := \min_{\substack{\alpha_j \\ j \neq i}} \|\mathbf{y}_i^* - \mathbf{y}_i\|_2 \quad (6.53)$$

definiert. Je größer die Maßzahl $\beta(x_i)$ ist, desto besser läßt sich die Zustandsvariable x_i aus den Meßwerten rekonstruieren. Für den Fall $\beta(x_i) = 0$ ist die Zustandsvariable x_i nicht beobachtbar.

Berechnung des Beobachtbarkeitsmaßes

Das Beobachtbarkeitsmaß läßt sich für lineare, zeitinvariante, zeitdiskrete Systeme in geschlossener Form für den Fall berechnen, daß die Anzahl der Zeitschritte K gegen ∞ strebt.

Nach [15] gilt für das Beobachtbarkeitsmaß $\beta(x_i)$ der i -ten Zustandsvariable

$$\beta(x_i) = ((\mathbf{R}^{-1})_{ii})^{-\frac{1}{2}}, \quad (6.54)$$

$(\mathbf{R}^{-1})_{ii}$ ist hierbei das i -te Diagonalelement der Inversen der Matrix \mathbf{R} .

Die Matrix \mathbf{R} ist für asymptotisch stabile Systeme positiv-(semi-)definit und symmetrisch. Sie ist die Lösung der Ljapunov-Gleichung

$$\mathbf{F}^T \mathbf{R} \mathbf{F} - \mathbf{R} = -\mathbf{H}^T \mathbf{H}. \quad (6.55)$$

Für nichtlineare, zeitinvariante, zeitdiskrete Systeme sind die Systemmatrizen \mathbf{F} und \mathbf{H} die Jacobi-Matrizen der Vektorfunktionen $\mathbf{f}(\cdot)$ und $\mathbf{g}(\cdot)$ in vorgegebenen Arbeitspunkten.

Die Berechnung des Beobachtbarkeitsmaßes beschränkt sich somit auf die Lösung der Ljapunov-Gleichung (6.55) und einer Matrixinversion (6.54). Die Anwendung des Beobachtbarkeitsmaßes auf das Verkehrsflußmodell soll im folgenden im Hinblick auf die Erweiterung der Messungen der Fahrzeugreisezeiten diskutiert werden.

6.3.2 Analyse des erweiterten Verkehrsflußmodells

Für das Verkehrsflußmodell ist das Beobachtbarkeitsmaß abhängig von der Segmentanzahl n (und damit von der Dimension des Systems) und von der Anzahl der Meßgrößen m (Dimension des Meßvektors). Der gewählte Arbeitspunkt der linearisierten Systembeschreibung hingegen hat wenig Einfluß auf die Beobachtbarkeit der einzelnen Variablen [22]. Die Qualität des Schätzverfahrens soll nun für unterschiedliche Konfigurationen durch Anwendung des Beobachtbarkeitsmaßes untersucht werden. Damit die Zahlenwerte der makroskopischen Verkehrsgrößen gleiche Größenordnungen besitzen, werden für alle Zahlenangaben die Werte wie folgt normiert:

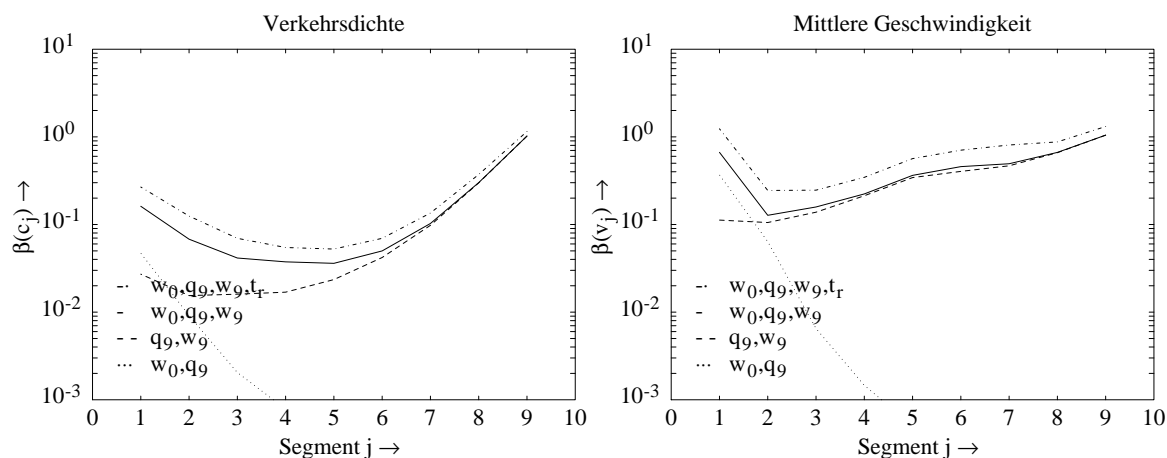


Bild 6.6: Beobachtbarkeitsmaß für die Verkehrsdichten und mittleren Geschwindigkeiten bei Auswertung verschiedener Kombinationen von Meßgrößen für 9 Segmente (4,5 km)

Verkehrsdichte:	$[Fz/100 \text{ m}]$,
Mittlere Geschwindigkeit:	$[100 \text{ m}/5 \text{ s}]$,
Verkehrsstärke:	$[Fz/5 \text{ s}]$.

Abhängigkeit von der Anzahl der Meßgrößen

Nach [22] werden die Größen w_0 , q_n und w_n als Meßgrößen und die Verkehrsstärke q_0 als die den Verkehrsablauf bestimmende Eingangsgröße betrachtet. Durch die Modellerweiterung liegt mit der mittleren Reisezeit t_r eine weitere Meßgröße vor.

Bild 6.6 zeigt den Verlauf des Beobachtbarkeitsmaßes für die Verkehrsdichten und die mittleren Geschwindigkeiten für verschiedene Kombinationen von Meßgrößen bei einem Arbeitspunkt von 2000 Fz/h. Der betrachtete zweispurige Streckenabschnitt ist in 9 Segmente unterteilt und besitzt eine Länge von 4,5 km. Deutlich erkennbar ist, daß die Erweiterung der Meßgrößen durch die Reisezeit t_r die Beobachtbarkeit des Systems entscheidend verbessert. Gleiche Ergebnisse ergeben sich auch bei anderen Streckenlängen.

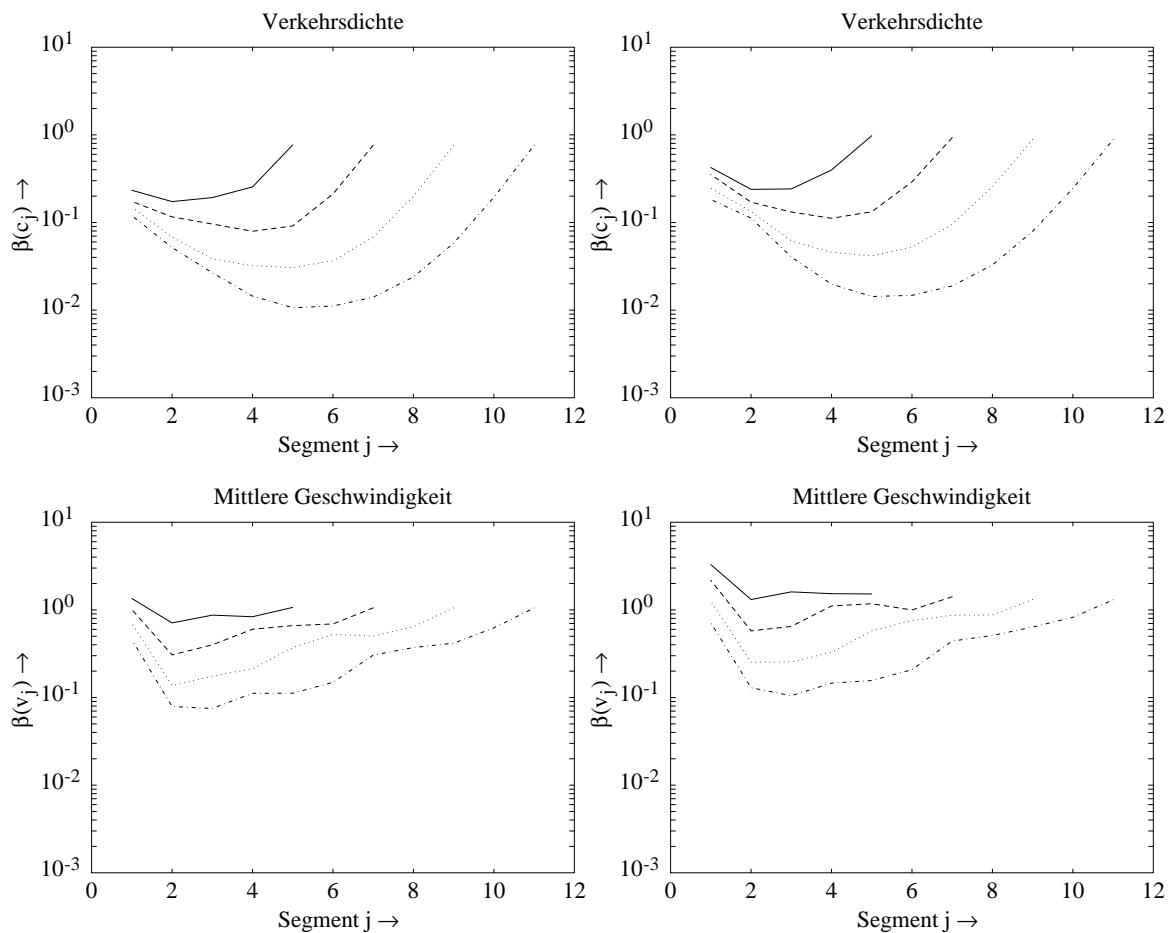


Bild 6.7: Beobachtbarkeitsmaß für die Verkehrsdichten (oben) und die mittleren Geschwindigkeiten (unten) in Abhängigkeit der Segmentanzahl für 2000 Fz/h ohne Reisezeitinformation (links) und mit Reisezeitinformation (rechts)

Abhängigkeit von der Segmentanzahl

In Bild 6.7 wird der Verlauf des Beobachtbarkeitsmaßes für die Verkehrsdichten und die mittleren Geschwindigkeiten für drei und vier Meßgrößen (mit Reisezeit) in Abhängigkeit der Anzahl der Segmente bei einem Arbeitspunkt von 2000 Fz/h gezeigt. Deutlich wird, daß die Beobachtbarkeit in beiden Fällen mit zunehmender Streckenlänge sinkt. Außerdem wirkt sich die Tendenz, daß die Zustandsvariablen im Innern des Streckenabschnitts schlechter beobachtbar sind als an den Rändern, mit zunehmender Segmentanzahl aus. Aus dem Vergleich der Werte ohne und mit

Reisezeitinformation zeigt sich allerdings, daß bei einer Erweiterung der Meßgrößen um die Reisezeit eine Vergrößerung der Sensorabstände um einige Segmente theoretisch die Beobachtbarkeit einzelner Zustandsgrößen kaum mindert.

6.4 Simulationsergebnisse

Zur Durchführung der Filtererprobung wird ein mikroskopisches Verkehrsmodell herangezogen, welches für verschiedenartige Verkehrszustände makroskopische Verkehrsgrößen generiert. Die aus dem mikroskopischen Modell hervorgehenden makroskopischen Größen gehen zum einen als Meßgrößen in die Filtergleichungen ein, zum anderen werden die Zustandsgrößen mit den vom erweiterten Kalman-Filter berechneten Schätzwerten verglichen. Somit ist eine Aussage über die Schätzgüte möglich. Vor einem Vergleich der erzielbaren Genauigkeiten der Zustandsschätzung ohne und mit Reisezeitinformation und der Darstellung der Störungsdetektion mit Hilfe streckenbezogener Verkehrsdaten muß die Varianz der Reisezeitmessung bestimmt werden.

6.4.1 Parameterfestlegung

Als freie Parameter des Schätzverfahrens sind die Varianzen des Störprozesses am Systemausgang anzusehen, die in der Kovarianzmatrix $\mathbf{R}_{nn}(k)$ nach Gleichung (6.46) aufgeführt sind. Anhaltswerte für die Varianzen $E\{\xi_n(k)\}^2$, $E\{\zeta_{w_0}(k)\}^2$, $E\{\zeta_{q_n}(k)\}^2$ und $E\{\zeta_{w_n}(k)\}^2$ finden sich in [22]. Die aus den Meßfehlern und den Abweichungen von den tatsächlichen Werten resultierende Varianz $E\{\zeta_{t_r}(k)\}^2$ kann durch die Filtererprobung festgelegt werden.

Streng genommen ist die Varianz der mittleren Reisezeit von der Verkehrsdichte und vom Verkehrszustand abhängig. Im Rahmen der folgenden Betrachtungen wird allerdings von einer konstanten Varianz ausgegangen. Eine zu groß gewählte Varianz der Reisezeiten bewirkt, daß der Unterschied zur Schätzung ohne Reisezeit nicht mehr erkennbar ist. Eine

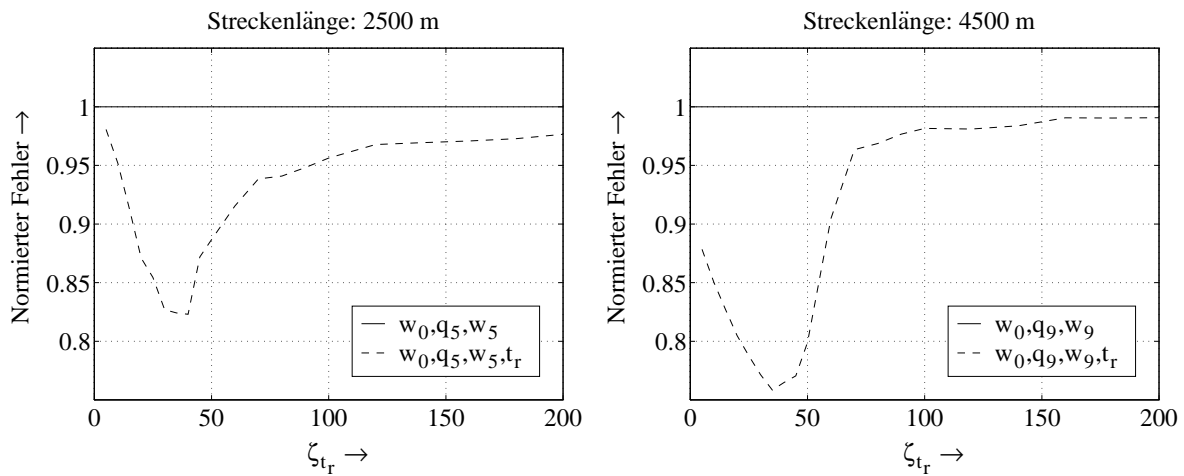


Bild 6.8: Normierter quadratischer Schätzfehler in Abhängigkeit von ζ_{t_r} für zwei verschiedene Streckenabschnitte

zu klein gewählte Varianz der Reisezeit führt zu einer zu starken Gewichtung der Reisezeitinformation. Eine Optimierung der Varianz kann durch die Berechnung des normierten quadratischen Schätzfehlers der Zustandsvariablen anhand mehrerer Simulationen durchläufe verschiedener Topologien erfolgen. Bild 6.8 zeigt beispielhaft den normierten quadratischen Schätzfehler für Schätzungen mit und ohne Reisezeitinformation in Abhängigkeit von ζ_{t_r} für Simulationen bei einer Streckenlänge von 2,5 km und 4,5 km bei einer Eingangsverkehrsstärke von 2500 Fz/h. Da hier nur die Relation der Schätzfehler interessiert, wurde der Schätzfehler ohne Reisezeitinformation auf Eins normiert.

Aus mehreren Simulationen folgt, daß das Minimum im Bereich von $\zeta_{t_r} \approx 40$ s liegt, d.h. die Varianz kann mit $E\{\zeta_{t_r}^2(k)\} \approx 1600$ s² festgelegt werden. Die Varianz der mittleren Reisezeit, die aus den Meßfehlern resultiert, liegt bei den in Kapitel 5 beschriebenen Verfahren bei ca. 36 s². Der übrige Anteil an der Varianz resultiert aus den Abweichungen der tatsächlichen mittleren Reisezeiten zu den aus Gleichung (6.16) errechneten Werten. Da die mittlere Reisezeit nicht auf die Streckenlänge normiert ist, muß dieser Wert je nach Streckenlänge geringfügig verkleinert oder vergrößert werden.

6.4.2 Ergebnisse der Verkehrszustandsschätzung

In diesem Abschnitt wird exemplarisch eine Simulation für einen zweispurigen, 5,5 km langen Streckenabschnitt betrachtet und den theoretischen Untersuchungen gegenübergestellt. Es wird eine Simulation von 30 Minuten Dauer zugrunde gelegt. Nach 10 Minuten wird für die Dauer von 10 Minuten ein Unfall durch die Sperrung einer Fahrspur simuliert. Der Störbereich befindet sich im siebten von 11 Segmenten. Die Eingangsverkehrsstärke ist mit 2500 Fz/h relativ hoch, so daß es durch die Störung zu einem Stau im Streckenabschnitt kommt. Dieses wirkt sich nach 10 Minuten direkt auf die Zustandsgrößen im sechsten Segment aus. Es folgt zu einer Abnahme der mittleren Geschwindigkeit und einer Zunahme der Verkehrsdichte. Mit entsprechender Zeitverzögerung gilt das gleiche für die Zustandsgrößen in den anderen eingangsseitigen Segmenten. Das resultierende Verkehrsdichte- und Geschwindigkeitsprofil ist im Bild 6.9 dargestellt.

Im Bild 6.9 sind zudem die ohne und mit Reisezeitinformation geschätzten Werte der Verkehrsdichten und der mittleren Geschwindigkeiten aufgetragen. Man erkennt, daß beide Schätzverfahren auf die Störung im Streckenabschnitt reagieren. Deutlich wird aber, daß der Anstieg der Verkehrsdichten am Anfang des Streckenabschnitts beim Schätzverfahren ohne Reisezeitinformation nicht erkannt wird. Auch die Abnahme der mittleren Geschwindigkeiten wird nicht so gut geschätzt wie bei Kenntnis der Reisezeit. Aus den Schätzergebnissen mit Reisezeitinformation hingegen zeigt sich, daß auch bei großen Streckenlängen noch gute Schätzergebnisse für den Verkehrszustand erzielt werden können. Diese Ergebnisse decken sich mit den Resultaten aus der Beobachtbarkeitsanalyse in den Bildern 6.6 und 6.7.

Auch weitere Simulationen für verschiedene Konstellationen belegen, daß die Erweiterung mit streckenbezogenen Verkehrsdaten zu einer Verbesserung der Verkehrszustandsschätzung führt [83].

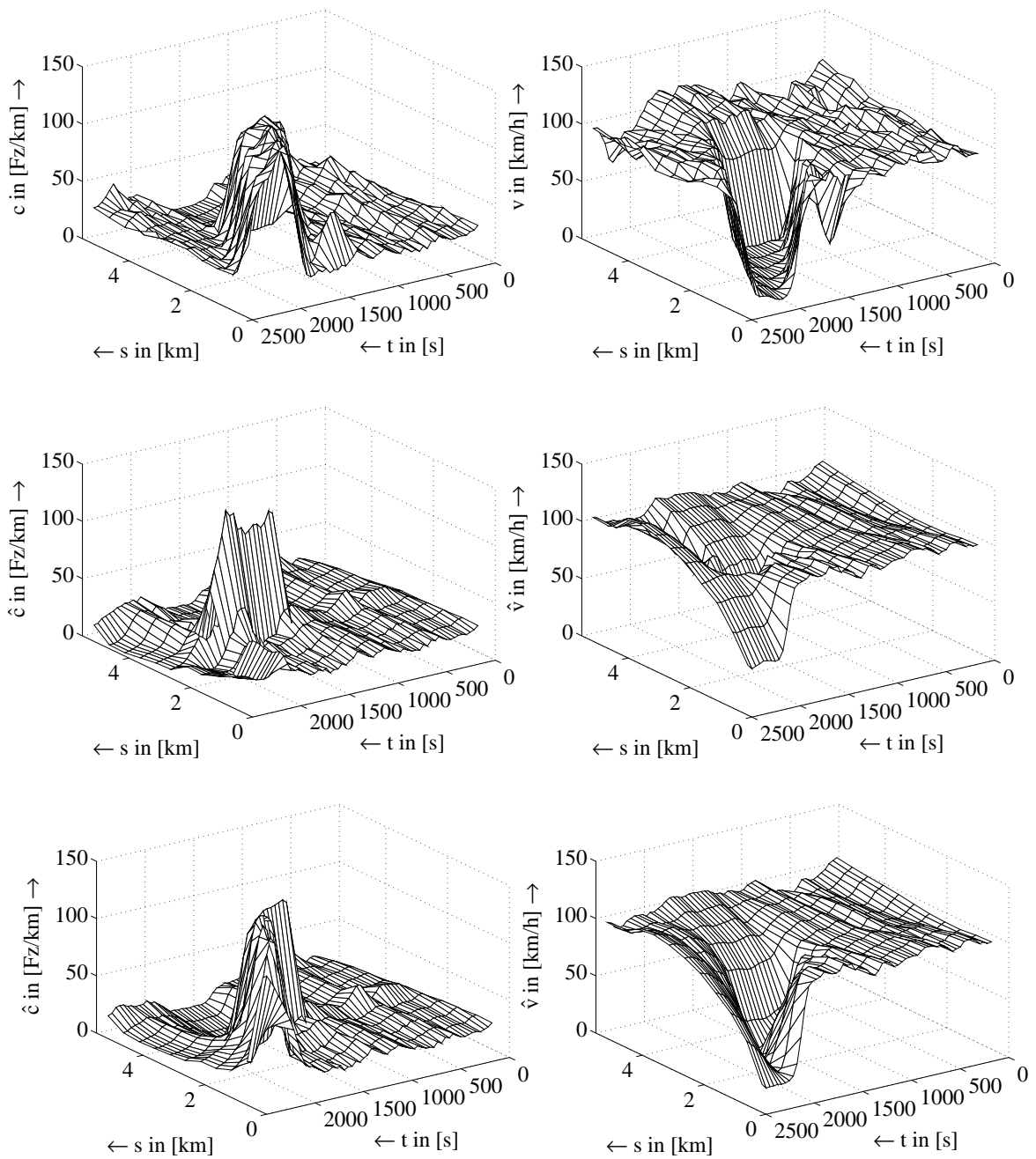


Bild 6.9: Von oben nach unten: Wahres, ohne und mit Reisezeitinformation geschätztes Verkehrsdichteprofil (links) und Geschwindigkeitsprofil (rechts)

6.4.3 Ergebnisse der Störungserkennung

Ein Vorteil der Modellerweiterung mit dem hypothetischen Zufluß nach Abschnitt 6.1.3 liegt darin, daß neben der Schätzung des aktuellen Verkehrszustandes auch noch eine Störungserkennung durchgeführt werden kann. Grundlage der nachfolgenden Darstellungen ist wiederum eine Simulation von 30 Minuten Dauer, bei der nach 10 Minuten durch die Sperrung einer Fahrspur im mittleren Streckenbereich für den Zeitraum von 10 Minuten eine Störung erzeugt wird. Der Streckenabschnitt besitzt in diesem Fall eine Länge von 2,5 km und die Eingangsverkehrsstärke variiert um den Wert von 2500 Fz/h.

In [73] wurde ein Vergleich des Störungserkennungsverfahrens mit dem hypothetischen Zufluß mit zwei anderen Verfahren durchgeführt. Diese Verfahren, der *California-Algorithmus* [82] und die *räumliche Verkehrsstärkeprognose* [8], werden in [16] für den vorgegebenen Detektorabstand vorgeschlagen. Aus dem Vergleich in [73] geht hervor, daß die Schätzung des hypothetischen Zuflusses zur Störungserkennung für die hier vorliegende Simulation am besten geeignet ist.

Gleichzeitig kann aber auch die Reisezeitinformation zur Störungserkennung herangezogen werden. Ein deutlicher Anstieg der mittleren Reisezeiten ist signifikant für das Vorliegen einer Störung. Der Abfall der mittleren Reisezeiten kann als Störungsende interpretiert werden. In Bild 6.10 ist die aus der Simulation hervorgegangene mittlere Reisezeit der Fahrzeuge aufgetragen. Zur besseren Übersicht sind die zeitdiskreten Werte als Polygonzug dargestellt und der Zeitraum der Störung ist grau gekennzeichnet. Zudem ist die Reisezeit durch eine gleitende Mittelwertbildung über drei Abtastwerte geglättet worden.

Bildet man eine Differenz dieser Reisezeitwerte, $t_r(k) - t_r(k - \ell)$, so kann diese Reisezeitdifferenz bei günstiger Wahl der Zeitverzögerung sehr gut zur Störungserkennung eingesetzt werden. Für eine Zeitverzögerung von $\ell T = 2$ Minuten ist im Bild 6.11 die aus dieser Simulation resultierende Reisezeitdifferenz zusammen mit dem geglätteten hypothetischen Zufluß aufgetragen. Aus dem Bild wird deutlich, daß die Reisezeitdifferenz ungefähr das gleiche Zeitverhalten und einen ähnlichen Amplitu-

denanstieg liefert wie der hypothetische Zufluß. Das Zeitverhalten wird hierbei vom Abstand der Fahrspursperrung zum stromabwärtigen Meßquerschnitt und der damit verbundenen Verzögerung der Fahrzeugdetektion bestimmt. Zum besseren Vergleich der beiden Verfahren sind im Bild lediglich die positiven Werte der Reisezeitdifferenz aufgetragen.

Der Vorteil der Störungserkennung mit der Reisezeitinformation liegt jetzt darin, daß mit der mittleren Reisezeit der Fahrzeuge auch andersartige Störungen, die z.B. durch die Überlastung des Verkehrs entstehen, identifiziert werden können. In solchen Fällen ist eine sichere Störungserkennung mit dem hypothetischen Zufluß nicht mehr gewährleistet, da der fiktive Zufluß nur die herabgesetzte Kapazität durch die Sperrung einer Fahrspur im Zustandsmodell ausgleichen soll.

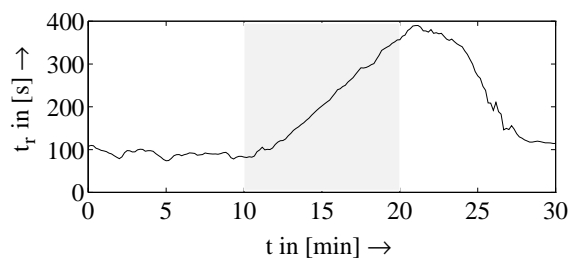


Bild 6.10: Mittlere Reisezeit

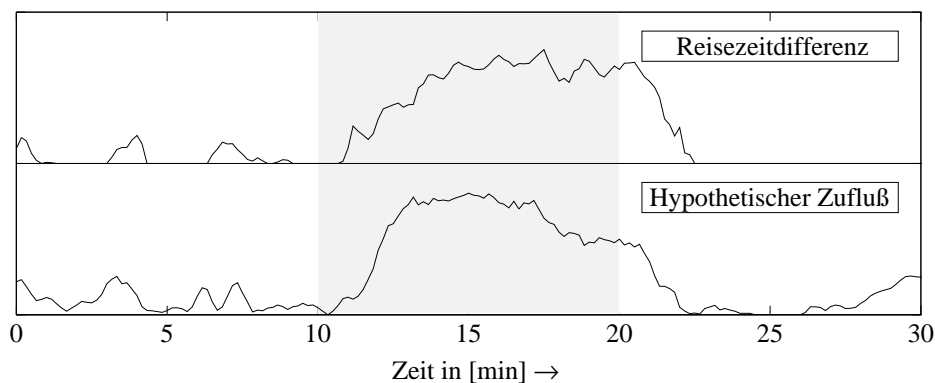


Bild 6.11: Störungserkennung mit der Reisezeitdifferenz und dem hypothetischen Zufluß

7 Zusammenfassung

Die Untersuchungen der vorliegenden Arbeit zeigen, daß für die Klassifikation und Wiedererkennung von Fahrzeugen mit Signalen von Induktionsschleifen effiziente Verfahren auf der Basis von Mustererkennungssystemen angegeben werden können. Anhand eines entscheidungstheoretischen Ansatzes wird die Gesamtaufgabe durch die Aufteilung in verschiedene Verarbeitungsstufen gegliedert. Beide Probleme basieren auf der gleichen Signalvorverarbeitung und einer ähnlichen Merkmalsextraktion, so daß diese Stufen in einem System zusammen realisiert werden könnten. Der Unterschied beider Probleme liegt in der Klassifikation, bei der entweder die Entscheidung für eine vorgegebene Fahrzeugklasse oder für ein Fahrzeug im Streckenabschnitt zur Ermittlung der streckenbezogenen Daten erwünscht ist.

Ausgangspunkt für die Klassifikation von Fahrzeugen ist die Bayes-Klassifikation, für die zur Bestimmung der optimalen Entscheidungsfunktionen die A-posteriori-Wahrscheinlichkeiten der Fahrzeugklassen bzw. alternativ die bedingten Dichten der Merkmalsvektoren sowie die A-priori-Wahrscheinlichkeiten der Klassen ermittelt werden müssen. Da die Schätzung der A-posteriori-Wahrscheinlichkeiten problematisch ist, beschränkt man sich auf die Schätzung der bedingten Dichten, die mit parametrischen oder nichtparametrischen Verfahren erfolgen kann. Anhand der wichtigsten Verfahren wird die Einsatzmöglichkeit verschiedener Klassifikatoren bezüglich der genannten Problemstellung untersucht und bewertet. Bezüglich der Merkmalsextraktion erweist sich die Karhunen-Loève-Transformation im Vergleich zu einer Diskriminanzanalyse für den vorliegenden Fall als die beste Methode. Es zeigt sich, daß mit einem zweistufigen Verfahren mit neuronalen Netzwerken aufgrund der guten binären Trennbarkeit der Fahrzeugklassen eine Fehlerrate des Klassifikators erreicht werden kann, die dem minimalen Bayes-Fehler sehr nahe kommt.

Die Schätzung streckenbezogener Verkehrsdaten gelingt mit der Wiedererkennung von Fahrzeugen oder Fahrzeugkollektiven im Streckenabschnitt. Im Vergleich zu einem Referenzverfahren, das mit einer Korrelationsanalyse zur Wiedererkennung von Fahrzeugkollektiven arbeitet, werden zwei verschiedene, neuartige Methoden entwickelt, die beide auf der Basis der individuellen Wiedererkennung von Fahrzeugen arbeiten. Das erste Verfahren arbeitet mit Kostenfunktionen, in die neben der Metrik der Merkmalsvektorendifferenz weitere charakteristische Fahrzeugdaten einbezogen werden. Zur Bestimmung der Kostenfunktionen werden die statistischen Abweichungen zwischen wahren und prognostizierten Werten analysiert und anschließend die optimalen Gewichtungsfaktoren bestimmt. Eine zweite vollständig andere Methodik berücksichtigt die statistischen Bindungen der Fahrzeuge im Streckenabschnitt, die mit Hidden-Markov-Modellen modelliert werden können. Eine unbekannte Fahrzeugsequenz wird anhand der beobachteten Merkmalsvektoren mit der dynamischen Programmierung mit maximaler Wahrscheinlichkeit ermittelt. Beide Verfahren werden mit Simulations- und Meßdaten bewertet. Aus einem Vergleich mit dem Referenzverfahren werden die Vorteile der neu entwickelten Verfahren abgeleitet und die Unterschiede der beiden neuen Verfahren aufgedeckt. Es zeigt sich, daß das Schätzverfahren mit Kostenfunktionen den geringsten Rechenaufwand besitzt, während das Verfahren mit Hidden-Markov-Modellen die geringste Verzögerungszeit zwischen Fahrzeugdetektion und Schätzung der Fahrzeugreisezeit hat.

Die Einsatzmöglichkeit der streckenbezogenen Verkehrsdatenerfassung wird in dieser Arbeit abschließend anhand der Verkehrszustandsschätzung mit einem modellgestützten Schätzverfahren beschrieben. Dieses bislang vorwiegend mit querschnittsbezogenen Meßdaten arbeitende makroskopische Abschnittsverfahren wird um streckenbezogene Meßdaten, die aus der mittleren Reisezeit der Fahrzeuge im betrachteten Streckenabschnitt bestehen, erweitert. Anhand einer theoretischen Analyse mit einem Beobachtbarkeitsmaß und Ergebnissen aus Simulationen mit und ohne Reisezeitinformation kann eine Verbesserung der Verkehrszustandsschätzung aufgezeigt werden. Gleichzeitig kann die Reisezeitinformation zur Detektion von Verkehrsstörungen eingesetzt werden.

Abkürzungs- und Symbolverzeichnis

Abkürzungen

DCT	Diskrete Cosinus-Transformation
DIS	Diskriminanzanalyse
LIN	Linearer Gaußklassifikator
HIST	Histogrammklassifikator
HMM	Hidden-Markov-Modell
IDCT	Inverse diskrete Cosinus-Transformation
KLT	Karhunen-Loève-Transformation
MAP	Maximum a posteriori
ML	Maximum Likelihood
MLP	Multi-Layer-Perceptron
NN	Nächster-Nachbar-Klassifikator
POLY	Polynomklassifikator
QUAD	Quadratischer Gaußklassifikator
SLP	Single-Layer-Perceptron
WT	Whitening-Transformation

Symbole

$a(t)$	Kontinuierliche Funktion der Zeit t
$a(n)$	Zeitdiskrete Folge
\mathbf{a}	Spaltenvektor
\mathbf{a}^T	Transponierter des Vektors \mathbf{a}
\mathbf{A}	Matrix
\mathbf{A}^T	Transponierte der Matrix \mathbf{A}
\mathbf{A}^{-1}	Inverse der Matrix \mathbf{A}
$\langle \mathbf{a}, \mathbf{b} \rangle$	Skalarprodukt der Vektoren \mathbf{a} und \mathbf{b}
$\ \mathbf{a}\ $	Norm des Vektors \mathbf{a}
$\hat{\mathbf{a}}$	Schätzung des Vektors \mathbf{a}
\mathbf{a}	Realisation eines Prozesses \mathbf{a}

$\mathbf{A} = \{a_{ij}\}$	Transitionsmatrix
$\mathbf{B} = \{b_j(\mathbf{o})\}$	Erzeugungswahrscheinlichkeitsmatrix
C_{ik}	Kostenfaktor
$c_j(k)$	Verkehrsdichte
\mathbf{c}	Zielvektor der Polynomklassifikation
const	Konstante
d	Kopfabstand der Induktionsschleifen
$d(\mathbf{x}, \mathbf{y}) = \ \mathbf{x} - \mathbf{y}\ $	Metrik der Vektoren \mathbf{x} und \mathbf{y}
$d_i(\mathbf{m})$	Entscheidungsfunktion
d_{opt}	Optimale Entscheidungsschwelle
$\det\{\mathbf{X}\}, \mathbf{X} $	Determinante der Matrix \mathbf{X}
e	Eulersche Zahl
\mathbf{e}	Schätzfehler
$E\{\cdot\}$	Erwartungswertoperator
\mathbf{g}	Gewichtsvektor
$h(n)$	Impulsantwort
H_i	Hypothese für das Eintreffen von k_i
\mathbf{I}	Einheitsmatrix
$J\{\cdot\}$	Gütemaß
k_i	Fahrzeugklasse
K	Anzahl der Fahrzeugklassen
$\mathbf{k}(\mathbf{x}, \mathbf{y})$	Kostenvektor
\mathbf{K}	Matrix der DCT
l_{Fz}	Fahrzeuglänge
l_S	Länge der Induktionsschleifen
μ	Mittelwert
$\boldsymbol{\mu}$	Mittelwertsvektor
\mathbf{m}	Merkmalsvektor
max	Maximalwert
min	Minimalwert
Median(\cdot)	Median eines Vektors
$\mathbf{O} = \{\mathbf{o}_1, \mathbf{o}_2, \dots, \mathbf{o}_T\}$	Beobachtungssequenz
p_k	A-priori-Wahrscheinlichkeit der Klasse k
$p(n, m)$	Geglättete Korrelationsfolge

$P(H_i, k_k)$	Verbundwahrscheinlichkeit für $\{H_i, k_k\}$
$p_{\mathbf{m}}(\mathbf{m})$	Verbunddichte der Komponenten eines zufälligen Vektors \mathbf{m}
$p_{\mathbf{m} H_i}(\mathbf{m} H_i)$	Dichte eines zufälligen Vektors \mathbf{m} unter der Bedingung, daß H_i eingetreten ist
$P(q_t = i q_{t-1} = j)$	Zustandsübergangswahrscheinlichkeit
$q_j(k)$	Verkehrsstärke
$Q = \{q_1, q_2, \dots, q_N\}$	Zustandssequenz
$\mathbf{r}(kT)$	Signalvektor mit Abtastwerten
$\mathbf{r}_N(kT_N)$	Normierter Signalvektor mit Abtastwerten
$r_{xy}(\tau)$	Kreuzkorrelationsfunktion von $x(t)$ und $y(t)$
$r_{xy}(m)$	Kreuzkorrelationsfolge von $x(n)$ und $y(n)$
$r(n, m)$	Zweidimensionale Korrelationsfolge
R_i	Raumgebiet von k_i
\mathbf{R}_{xx}	Autokovarianzmatrix des Prozesses \mathbf{x}
\mathbf{R}_{xy}	Kreuzkovarianzmatrix der Prozesse \mathbf{x}, \mathbf{y}
\mathcal{R}	Risiko
$\text{spur}\{\mathbf{X}\}$	Spur einer Matrix \mathbf{X}
\mathbf{s}	Sondermerkmalsvektor
\mathbf{S}_B	Interklassen-Scattermatrix
\mathbf{S}_M	Gemeinsame Scattermatrix
\mathbf{S}_W	Intraklassen-Scattermatrix
$t_{r_{fz}}$	Reisezeit eines Fahrzeugs
$t_r(k)$	Mittlere Reisezeit
T	Abtastintervall
\mathbf{u}	Merkmalsvektor zur Wiedererkennung mit Kostenfunktionen
\mathbf{U}	Matrix der Karhunen-Loève-Transformation
$v_j(k)$	Mittlere Geschwindigkeit
v_{Fz}	Fahrzeuggeschwindigkeit
$V(c_j)$	Geschwindigkeits-Dichte-Charakteristik
\mathbf{v}	Merkmalsvektor zur Wiedererkennung
$w_j(k)$	Lokale Geschwindigkeit
w_{ij}	Gewichte eines neuronalen Netzwerks

W	Wartezeit
$\mathbf{w}(m)$	Polynomvektor
\mathbf{W}	Matrix der Whitening-Transformation
\mathbf{y}	Merkmalsvektor aus Karhunen-Loève-Transformation
$\alpha_t(i)$	Vorwärts-Variable
$\beta_t(i)$	Rückwärts-Variable
$\beta(x_i)$	Beobachtbarkeitsmaß für x_i
δ_{ij}	Kronecker-Delta
$\delta_t(i)$	Maximale Wahrscheinlichkeit
Δ_j	Segmentlänge
ΔL	Induktivitätsänderung
ε	Fehlermaß
ε_B	Bayes-Fehler
$\varepsilon_C(s)$	Chernoff-Grenze
$\Gamma(k)$	Gammafunktion
$\lambda = (\mathbf{A}, \mathbf{B}, \boldsymbol{\pi})$	Hidden-Markov-Modell
λ_i	i -ter Eigenwert einer quadratischen Matrix
$\boldsymbol{\pi} = \{\pi_i\}$	Anfangswahrscheinlichkeitsvektor
$\varphi(v_j)$	Aktivierungsfunktion
$\psi_t(i)$	Argument
τ_{Fz}	Signallänge
τ_S	Zeitdifferenz
Θ_j	Schwellwert oder Bias
\mathbb{C}	Menge der komplexen Zahlen
\mathbb{N}	Menge der natürlichen Zahlen
\mathbb{R}	Menge der reellen Zahlen
\mathbb{Z}	Menge der ganzen Zahlen
$\mathbf{0}$	Nullvektor, Nullmatrix

Literaturverzeichnis

- [1] N. Ahmed, T. Natarajan und K. R. Rao: "Discrete cosine transform," *IEEE Trans. Computers*, Vol. 23, S. 88–93, Januar 1974.
- [2] M. S. Ahmed und A. R. Cook: "Analysis of freeway traffic time series data by using Box-Jenkins techniques," *Transp. Res. Rec.*, No. 722, S. 1–9, 1979.
- [3] B. D. Anderson und J. B. Moore: *Optimal Filtering*. Englewood Cliffs, NJ: Prentice Hall, 1979.
- [4] L. R. Bahl, F. Jelinek und R. L. Mercer: "A maximum likelihood approach to continuous speech recognition," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. PAMI-5, S. 179–190, 1983.
- [5] L. E. Baum und T. Petrie: "Statistical inference for probabilistic functions of finite state Markov chains," *Ann. Math. Stat.*, Vol. 37, S. 1554–1563, 1966.
- [6] L. E. Baum, T. Petrie, G. Soules und N. Weiss: "A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains," *Ann. Math. Stat.*, Vol. 41, No. 1, S. 164–171, 1970.
- [7] L. E. Baum: "An inequality and associated maximization technique in statistical estimation for probabilistic functions of a Markov process," *Inequalities*, Vol. 3, No. 1, S. 1–8, 1972.
- [8] R. Böttger: "Ein Verfahren zur meßtechnischen Feststellung von Verkehrsstörungen auf Fernstraßen und Autobahnen," *Straßenverkehrstechnik*, Juni 1979.
- [9] P. Böhnke: *Beitrag zu einer Systemtheorie von Objektfunktionen*. Dissertation, RWTH Aachen, 1980.
- [10] R. Böker: *Physikalische und informationstheoretische Untersuchung von Induktionsschleifen zur Wiederentdeckung von Fahrzeugen*. Dissertation, RWTH Aachen, 1984.
- [11] P. Böhnke, R. Böker, E. Pfannerstill, W. Stremme und R. Ziegler: "Ein Verkehrsdatenerfassungssystem für das Straßennetz – Konzept und Realisierung," *Frequenz*, Vol. 41, S. 88–94, April 1987.

-
- [12] K. Brammer und G. Siffing: *Stochastische Grundlagen des Kalman-Bucy-Filters*. München: Oldenbourg, 2. Auflage, 1986.
- [13] K. Brammer und G. Siffing: *Kalman-Bucy-Filter*. München: Oldenbourg, 3. Auflage, 1989.
- [14] M. Bredehöft: "Entwurf eines Beobachtersystems für zeitvariable Strömungsvorgänge zur Überwachung der Schadstoffbelastung in Flüssen," Schriftenreihe der Arbeitsgruppe Automatisierungstechnik, Heft 3, TU Hamburg-Harburg, 1988.
- [15] M. Bredehöft und M. Cremer: "Ein physikalisches Beobachtbarkeitsmaß für den Entwurf von Zustandsbeobachtern für Systeme von größerer Komplexität," *Automatisierungstechnik*, No. 2, 1988.
- [16] F. Busch: *Automatische Störungserkennung auf Schnellverkehrsstraßen*. Dissertation, Universität Karlsruhe, 1986.
- [17] F. Busch und M. Fellendorf: "Automatic incident detection on motorways," *Traffic Engineering and Control*, S. 221–227, April 1990.
- [18] R. L. Cheu, S. G. Ritchie, W. W. Recker und B. Bavarian: "Investigation of a neural network model for freeway incident detection," in *Second International Conference on the Application of Artificial Intelligence Techniques to Civil and Structural Engineering*, Oxford, September 1991, S. 3–5.
- [19] A. R. Cook und D. E. Cleveland: "The detection of freeway capacity-reducing incidents by traffic stream measurements," *Transp. Res. Rec.*, No. 495, S. 1–11, 1974.
- [20] T. M. Cover und P. E. Hart: "Nearest neighbor pattern classification," *IEEE Trans. Information Theory*, Vol. IT-13, S. 21–27, Januar 1967.
- [21] M. Cremer: "A new scheme for traffic flow estimation and control with a two component model," in *Proc. of the IFAC/IFIP/IFORS 3rd Symposium on Control in Transportation Systems*, Columbus, Ohio, August 1976, S. 29–37.
- [22] M. Cremer: *Der Verkehrsfluß auf Schnellstraßen*. Berlin: Springer, 1979.
- [23] M. Cremer: "Incident detection of freeways by filtering techniques," in *Preprints of the 8th IFAC Congress*, Kyoto, 1981.

-
- [24] M. Cremer und M. Papageorgiou: "Parameter identification for a traffic flow model," *Automatica*, Vol. 17, No. 6, S. 837–843, 1981.
- [25] M. Cremer und K. Putensen: "Monitoring of traffic density profiles using measurements from loop detectors and vehicle trip data," in *Proc. Int. Conf. on Advanced Traffic Management Systems*, St. Petersburg, Florida, Oktober 1993, S. 241–258.
- [26] M. Cremer und K. Putensen: "An approach for traffic state estimation using invehicle information," in *Proc. 7th Symposium on Transportation Systems*, Tianjin, China, August 1994, S. 643–648.
- [27] M. Cremer und S. Schrieber: "Monitoring traffic load profiles with heterogeneous data source configurations," in *Proc. 13th Int. Symposium on Transport and Traffic Theory*, Lyon, Juli 1996.
- [28] M. Dalgeish und R. Tuthill: "The development of a microprocessor-based vehicle classifier," *Traffic, Engineering, and Control*, Vol. 19, No. 3, 1978.
- [29] A. P. Dempster: "Maximum likelihood from incomplete data via the EM algorithm," *J. Royal Stat. Soc. B.*, Vol. 39, S. 1–18, 1977.
- [30] P. A. Devijver und J. Kittler: *Pattern Recognition: A Statistical Approach*. London: Prentice-Hall, 1982.
- [31] R. O. Duda und P. E. Hart: *Pattern Classification and Scene Analysis*. New York: Wiley, 1973.
- [32] R. J. Elliott, L. Aggoun und J. B. Moore: *Hidden Markov Models*. New York: Springer, 1995.
- [33] N. Fliege: *Systemtheorie*. Stuttgart: Teubner, 1991.
- [34] N. Fliege: *Multiraten-Signalverarbeitung*. Stuttgart: Teubner, 1993.
- [35] G. D. Forney: "The Viterbi algorithm," *Proc. IEEE*, Vol. 61, S. 268–278, März 1973.
- [36] K.-S. Fu: *Digital Pattern Recognition*. Berlin: Springer, 1980.
- [37] K. Fukunaga: *Introduction to Statistical Pattern Recognition*. Boston: Academic Press, 1990.

- [38] G. J. Gibson und C. F. N. Cowan: "On the decision regions of multilayer perceptrons," *Proc. IEEE*, Vol. 78, S. 1590–1594, Oktober 1990.
- [39] F. L. Hall: "An interpretation of speed-flow-concentration relationships using catastrophe theory," *Transportation Research A*, No. 21A, S. 191–201, 1987.
- [40] J. C. Hancock und P. A. Wintz: *Signal Detection Theory*. New York: McGraw-Hill, 1966.
- [41] S. S. Haykin: *Neural Networks*. New York: Macmillan, 1994.
- [42] F. Jelinek: "Continuous speech recognition by statistical methods," *Proc. IEEE*, Vol. 64, S. 532–536, April 1976.
- [43] B. H. Juang und L. R. Rabiner: "The segmental k -means algorithm for estimating parameters of hidden Markov models," *IEEE Trans. Acoustics, Speech, and Signal Processing*, Vol. ASSP-38, S. 1639–1641, September 1990.
- [44] K. D. Kammeyer: *Nachrichtenübertragung*. Stuttgart: Teubner, 1992.
- [45] J. F. Kingman: *Poisson Processes*. Oxford: Clarendon Press, 1993.
- [46] K. Kroschel: *Statistische Nachrichtentheorie, Band 2: Signalschätzung*. Berlin: Springer, 2. Auflage, 1988.
- [47] R. Kühne und S. Immes: "Freeway control systems using section-related traffic variable detection," in *Proc. Pacific Rim TransTech Conference*, Seattle, WA, Juli 1993, S. 56–62.
- [48] S. Y. Kung: *Digital Neural Networks*. Englewood Cliffs, NJ: Prentice-Hall, 1993.
- [49] M. Levin und G. M. Krause: "Incident detection: A Bayesian approach," *Transp. Res. Rec.*, No. 682, S. 52–58, 1978.
- [50] S. E. Levinson, L. R. Rabiner und M. M. Sondhi: "An introduction to the application of the theory of probabilistic functions of a Markov process to automatic speech recognition," *Bell Systems Technical Journal*, Vol. 62, No. 4, S. 1035–1075, 1983.
- [51] R. P. Lippmann: "An introduction to computing with neural networks," *IEEE ASSP Magazine*, Vol. 4, S. 4–22, April 1987.

-
- [52] R. P. Lippmann: "Pattern classification using neural networks," *IEEE Communications Magazine*, S. 47–64, November 1989.
- [53] H.-L. Lou: "Implementing the Viterbi algorithm," *IEEE Signal Processing Magazine*, Vol. 12, S. 42–52, September 1995.
- [54] A. Mertins: *Signaltheorie*. Stuttgart: Teubner, 1996.
- [55] H. A. Moland: "Pre-Processing and Feature Extraction for Pattern Recognition in Traffic Control Systems," Diplomarbeit, TU Hamburg-Harburg, Arbeitsbereich Nachrichtentechnik, 1992.
- [56] F. P. D. Navin: "Traffic congestion catastrophes," *Transportation Planning and Technology*, No. 11, S. 19–25, 1986.
- [57] J. T. Olesik: *Use of Loop Detector Phase Signatures for Vehicle Speed and Type Classification*. Dissertation, M.I.T., Department of Electrical Engineering and Computer Science, 1976.
- [58] Y.-h. Pao: *Adaptive Pattern Recognition and Neural Networks*. Reading, Mass.: Addison-Wesley, 1989.
- [59] H. J. Payne: "Models of freeway traffic control," *Simulation Council Proc.*, Vol. 1, S. 51–61, 1971.
- [60] S. Peng und C. Harlow: "Systems for vehicle classification from range imagery," in *Proc. of the 28th Southeastern Symposium on System Theory*, Baton Rouge, LA, März 1996, S. 327–330.
- [61] E. Pfannerstill: *Ein Verfahren zur Merkmalextraktion aus Linienmustern zur Zuordnung von Fahrzeugkollektiven*. Dissertation, RWTH Aachen, 1983.
- [62] E. Pfannerstill: "Measuring section-related traffic data by correlation methods: A new approach to traffic monitoring and control," in *Proc. 2nd International Conference on Applications of Advanced Technologies in Transportation Engineering*, Minneapolis, MN, August 1991.
- [63] A. B. Poritz: "Hidden Markov models: A guided tour," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, New York, 1988, S. 7–12.

-
- [64] L. R. Rabiner und B. H. Juang: "An introduction to hidden Markov models," *IEEE ASSP Magazine*, Vol. 3, S. 4–16, Januar 1986.
- [65] L. R. Rabiner: "A tutorial on hidden Markov models and selected applications in speech recognition," *Proc. IEEE*, Vol. 77, S. 257–285, Februar 1989.
- [66] L. R. Rabiner und B. H. Juang: *Fundamentals of Speech Recognition*. Englewood Cliffs, NJ: Prentice Hall, 1993.
- [67] K. R. Rao und P. Yip: *Discrete Cosine Transform*. Boston: Academic Press, 1990.
- [68] P. Reichelt: *Die Steuerung von dynamischen Verkehrsbeeinflussungsanlagen mit Hilfe der Reisezeit*. Dissertation, RWTH Aachen, 1980.
- [69] M. D. Richard und R. P. Lippmann: "Neural network classifiers estimate Bayesian a posteriori probabilities," *Neural Computation*, Vol. 3, S. 461–483, April 1991.
- [70] H. Roe und S. Hobson: "Classification of road vehicles from microwave profiles," in *6th International Conference on Road Traffic Monitoring and Control*, London, April 1992, S. 128–129.
- [71] D. E. Rumelhart und J. R. McClelland: *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Vol. 1: Foundations*. Cambridge, Mass.: MIT Press, 1986.
- [72] J. Schürmann: *Polynomklassifikatoren für die Zeichenerkennung*. München: Oldenbourg, 1977.
- [73] H. Schütt und M. Cremer: "On-line-Identifizierung von Störungen und wesentlichen Parameteränderungen im Autobahnverkehr," Schriftenreihe der Arbeitsgruppe Automatisierungstechnik, Heft 4, TU Hamburg-Harburg, 1990.
- [74] S. Schrieber und M. Cremer: "A scheme for monitoring the traffic state on motorways using delayed data," in *Proc. 3rd World Congress on Intelligent Transport Systems*, Orlando, Florida, Oktober 1996.
- [75] D. L. Snyder und M. I. Miller: *Random Point Processes in Time and Space*. New York: Springer, 1991.

-
- [76] M. Sprotte: "Signalvorverarbeitung zur streckenbezogenen Verkehrsdatenerfassung," Diplomarbeit, TU Hamburg-Harburg, Arbeitsbereich Nachrichtentechnik, 1993.
- [77] C. W. Therrien: *Decision, Estimation, and Classification*. New York: Wiley, 1989.
- [78] J. T. Tou und R. C. Gonzalez: *Pattern Recognition Principles*. Reading, Mass.: Addison-Wesley, 1974.
- [79] H. L. VanTrees: *Detection, Estimation, and Linear Modulation Theory, Vol. 1*. New York: Wiley, 1968.
- [80] A. J. Viterbi: "Error bounds on convolutional codes and an asymptotically optimum decoding algorithm," *IEEE Trans. Information Theory*, Vol. IT-13, S. 260–269, April 1967.
- [81] H. Voß: "Mathematik für Ingenieure I," Vorlesungsskript, TU Hamburg-Harburg, 1984.
- [82] J. T. West: "California makes its move," *Traffic Engineering*, Vol. 41, No. 4, 1971.
- [83] J. Wohlers: "Verwendung streckenbezogener Verkehrsdaten bei der modellgestützten Schätzung kritischer Belastungsprofile auf Autobahnen," Diplomarbeit, TU Hamburg-Harburg, Arbeitsgruppe Automatisierungstechnik und Arbeitsbereich Nachrichtentechnik, 1991.
- [84] J. Wohlers, A. Mertins und N. Fliege: "Time-delay estimation for compound point-processes using hidden Markov models," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, Atlanta, GA, Mai 1996.
- [85] T. Y. Young und T. W. Calvert: *Classification, Estimation, and Pattern Recognition*. New York: American Elsevier, 1974.
- [86] X. Yuan, Y.-J. Lu und S. Sarraf: "Computer vision system for automatic vehicle classification," *ASCE*, Vol. 120, S. 861–876, November 1994.
- [87] U. Zölzer: *Digitale Audiosignalverarbeitung*. Stuttgart: Teubner, 1996.

Lebenslauf

Persönliche Daten:

Name: Jens Wohlers
Geburtsdatum: 8. Oktober 1963
Geburtsort: Rendsburg

Ausbildung und Beruf:

08/70 - 06/74 Grundschule Westerrönfeld
07/74 - 06/83 Gymnasium Herderschule Rendsburg
07/83 - 06/85 Wehrdienst
09/85 - 02/92 Studium der Elektrotechnik an der
Technischen Universität Hamburg-Harburg
seit 04/92 Wissenschaftlicher Mitarbeiter im
Arbeitsbereich Nachrichtentechnik der
Technischen Universität Hamburg-Harburg