Kapitel 4

Optische Fahrzeugalarmsysteme

Inhaltsangabe

4.1 Mot	tivation	81		
4.2 Pixelorientierte Verarbeitung 83				
4.2.1	${\rm Referenzbild-Verarbeitung}\ .\ .\ .\ .\ .\ .\ .\ .\ .\ .\ .\ .\ .\$	84		
4.2.2	Differenzbild-Verarbeitung			
4.2.3	Zeitrekursive-Verarbeitung			
	4.2.3.1 Einfache Zeitrekursion	89		
	4.2.3.2 Erweiterte Zeitrekursion	91		
4.2.4	Kalman-Filter	93		
4.2.5	Adaptive Schwellwertbestimmung $\ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots 1$	01		
	4.2.5.1 Methode nach Otsu	.03		
	4.2.5.2 Iterative Schwellwertbestimmung 1	.06		
4.3 Regionenorientierte Bildauswertung				
4.4 Zusammenfassung				

4.1 Motivation

Bei der Analyse von Bildfolgen ist die Lokalisierung von Bewegung eines der Hauptaufgaben. Bei der Bewegung von Objekten wird eine Unterscheidung zwischen starren Objekten (rigid) und Objekten mit veränderlicher Hülle (nonrigid) vorgenommen. Je nach der Verformung der äußeren Hülle des Objekts oder Teile des Objekts, wird eine Unterteilung der zweiten Form in

- artikulierte,
- elastische und
- flüssige

Bewegung vorgenommen [2]. Eine artikulierte Bewegung wird von einem Objekt ausgeführt, das sich aus starren Teilen zusammensetzt und das aufgrund der Bewegung dieser Teile zueinander eine nicht starre Bewegung ausführt. Die Bewegung eines Roboters zum Beispiel fällt unter die Definition artikulierter Bewegung. Bei der elastischen Bewegung geht man zwar von nicht starren Teilen des Objektes aus, setzt aber für die Bewegung dieser Teile einen gewissen Grad an Kontinuität voraus. Unter dieser Art versteht man z.B. die Bewegung von Kleidung. Bei der flüssigen Bewegung entfällt selbst die Kontinuitätsbedingung der elastischen Bewegung. In [34] gibt Huang einen umfassenden Überblick über die Definition der unterschiedlichen Bewegungsarten. Da bei der Bewegungsdetektion, wie sie in diesem Kapitel vorgestellt wird, keine *a priori* Information über die Art der Bewegung der Objekte vorliegt, wird die Annahme flüssiger Bewegung gemacht.

Die Ziele bei der Klassifizierung der Bewegungsarten sind:

- Rekonstruktion der 3D-Hülle
- Detektion/Erkennung von Objekten
- Verfolgung von Objekten
- Detektion/Erkennung von Gestik

Ziel dieses Kapitels sind Algorithmen, die zur Detektion/Erkennung von Objekten und die Verfolgung dieser, eingesetzt werden können. Hierbei werden die Algorithmen hauptsächlich in Bezug auf den optischen Innenraumschutz ausgesucht und entwickelt. Im folgenden soll unter der Bewegungsdetektion die Verfolgung dieser Ziele verstanden werden. Generell wird unter der Bewegungsdetektion die Trennung von bewegten Objekten, die den sog. Vordergrund darstellen, von einem unbewegten Hintergrund verstanden (Segmentierung). Daher kann die Bewegungsdetektion auch Vorder- bzw. Hintergrund Detektion genannt werden. Dieser Definition liegt zugrunde, daß in den Bildfolgen, die verarbeitet werden, die Kamera stationär montiert ist und somit jede Veränderung durch einer Bewegung eines Objektes (Vordergrund) verursacht wird.

Für die Anwendung des optischen Innenraumschutzes (OIS) werden an die Bildverarbeitung folgende Anforderungen gestellt:

- 1. Aus einer Folge von Bildern ist jede Bewegung eines Objektes in den Innenraum und um das Fahrzeug zu detektieren.
- 2. Der Vordergrund, der als bewegtes Objekt detektiert wurde, muß vom Hintergrund segmentiert werden.
- 3. Aus dem Vordergrund müssen einzelne Objekte separiert werden.
- 4. Der Umriß, die Größe, der Weg und die Richtung von jedem Objekt müssen je nach Anforderung bestimmt werden.
- 5. Objekte, die ihre Bewegung unterbrechen, müssen weiterhin detektiert werden.
- 6. Beleuchtungsänderungen durch Objekte dürfen nicht als Vordergrund detektiert werden.

Die Verfahren, die für diese Anforderungen in Frage kommen, können generell in regionenorientierte und pixelorientierte Verarbeitung unterteilt werden. Die pixelorientierten Verfahren nehmen eine Unterscheidung zwischen Vorder- und Hintergrund ausschließlich aus dem zeitlichen Grauwertverlauf eines Bildpunktes und den daraus ableitbaren Eigenschaften vor (siehe [19, 44, 82]). Sie eignen sich hervorragend als Vorverarbeitung für regionenorientierte Verfahren, die die Auswertung des Vordergrundes vornehmen. Durch diese Vorgehensweise ergibt sich eine wesentliche Einsparung an Rechenleistung, da die rechenaufwendigen Methoden nur auf einige wenige Regionen angewandt werden [92].

Die regionenorientierten Verfahren nutzen nicht nur den Grauwertverlauf eines Bildpunktes, sondern auch die der benachbarten Bildpunkte, um Veränderungen und somit die Bewegung von Objekten zu detektieren (siehe [62,105]). Diese Vorgehensweise folgt aus der grundsätzlichen Annahme, daß Objekte zusammenhängend sind, sich also aus mehreren Bildpunkten zusammensetzen müssen. Die Standardmethoden der regionenorientierten Verfahren sind u.a. Bereichswachstumsverfahren in [100] oder Pyramidenverknüpfungsalgorithmen in [42]. Diese Verfahren sind in der Regel aufgrund der Merkmalsauswahl und der daraus folgenden Merkmalsextraktion sehr rechenaufwendig und eignen sich nur bedingt als Vorverarbeitung für mehrere unterschiedliche Anwendungen gleichzeitig.

In der Literatur existieren unterschiedliche Ansätze zur Bewegungsdetektion. Bei Jähne in [42] werden die Segmentierungsverfahren – ähnlich wie in dieser Arbeit – in punkt- und regionenorientierte Verfahren unterteilt. Nagel hingegen führt in [58] und [59] die Änderungsdetektion entweder durch einen Vergleich zwischen Bildpunkten oder Bildauschnitten durch.

In Abschnitt 4.2 werden die Algorithmen untersucht, die hinsichtlich einer Bewegungsdetektion in Abschnitt 4.3 als Vorverarbeitung in Frage kommen. Abschnitt 4.2.5 untersucht Möglichkeiten die Koeffizienten adaptiv an die jeweilige Situation anzupassen. Hierzu werden einige aus der Literatur bekannte Verfahren an die Algorithmen aus dem folgenden Abschnitt angepaßt.



Abbildung 4.1: Diebstahl-Sequenz

4.2 Pixelorientierte Verarbeitung

Die in diesem Kapitel vorgestellten Methoden beruhen auf der Berechnung eines Hintergrundbildes und der Differenzbildung des aktuellen Bildes von diesem Hintergrundbild. Hierzu wird folgende Nomenklatur eingeführt:

- p(x, y) Pixel an der Koordinate (x, y), wird mit p abgekürzt
- i(p,t) Grauwert i des Pixels p zum Zeitpunkt t
- B(t) Hintergrundbild zum Zeitpunkt t
- I(t) Eingangsbild
- D(t) Differenzbild
- M(t) Maskenbild
- T Schwellwert zur Binarisierung

Das nach einer Schwellwertbildung erhaltene binäre Maskenbild M(p,t) kennzeichnet den Vordergrund mit einer 1 und den Hintergrund mit einer 0. Die Verfahren unterscheiden sich generell nur in der Berechnung des Hintergrundbildes. Für den optischen Innenraumschutz (OIS) bieten die pixelorientierten Verfahren, die Möglichkeit die Regionen zu detektieren, die verändert wurden. Durch die Auswertung der Maske kann somit der OIS einfach realisiert werden, obwohl erst die regionenorientierte Bildauswertung (Abschnitt 4.3) die Bestimmung des Zustandes durchführen kann.

In Abb. 4.1 sind einige Frames aus einer Sequenz von 50 Frames dargestellt. Diese Sequenz wurde aus der Position Nr. 6 (Tabelle 3.1) mit der hochdynamischen CMOS-Kamera (Abschnitt 3.5) mit einer Bildrate von ca. 10 Frames/s aufgenommen. Das Objektiv hat einen Gesamt-Feldwinkel von $2w = 150^{\circ}$ in der Diagonalen bei einem 1/3 Zoll Sensor. Die Auflösung der Frames ergibt sich aus der Auflösung des Sensors, der Grösse des



Abbildung 4.2: Fehlalarm-Sequenz

Innenraums bei der optischen Abbildung des Objektivs und beträgt in diesem Fall 220 × 160 Pixel. Die Sequenz stellt nach Abschnitt 2.2 eine typische Diebstahl-Sequenz dar, die zur Prüfung des Innenraumschutzes nach [94] geeignet ist. Auf dem Beifahrersitz befindet sich ein Objekt mit den Abmessungen 10 cm × 15 cm und einer Fläche \approx 100 Pixel, das von einer Person durch das *offene* Fenster aus dem Fahrzeug entwendet wird.

Abb. 4.2 zeigt eine Sequenz, die als Fehlalarm-Sequenz für die Algorithmen in dieser Arbeit verwendet wird und mit der gleichen Anordnung wie die Diebstahl-Sequenz aufgenommen wurde. Es zeigt eine Person, die sich auf der Beifahrerseite dem Fahrzeug nähert und interessiert in das Fahrzeug schaut. Um Frame 87 bewegt die Person den Kopf leicht in den Innenraum des Fahrzeugs. Mit dieser Sequenz soll geprüft werden, wie anfällig die Analyse in den einzelnen Verarbeitungs-Schritten ist. Nach den geltenden Richtlinien sollte ein Innenraumschutz – wenn überhaupt – erst um Frame 87 einen Alarm melden.

Die folgenden Unterabschnitte behandeln unterschiedliche Verfahren zur Bestimmung eines Hintergrundbildes. Nach einer kurzen Beschreibung der Referenz- und Differenz-Verfahren, die in der Literatur vollständig beschrieben sind, aber der Vollständigkeit halber erwähnt werden, werden rekursive Verfahren zur Schätzung des Hintergrundbildes vorgestellt. Den Abschluß bildet das Kalman-Filter, das als Prädiktionsfilter für das Hintergrundbild in dieser Arbeit das aufwendigste Verfahren darstellt.

4.2.1 Referenzbild-Verarbeitung

Die einfachste Methode zur Detektion des Vordergrundes ist die Referenzbild-Methode. Zur Berechnung wird zum Zeitpunkt t_0 ein Hintergrundbild B(t) gespeichert, von dem das aktuelle Eingangsbild I(t) abgezogen wird. Nach der Betrags- und Schwellwertbildung steht die binäre Maske M(t) zur Verfügung (Abb. 4.3). Die Berechnung der Maske ergibt sich somit folgendermaßen:

$$B(t) = I(t_0)$$
 für $t > t_0$ (4.1)

$$D(t) = B(t) - I(t) = I(t_0) - I(t)$$
(4.2)



Abbildung 4.3: Schema der Referenzbild-Verarbeitung



Abbildung 4.4: Ergebnis: Referenzbild-Verarbeitung Diebstahl-Sequenz

$$M(t) = \begin{cases} 1 & \text{für } |D(t)| > T \\ 0 & \text{sonst.} \end{cases}$$
(4.3)

Wie aus Gleichung (4.1) bis Gleichung (4.3) hervorgeht, hängt die Güte der Detektion entscheidend von dem Referenzbild und somit vom Zeitpunkt t_0 und der innerhalb dieser Spanne aufgetretenen bleibenden Veränderungen ab. Solche Veränderungen können Änderungen der Beleuchtung und/oder Objekte sein, deren Position verändert wurden. Nach einer solchen Änderung bleibt diese in der Maske solange erhalten, bis ein neues Referenzbild aufgenommen wird. Die Referenzbild-Methode eignet sich entweder nur in Umgebungen die keinen Veränderungen, außer den Objekten, die detektiert werden sollen, unterliegen oder in Umgebungen, in denen die Zeitpunkte für die Aufnahme eines Referenzbildes bekannt sind.

Für die Simulation der Sequenzen, deren Ergebnisse in den Abbildungen 4.4 und 4.5 dargestellt sind, wurde als Referenzbild jeweils das Frame 0 genommen. Es enthält das leere Fahrzeug ohne Personen in der Umgebung. Der Schwellwert T wurde statisch auf 80 ge-



Abbildung 4.5: Ergebnis: Referenzbild-Verarbeitung Fehlalarm-Sequenz

setzt, damit Änderungen in den Simulationsergebnissen nicht vom Schwellwert erzeugt werden. Zur Darstellung wurden die Werte M(t) = 1 in schwarz und M(t) = 0 in weiß gekennzeichnet. In Frame 3 ist schon zu erkennen, daß auf der Beifahrerseite außerhalb des Fahrzeugs eine Bewegung stattfindet. In Frame 15 hat die Person das Fahrzeug erreicht und greift mit der Hand durch das Fenster. Die Hand befindet sich genau im Fensterbereich. Durch den Schatten, den die Person in das Fahrzeug wirft, sinken die Grauwerte auf der Beifahrerseite. Während die Hand zum Objekt greift wird fast die gesamte Beifahrerseite als Vordergrund detektiert. In Frame 32 hat die Person das Objekt vom Beifahrersitz gehoben und zieht die Hand mit Objekt zurück. Das Objekt ist zwischen den Frames 27 bis 34 kaum zu erkennen. Wie die Abb. 4.4 eindeutig zeigt, segmentiert die Referenzbild-Verarbeitung Flächen, die sich nicht durch Bewegung verändert haben. Die Fehlalarm-Sequenz zeigt diesen Effekt noch deutlicher. Ab Frame 16 wird die gesamte Beifahrerseite als Vordergrund detektiert, obwohl sich die Person nur außerhalb des Fahrzeugs befindet.

Die Simulationsergebnisse zeigen, daß der Zeitpunkt zu dem das Referenzbild gewählt wird eine entscheidenden Einfluß hat. Da das Referenzbild während der Verarbeitung unverändert bleibt, ist diese Methode sehr empfindlich gegenüber zeitlichen Rauschen, das durch Rauschen des Bildsensors und durch Rauschen der Beleuchtung erzeugt werden kann. Solange das zeitliche Rauschen unter dem Schwellwert bleibt, werden die Ergebnisse dieser Methode nicht verfälscht. Es zeigt sich, daß einerseits die Bestimmung des Referenzbildes stark die Qualität der Detektionsergebnisse beeinflußt und andererseits die Detektion der Objekte in den Test-Sequenzen zu empfindlich ist. Aus diesen Gründen wird die Referenzbild-Verarbeitung nicht weiter verwendet.

4.2.2 Differenzbild-Verarbeitung

Mit der Differenzbild-Verarbeitung wird versucht, das in Abschnitt 4.2.1 genannte Problem, ein gutes Referenzbild aufzunehmen, zu lösen. Das Schema der Differenzbild-Verarbeitung ist in Abb. 4.6 abgebildet. Das Hintergrundbild wird ständig neu in zeitlichen Abständen von Δt aufgenommen und paßt sich somit den Änderungen, die in diesem



Abbildung 4.6: Schema der Differenzbild-Verarbeitung



Abbildung 4.7: Beispiel Differenzbild-Verarbeitung

Zeitraum aufgetreten sind, sofort an. Meistens wird für Δt ein Frame gewählt.

$$B(t) = I(t - \Delta t) \tag{4.4}$$

$$D(t) = B(t) - I(t) = I(t - \Delta t) - I(t)$$
(4.5)

Die Maske M(t) wird in Gleichung (4.6) anhand der Differenz mit Hilfe des Schwellwertes bestimmt.

$$M(t) = \begin{cases} 1 & \text{für } |D(t)| > T \\ 0 & \text{sonst} \end{cases}$$
(4.6)

Ein Beispiel einer Differenzbild-Verarbeitung ist in Abb. 4.7 dargestellt. Die Bilder wurden mit einer handelsüblichen CCD-Kamera aufgenommen. Zur Simulation wurde der Schwellwert auf T = 30 festgesetzt. Der Schwellwert mußte im Vergleich zu dem Schwellwert der Diebstahl- und Fehlalarm-Sequenz niedriger gewählt werden, da die CCD-Kamera Grauwerte mit 8 Bit statt mit 20 Bit darstellt. Zwischen den Frames 25 und 26 sind kaum Unterschiede festzustellen. Anhand des Differenzbildes werden die Unterschiede sichtbar. Wie an dem Beispiel deutlich wird, ist die Differenzbild-Methode besonders zur ortsgenauen Bestimmung von Veränderungen geeignet. Aus diesem Grund wird diese Methode oft



Abbildung 4.8: Ergebnis: Differenzbild-Verarbeitung Diebstahl-Sequenz

in der Vorverarbeitung in Datenkompressionsverfahren, wie z.B. beim MPEG-Verfahren (Motion Picture Expert Group) verwendet.

Trotz der hohen Ortsauflösung kann die Differenzbild-Verarbeitung nicht ohne weiteres zur Detektion von bewegten Objekten eingesetzt werden, wie aus der Maske in Abb. 4.7 klar hervorgeht. Durch die im Vergleich zur Bewegung hohe Bildrate ergeben sich drei Regionen [19], die auch Fronten genannt werden. Die vordere Front ist die Fläche, in der der Vordergrund sich über den Hintergrund bewegt hat. Die hintere Front ist die Fläche, die der Vordergrund wieder frei gibt. Die mittlere Region ist die Fläche, in der sich der Vordergrund über sich selbst bewegt. Aus diesem Grund sind die Differenzen in der mittleren Region so niedrig, daß sie unter dem Schwellwert liegen. Das Objekt wird nicht in einer geschlossenen Region segmentiert. Als ein weiterer Nachteil kann aufgeführt werden, daß die Objekte, die in ihrer Bewegung anhalten, innerhalb einer Iteration aus dem Vordergrund in den Hintergrund aufgenommen werden (siehe Gleichung (4.4)).

Die Simulationsergebnisse der Diebstahl-Sequenz in Abb. 4.8 zeigen eine hohe Unempfindlichkeit gegenüber Beleuchtungsänderungen. Der Schwellwert wurde – wie bei der Referenzbild-Methode – fest auf T = 80 gesetzt. Im Vergleich zu den Ergebnissen der Referenzbild-Verarbeitung wird der Schatten der Person nicht als Veränderung detektiert. Der Grund liegt darin, daß das Hintergrundbild ständig erneuert wird und sich somit an die Veränderungen, verursacht durch den Schatten, anpaßt. In den Frames 18 bis 34 enthalten die Regionen, die als Vordergrund erkannt wurden, kleine Löcher, die auch Blobs genannt werden. Die Löcher sind, wie in Abb. 4.7 gezeigt wurde, Regionen, in denen sich das Objekt über sich selbst bewegt. Durch die Blobs kann es bei langsamer Bewegung zu fehlerhaften Ergebnissen kommen. Bei der Fehlalarm-Sequenz in Abb. 4.9 zeigt sich die Differenzbild-Verarbeitung zwar unempfindlich gegenüber den Beleuchtungsänderungen, aber es kommt trotzdem zu Fehlern in Frame 83, da die Änderungen in diesem Bereich zu schnell sind. Weiterhin wird die Person außerhalb des Fahrzeugs nicht in einer kohärenten Region segmentiert.

Aus dem Grund, daß besonders langsame Objekte von der Differenzbild-Verarbeitung nicht in geschlossenen Flächen detektiert werden, wird die Differenzbild-Verarbeitung an



Abbildung 4.9: Ergebnis: Differenzbild-Verarbeitung Fehlalarm-Sequenz

dieser Stelle für den optischen Innenraumschutz nicht weiter eingesetzt. Um diesen Effekt entgegenzuwirken, könnte Δt größer als ein Frame gewählt werden, wobei dann Effekte der Referenz-Methode auftreten. Gegenüber zeitlichen Rauschen ist die Differenz-Methode unempfindlich, solange die Werte und unter den des Schwellwertes liegen.

4.2.3 Zeitrekursive-Verarbeitung

Faßt man die Nachteile der in Abschnitt 4.2.2 und Abschnitt 4.2.1 vorgestellten Methoden zusammen, ist für eine pixelorientierte Vordergrund-Detektion die Bestimmung des Hintergrundbildes das entscheidende Kriterium. Objekte, die sich bewegen oder ihre Bewegung für eine kurze Zeit anhalten, sollten als kohärent – auch über größere Zeiträume – segmentiert werden, wobei dauerhafte Veränderungen in das Hintergrundbild aufgenommen werden sollten. In den folgenden Unterabschnitten werden Methoden zur Berechnung eines *optimalen* Hintergrundbildes vorgestellt.

4.2.3.1 Einfache Zeitrekursion

Bei der einfachen Zeitrekursion (siehe Abb. 4.10) berechnet sich das Hintergrundbild in Abhängigkeit der Differenz und eines Adaptionskoeffizienten α . Die Berechnung des Hintergrundbildes läßt sich folgendermaßen beschreiben:

$$B(t) = (1 - \alpha) B(t - \Delta t) + \alpha I(t - \Delta t) \quad \text{mit } 0 \le \alpha \le 1.$$

$$(4.7)$$

Das Eingangsbild I(t) wird nun mit der Gewichtung α in das Hintergrundbild adaptiert, wobei das vorherige Hintergrundbild mit $1 - \alpha$ eingeht. Mit

$$D(t) = B(t - \Delta t) - I(t)$$
(4.8)

nach I(t) umgeformt und in Gleichung (4.7) eingesetzt, ergibt sich für die Bestimmung des aktuellen Hintergrundbildes:

$$B(t) = B(t - \Delta t) - \alpha D(t)$$
(4.9)



Abbildung 4.10: Schema der einfachen Zeitrekursion

Betrachtet man Gleichung (4.7) genauer, so sieht man, daß für $\alpha = 1 B(t) = I(t)$ wird und Gleichung (4.8) in Gleichung (4.5) übergeht. Ist $\alpha = 0$ übernimmt die einfache Zeitrekursion keine Veränderungen aus dem Eingangsbild und das Hintergrundbild bleibt unverändert. Die Gleichung (4.7) verhält sich in diesem Fall wie die Referenzbild-Verarbeitung, wobei das Referenzbild nicht definiert ist. Die zeitrekursive Verarbeitung stellt somit ein Bindeglied zwischen der Referenzbild-Verarbeitung und der Differenzbild-Verarbeitung dar. Durch die Anpassungsfähigkeit, die durch α eingestellt wird, ist der Algorithmus in der Lage, sich an Änderungen innerhalb einer Sequenz anzupassen. Karmann *et al* in [44] haben gezeigt, daß die einfache Zeitrekursion zur Segmentierung von Fahrzeugen im Verkehr geeignet ist.

Für die Simulationen wurde der Schwellwert statisch auf T = 80, der Adaptionskoeffizient $\alpha = 0,8$ und Δt auf ein Frame gesetzt. Bei unterschiedlichen Sequenzen haben sich diese Werte bewährt. Die Ergebnisse der einfachen Zeitrekursion, die in Abb. 4.11 dargestellt sind, erzielen bei der Diebstahl-Sequenz ähnliche Ergebnisse, wie die Differenzbild-Methode. Unterschiede werden nur bei Bewegung von Hintergrund-Objekten deutlich, wie in Frame 34. Beim Entfernen des Objekts wird eine Region, wie ein Schweif, nachgezogen, weil es an diesen Stellen zu Anpassungen im Hintergrundbild kommt. Diese Eigenschaft gibt der einfachen Zeitrekursion einen leichten Vorteil gegenüber der Differenzbild-Verarbeitung, da Objekte, die im Hintergrund-Bild enthalten sind und entfernt werden über mehrere Frames als Veränderung angezeigt werden. Die Simulationen der Fehlalarm-Sequenz in Abb. 4.12 zeigen ab dem Frame 20 Regionen im Innenraum als verändert an. Ursache ist die im Vergleich zur Bewegung des Schattens langsame Anpassung des Hintergrundbildes, ähnlich wie bei der Referenzbild-Verarbeitung. Die Flächen, die falsch detektiert werden, sind aber kleiner als bei der Referenzbild-Methode (siehe Abb. 4.5).

Auch wenn durch den Adaptionskoeffizienten eine Anpassung an die Veränderungsgeschwindigkeit gegeben ist, kann nicht zwischen einer Veränderung, erzeugt durch ein



Abbildung 4.11: Ergebnis: Einfache Zeitrekursion Diebstahl-Sequenz



Abbildung 4.12: Ergebnis: Einfache Zeitrekursion Fehlalarm-Sequenz

Vordergrundobjekt, und einer Veränderung, die durch eine Bewegung im Hintergrund hervorgerufen wurde, unterschieden werden. Beide Ereignisse werden mit dem gleichen Koeffizienten in das Hintergrundbild aufgenommen. Durch die fehlende Diskriminierung kann die Einstellung des Adaptionskoeffizienten nur suboptimal erfolgen.

4.2.3.2 Erweiterte Zeitrekursion

Die erweiterte Rekursion unterscheidet sich von der einfachen Rekursion durch die Einführung eines zweiten Adaptionskoeffizienten (Abb. 4.13), der die Adaption des Vordergrundes bestimmt. Bei der erweiterten Rekursion wird als erstes das Differenzbild bestimmt:

$$D(t) = B(t - \Delta t) - I(t)$$
(4.10)

Von dem Ergebnis des Schwellwertbilders hängt die Entscheidung ab, mit welchem Satz von Koeffizienten die Schätzung des Hintergrundbildes erfolgen soll. Die erweiterte Rekursion für die Berechnung des Hintergrundbildes ist somit vom Maskenbild M(t) abhängig.



Abbildung 4.13: Schema: Erweiterte Zeitrekursion

Dies wurde in Abb. 4.13 durch die gestrichelte Linie angedeutet, die den Signalpfad am Wechselschalter bestimmt. Die Gleichung zur Berechnung des Hintergrundbildes ist:

$$B(t) = \begin{cases} (1-\alpha)B(t-\Delta t) + \alpha I(t-\Delta t) & \text{für } |D(t-\Delta t)| \le T\\ (1-\beta)B(t-\Delta t) + \beta I(t-\Delta t) & \text{für } |D(t-\Delta t)| > T. \end{cases}$$
(4.11)

Je nachdem, ob es sich um einen Hintergrund- oder Vordergrund-Pixel handelt, ändert sich die Berechnung des Hintergrundbildes für die nächste Rekursion. Mit $\beta \ll \alpha$ werden Regionen, die als Hintergrund erkannt werden, schnell mit dem Koeffizienten α in das Hintergrundbild adaptiert und Flächen, die als Vordergrund erkannt wurden, nur sehr langsam mit dem Koeffizienten β in das Hintergrundbild aufgenommen. Dadurch ist einerseits gewährleistet, daß leichte Beleuchtungsänderungen nicht als Bewegungen detektiert werden, und andererseits, daß langfristige Änderungen nach einer gewissen Anzahl an Rekursionen im Hintergrund sind [66]. Auch in dem Fall wenn ein Objekt (Vordergrund) über einen Zeitraum sich in der Szene aufhält, wird gewährleistet, daß an dieser Stelle die Adaption mit dem wesentlich kleineren Wert β erfolgt und nicht wie bei der einfachen Rekursion mit dem wesentlich größeren Koeffizienten α . Verläßt nun das Objekt die Szene, so werden die Flächen, die von dem Objekt belegt wurden, richtig als Hintergrund erkannt.

Für die Simulationen wurden der Schwellwert T = 80, die Adaptionskoeffizienten $\alpha = 0, 8$ und $\beta = 0,001$ und Δt auf ein Frame gesetzt. Die Simulationen der Test-Sequenzen in



Abbildung 4.14: Ergebnis: Erweiterte Zeitrekursion Diebstahl-Sequenz

den Abbildungen 4.14 und 4.15 zeigen bei der erweiterten Zeitrekursion die besten Ergebnisse, auch wenn in manchen Situation Kompromisse geschlossen werden müssen. Die Person in der Diebstahl-Sequenz in Abb. 4.14 wird exakt und kohärent als Vordergrund angezeigt. Die Stelle an der das entwendete Objekt lag, wird noch einige Frames nach dem Entwenden als veränderte Stelle gekennzeichnet. Die erweiterte Zeitrekursion zeigt an einigen Stellen stärkeres Rauschen, das sich durch kleine *Blobs* sichtbar macht. Diese sind, wie in der regionenorienten Bildauswertung noch gezeigt wird, für die Auswertung der Bilder nicht störend.

In der Fehlalarm-Sequenz in Abb. 4.15 werden durch die schnelle Bewegung des Schattens – wie bei den anderen Verarbeitungen vorher – ebenfalls Bereiche falsch segmentiert. Hier wirkt sich der zweite Adaptionskoeffizient aus. Wird ein Pixel in der Maske als Vordergrund gekennzeichnet, verlangsamt sich die Geschwindigkeit der Anpassung erheblich. Dies ist an den Stellen erwünscht, an denen tatsächlich ein Objekt ist, aber es führt in solchen Extremsitutationen zu fehlerhaften Ergebnissen.

Auch wenn die erweiterte Zeitrekursion in bestimmten Situationen noch Mängel aufweist, ist sie, wie der Abschnitt 4.2.4 noch zeigen wird, die effektivste Methode pixelintern eine Veränderung zu detektieren.

4.2.4 Kalman-Filter

Die Verfahren der zeitrekursiven Verarbeitung haben versucht, die Aufgabe ein optimales Hintergrundbild mit festen Adaptionskoeffizienten zu lösen. Geht man einen Schritt weiter kann die Berechnung eines optimalen Hintergrundbildes durch einen Filter erfolgen. Das Schema eines solchen Systems, in Abb. 4.16 dargestellt, unterscheidet sich von den vorangegangenen dadurch, daß für die Bestimmung des Hintergrundbildes ein Filter, in diesem Fall das Kalman-Filter, eingesetzt wird.

In diesem Abschnitt wird die Aufgabe betrachtet ein optimales Hintergrundbild das mit Hilfe eines Zustandsraummodells berechnet wird, welches die Entstehung der Klassen Hintergrund und Vordergrund modelliert, zu lösen. Hierzu werden Methoden zur Schätzung



Abbildung 4.15: Ergebnis: Erweiterte Zeitrekursion Fehlalarm-Sequenz



Abbildung 4.16: Prädiktion des Hintergrundbildes mit einem Kalman-Filter



Abbildung 4.17: Lineares System abgeleitet aus Gleichung (4.12)

und Filterung aus der Regelungstechnik entlichen. In der Regelungstechnik wird das Verhalten dynamischer Systeme im Zustandsraum beschrieben, wobei sichergestellt sein muß, daß das System beobachtbar ist (siehe [11]). Das Interesse an dieser Art der Beobachtung und Schätzung stieg Ende der 50er Jahre nicht nur durch die aufkommende Raumfahrt, sondern auch durch die Entwicklung leistungsstarker Digitalrechner, mit der Möglichkeit eine rekursive Schätzung in Echtzeit auszuführen, an. Zu dieser Zeit vollzog sich ein Wandel in der Betrachtung von Filtern, die zu Zeiten von Wiener, Bode und Shannon mit Hilfe des Frequenzganges und der Übertragungsfunktion charakterisiert wurden, zum Entwurf von Algorithmen zur Echtzeitbestimmung von Gaußschen, Gauß-Markoffschen bzw. Minimum-Varianz-Schätzwerten, von bedingten Erwartungswerten oder sogar ganzer bedingter Verteilungsdichten.

Die um 1940 von Kolmogoroff [49] und Wiener [102] unabhängig voneinander entstandenen Arbeiten, werden allgemein als die Basis der modernen Filtertheorie angesehen. Sie betrachteten stationäre Prozesse, unendlich lange Beobachtungsintervalle und zeitinvariante Filter. Als Gütekriterium dienten die mittleren Fehlerquadrate. Daraus wurde die Gewichtungsfunktion (Wiener-Hopfsche Integralgleichung) des Filters abgeleitet, welches das Gütekriterium optimierte. Die einschränkenden Bedingungen der Wiener-Kolmogoroffschen Theorie wurden von Booton [8] auf instationäre Prozesse und zeitvariable Filter gelockert. Booton war aber nicht in der Lage weder eine praktikable Methode zur Lösung der Integralgleichung noch zur Realisierung zeitvarianter Gewichtsfunktionen zur Verfügung zu stellen.

Zur Herleitung der Kalman-Gleichungen wird ein lineares System in diskreter Zeit für einen Bildpunkt betrachtet, d.h. das für jeden Bildpunkt in jedem Frame ein eigenes Kalman-Filter berechnet werden muß.

$$\vec{x}(k+1) = \mathbf{A}(k)\vec{x}(k) + \vec{v}(k) \qquad k \ge k_0$$

$$\vec{y}(k) = \mathbf{C}(k)\vec{x}(k) + \vec{w}(k) \qquad (4.12)$$

Nach dem mathematischen Modell aus Gleichung (4.12) kann Abb. 4.17 abgeleitet wer-

den, aus dem zu erkennen ist, daß der Zustandsvektor $\vec{x}(k)$ nicht zugänglich ist. Die Systemparameter $\mathbf{A}(k)$ und $\mathbf{C}(k)$ sind bekannte Matrizen, die das Verhalten des Systems beschreiben. Der Anfangswert $\vec{x}(k_0)$ wird als Zufallsvektor mit dem Erwartungswert Null und der Kovarianzmatrix $\mathbf{P}(k_0)$ angenommen.

Bezieht man das mathematische Modell auf die Aufgabe der Bestimmung eines optimalen Hintergrundbildes, so beschreibt das Modell die Beziehung zwischen dem realen Hintergrund, hier mit $\vec{x}(k)^1$ bezeichnet, dem Eingangswert $\vec{v}(k)$ des Modells und dem meßbaren Ausgangswert $\vec{y}(k)$. Der meßbare Ausgangswert $\vec{y}(k)$ entspricht hier einem Punkt des Eingangsbildes I(t), das von der Kamera aufgenommen wird. Der Eingangswert $\vec{v}(k)$, der im folgenden als Störvektor bezeichnet wird, kann in dieser Betrachtungsweise als Vordergrund gesehen werden. $\vec{w}(k)$ ist der durch den Prozeß der Aufnahme eingeführte Meßfehler.

Die Kovarianzmatrix ist allgemein wie folgt definiert:

$$E\left\{\left[\vec{x}(k_0) - E\left\{\vec{x}(k_0)\right\}\right]\left[\vec{x}(k_0) - E\left\{\vec{x}(k_0)\right\}\right]^T\right\} = \mathbf{P}(k_0),$$
(4.13)

wobei $E \{\vec{x}(k)\}$ der Erwartungswert von $\vec{x}(k)$ ist. Die Vektoren $\vec{v}(k)$ und $\vec{w}(k)$ sind vektorielle weiße Zufallsprozesse mit symmetrischen, positiv semidefiniten Kovarianzmatrizen.

$$E \{\vec{v}(k)\} \equiv \vec{0} , E \{\vec{v}(k)\vec{v}^{T}(n)\} = \mathbf{V}(k)\delta_{kn}$$

$$E \{\vec{w}(k)\} \equiv \vec{0} , E \{\vec{w}(k)\vec{w}^{T}(n)\} = \mathbf{W}(k)\delta_{kn}$$

$$\delta_{kn} = \begin{cases} 1 & \text{für } k = n \\ 0 & \text{für } k \neq n \end{cases}$$

$$(4.14)$$

Weiterhin wird vorausgesetzt, daß der Anfangszustand, der Stör- und der Meßprozesse gegenseitig unkorreliert sind.

$$E \{\vec{x}(k_0)\vec{v}^{T}(k)\} \equiv \mathbf{0}$$

$$E \{\vec{x}(k_0)\vec{w}^{T}(k)\} \equiv \mathbf{0}$$

$$E \{\vec{w}(k)\vec{v}^{T}(n)\} \equiv \mathbf{0}$$
(4.15)

Gesucht wird ein linearer, erwartungstreuer Schätzwert $\hat{\vec{x}}(k)$ (geschätzter Hintergrundwert) für $\vec{x}(k)$, der basierend auf dem Meßvektor $\vec{y}(k)$ berechnet werden soll. Der Schätzfehler $\tilde{\vec{x}}(k)$ ist definiert als die Differenz zwischen dem Schätzwert und dem tatsächlichen Wert und hat - aufgrund der Bedingung der Erwartungstreue - einen verschwindenden Erwartungswert. Die Kovarianzmatrix $\tilde{\mathbf{P}}(k)$ des Schätzfehlers ist somit:

$$\begin{aligned}
\tilde{\vec{x}}(k) &:= \vec{x}(k) - \vec{\hat{x}}(k) \\
\tilde{\mathbf{P}}(k) &:= E\left\{\tilde{\vec{x}}(k)\tilde{\vec{x}}^{T}(k)\right\}.
\end{aligned}$$
(4.16)

Zur Herleitung wird angenommen, daß der erste Schätzwert $\hat{\vec{x}}(k)$ schon optimal im Sinne der minimalen Varianz berechnet ist (siehe Anhang A). Mit der Gleichung A.14 aus Anhang A kann direkt gefolgert werden, daß

$$E\left\{\left[\vec{x}(k) - \hat{\vec{x}}(k)\right]\left[\vec{y}^{T}(k_{0}), \vec{y}^{T}(k_{0}+1), \cdots, \vec{y}^{T}(k)\right]\right\} = \mathbf{0}$$
(4.17)

 $^{^1 \}rm{Die}$ Bezeichnung wurde absichtlich so gewählt, um zu betonen, daß es sich um den Wert eines Bildpunktes handelt.

ist. Es wird nun der extrapolierte Wert $\vec{x}^*(k+1)$ gesucht, der die Bedingung der minimalen Varianz, unter der Kenntnis der bis zum Zeitpunkt k bekannten Meßwerte $(\vec{y}(k_0), \vec{y}, \cdots, \vec{y}(k))$, erfüllt. Für den extrapolierten Wert muß aufgrund derselben Voraussetzungen ebenfalls Gleichung (A.14) gelten.

$$E\left\{\left[\vec{x}(k+1) - \vec{x}^{*}(k+1)\right]\left[\vec{y}^{T}(k_{0}), \vec{y}^{T}(k_{0}+1), \cdots, \vec{y}^{T}(k)\right]\right\} = \mathbf{0}$$
(4.18)

Sobald der Meßwert $\vec{y}(k+1)$ verfügbar ist, kann der Schätzwert $\hat{\vec{x}}(k+1)$ in Gleichung (A.14) eingesetzt werden.

$$E\left\{\left[\vec{x}(k+1) - \hat{\vec{x}}(k+1)\right]\left[\vec{y}^{T}(k_{0}), \vec{y}^{T}(k_{0}+1), \cdots, \vec{y}^{T}(k), \vec{y}^{T}(k+1)\right]\right\} = \mathbf{0}$$
(4.19)

Ein Vergleich der Gleichung (4.17) mit Gleichung (4.19) legt die Vermutung nahe, $\vec{x}(k+1)$ durch $\mathbf{A}(k)\vec{x}(k) + \vec{v}(k)$ zu ersetzen (siehe Gleichung (4.12)). Da $\vec{v}(k)$ mit $\vec{w}(k)$ und $\vec{x}(k)$ unkorreliert ist und somit auch mit $\vec{y}(k)$, gestaltet sich Gleichung (4.18) zu:

$$E\left\{\left[\mathbf{A}(k)\vec{x}(k) - \vec{x}^{*}(k+1)\right]\left[\vec{y}^{T}(k_{0}), \vec{y}^{T}(k_{0}+1), \cdots, \vec{y}^{T}(k)\right]\right\} = \mathbf{0}$$
(4.20)

und kann durch $\vec{x}^*(k+1) = \mathbf{A}(k)\vec{x}^*(k)$ erfüllt werden, da nach Herausziehen der Matrix $\mathbf{A}(k)$ aus dem Erwartungswert die gültige Bedingung aus 4.17 vorliegt. Der Fehler der Extrapolation ist mit den bis hier hergeleiteten Gleichungen:

$$\vec{x}(k+1) - \vec{x}^{*}(k+1) = \mathbf{A}(k)\vec{x}(k) + \vec{v}(k) - \mathbf{A}(k)\vec{x}(k) = \mathbf{A}(k)\tilde{\vec{x}}(k) + \vec{v}(k)$$
(4.21)

und hat die Kovarianzmatrix:

$$\mathbf{P}^{*}(k+1) = E\left\{ \left[\vec{x}(k+1) - \vec{x}^{*}(k+1) \right] \left[\vec{x}(k+1) - \vec{x}^{*}(k+1) \right]^{T} \right\} \\ = E\left\{ \left[\mathbf{A}(k)\tilde{\vec{x}}(k) + \vec{v}(k) \right] \left[\tilde{\vec{x}}^{T}(k)\mathbf{A}^{T}(k) + \vec{v}^{T}(k) \right] \right\} \\ = \mathbf{A}(k)\tilde{\mathbf{P}}(k)\mathbf{A}^{T}(k) + \mathbf{V}(k)$$
(4.22)

In Gleichung (4.22) wurde die Eigenschaft, daß die Störgröße $\vec{v}(k)$ und der Schätzfehler $\tilde{\vec{x}}(k)$ unkorreliert sind, ausgenutzt. Somit ist die Extrapolation abgeschlossen und als nächstes muß eine Lösung für Gleichung (4.19) gefunden werden. Hierzu wird der folgende Ansatz verwendet (siehe [11]):

$$\hat{\vec{x}}(k+1) = \vec{x}^*(k+1) + \mathbf{K}(k+1) \left[\vec{y}(k+1) - \mathbf{C}(k+1)\vec{x}^*(k+1) \right]$$
(4.23)

 $\mathbf{K}(k)$ ist die sog. Kalman-Verstärkungsmatrix. Ersetzt man $\vec{y}(k+1)$ durch den entsprechenden Teil aus Gleichung (4.12), ergibt sich der folgende Schätzfehler.

$$\vec{x}(k+1) - \hat{\vec{x}}(k+1) = \begin{bmatrix} \mathbf{I} - \mathbf{K}(k+1)\mathbf{C}(k+1) \\ [\vec{x}(k+1) - \vec{x}^*(k+1)] - \mathbf{K}(k+1)\vec{w}(k+1) \end{bmatrix}$$
(4.24)

Der Ausdruck in 4.24 kann nun in die Bedingung 4.19 eingesetzt werden. Aus dem Grund, daß der eingesetzte Teil mit den alten Meßwerten unkorreliert ist, bleiben nur die zwei Terme

$$E\left\{ \left[\mathbf{I} - \mathbf{K}(k+1)\mathbf{C}(k+1) \right] \left[\vec{x}(k+1) - \vec{x}^*(k+1) \right] \vec{y}^T(k+1) \right\} - E\left\{ \mathbf{K}(k+1)\vec{w}(k+1)\vec{y}^T(k+1) \right\} = \mathbf{0}$$
(4.25)

über. Durch Ersetzen von $\vec{y}^{T}(k+1)$ mit der Transponierten von Gleichung (4.12) und der Tatsache, daß $\vec{w}(k+1)$ unkorreliert zu $\vec{x}(k+1)$ und $\vec{x}^{*}(k+1)$ ist, kann die Gleichung (4.25) vereinfacht werden.

$$[\mathbf{I} - \mathbf{K}(k+1)\mathbf{C}(k+1)] \underbrace{[\vec{x}(k+1) - \vec{x}^{*}(k+1)] \vec{x}^{T}(k+1)]}_{\mathbf{K}(k+1)\mathbf{C}(k+1)} \mathbf{C}^{T}(k+1) - \mathbf{K}(k+1)\vec{w}(k+1) \mathbf{K}(k+1) - \mathbf{K}(k+1)\vec{w}(k+1) = \mathbf{0}$$

$$[\mathbf{I} - \mathbf{K}(k+1)\mathbf{C}(k+1)] \mathbf{P}^{*}(k+1)\mathbf{C}^{T}(k+1) - \mathbf{K}(k+1)\mathbf{W}(k+1) = \mathbf{0}$$
(4.26)

In Gleichung (4.26) wurden die Gleichungen A.15 und 4.14 berücksichtigt. Dadurch steht nach einer Umformung die Vorschrift zur Berechnung der Kalman-Verstärkungsmatrix fest.

$$\mathbf{K}(k+1) = \mathbf{P}^{*}(k+1)\mathbf{C}^{T}(k+1)\left[\mathbf{C}(k+1)\mathbf{P}^{*}(k+1)\mathbf{C}^{T}(k+1) + \mathbf{W}(k+1)\right]^{-1} \quad (4.27)$$

Zur Bestimmung der Kovarianzmatrix des Schätzfehlers für den Zeitpunkt k + 1 wird Gleichung (4.23) mit $\vec{x}^{T}(k+1)$ von rechts erweitert und die Erwartung gebildet. Unter Berücksichtigung der Gleichung (A.15) – wie in Gleichung (4.26) – ergibt sich:

$$E\left\{\left[\vec{x}(k+1) - \hat{\vec{x}}(k+1)\right]\vec{x}^{T}(k+1)\right\} = \\E\left\{\tilde{\vec{x}}(k+1)\tilde{\vec{x}}^{T}(k+1)\right\} = \tilde{\mathbf{P}}(k+1)$$

$$\tilde{\mathbf{P}}(k+1) = E\left\{ \left[\mathbf{I} - \mathbf{K}(k+1)\mathbf{C}(k+1) \right] \left[\vec{x}(k+1) - \vec{x}^*(k+1) \right] \vec{x}^{T}(k+1) \right\} \\ = \left[\mathbf{I} - \mathbf{K}(k+1)\mathbf{C}(k+1) \right] \mathbf{P}^*(k+1).$$
(4.28)

Somit sind die Gleichungen des Kalman-Filters hergeleitet und können in dem folgenden Satz zusammengefasst werden.

Satz 4.2.1

Der lineare, erwartungstreue Schätzwert minimaler Varianz für den Zustand $\vec{x}(k)$ des zeitlich diskreten, stochastisch gestörten Systems

$$\vec{x}(k+1) = \mathbf{A}(k)\vec{x}(k) + \vec{v}(k)$$
 (4.29)

$$\vec{y}(k) = \mathbf{C}(k)\vec{x}(k) + \vec{w}(k) \tag{4.30}$$

bei dem die Meßwerte $\vec{y}(k)$ von k_0 bis k bekannt sind und die Gleichungen 4.13 und 4.14 gelten, ist durch den folgenden rekursiven Algorithmus gegeben:

1) Filtergleichungen

$$\vec{x}^{*}(k+1) = \mathbf{A}(k)\vec{x}^{*}(k) \tag{4.31}$$

$$\hat{\vec{x}}(k) = \vec{x}^*(k) + \mathbf{K}(k) [\vec{y}(k) - \mathbf{C}(k)\vec{x}^*(k)]$$
(4.32)

2) Kovarianzmatrizen und Verstärkungsmatrix

$$\mathbf{P}^*(k+1) = \mathbf{A}(k)\tilde{\mathbf{P}}(k)\mathbf{A}^T(k) + \mathbf{V}(k)$$
(4.33)

$$\mathbf{K}(k) = \mathbf{P}^{*}(k)\mathbf{C}^{T}(k)\left[\mathbf{C}(k)\mathbf{P}^{*}(k)\mathbf{C}^{T}(k) + \mathbf{W}(k)\right]$$
(4.34)

$$\tilde{\mathbf{P}}(k) = [\mathbf{I} - \mathbf{K}(k)\mathbf{C}(k)]\mathbf{P}^{*}(k)$$
(4.35)



Abbildung 4.18: Ergebnis: Kalman-Filter Diebstahl-Sequenz

3) Anfangsbedingung

$$\mathbf{P}^*(k_0) = \mathbf{P}(k_0) \tag{4.36}$$

Um die Komplexität der Simulationen niedrig zu halten, wurde die Systemmatrix $\mathbf{A}(k)$ konstant auf zwei Dimensionen beschränkt. Die Meßmatrix $\mathbf{C}(k)$ wurde ebenfalls konstant gehalten. Die Verstärkungsmatrix wurde in Abhängigkeit der binären Maske bestimmt. Um den Zusammenhang mit den Verfahren aus den anderen Abschnitten deutlich zu machen, wurden die Bezeichnungen für Hintergrund-, Eingangsbild, Maske und extrapolierten Hintergrundbild eingeführt. Hierbei ist zu beachten, daß das Hintergrundbild für jeden Bildpunkt zweidimensional ist.

$$\mathbf{B}^{*}(p,k+1) = \begin{bmatrix} 1 & a_{1,2} \\ 0 & a_{2,2} \end{bmatrix} \mathbf{B}^{*}(p,k)$$

$$\mathbf{K}(k) = \alpha M(p,k) + \beta (1 - M(p,k))$$

$$\mathbf{B}(p,k) = \mathbf{B}^{*}(p,k) + \mathbf{K}(k) \begin{bmatrix} I(p,k) - \begin{bmatrix} 1 & 0 \end{bmatrix} \mathbf{B}^{*}(p,k) \end{bmatrix}$$

$$(4.37)$$

Die Simulationen des Kalman-Filters mit den Koeffizienten $a_{1,2} = a_{2,2} = 0, 7, \alpha = 0,001$, $\beta = 0,01$ und T = 80, die in den Abbildungen 4.18 und 4.19 abgebildet sind, zeigen, daß das Kalman-Filter sehr empfindlich auf Veränderungen reagiert. Auch wenn die Koeffizienten in der Filtergleichung angepaßt werden, bleiben die Ergebnisse der Simulation nicht akzeptabel. Wird das Filter stark unempfindlich eingestellt, sind die Ergebnisse vergleichbar mit denen der Referenzbild-Methode. Nicht nur die Ergebnisse der Simulation, sondern auch der hohe Rechenaufwand, lassen das Kalman-Filter für die Vorverarbeitung für ein Fahrzeugalarmsystem ungeeignet scheinen. Da der Rechenaufwand mit den Systemkosten und dem Energieverbrauch verbunden ist, würden selbst bessere Ergebnisse als die der zeitrekursiven Methoden einen Einsatz im Fahrzeug für diese Anwendungszwecke nicht rechtfertigen.

In der Tabelle 4.1 wurden für die pixelorientierten Verfahren die Anzahl der Integer und Float Operation je Pixel und der Speicherbedarf berechnet. Das Kalman-Filter benötigt



Abbildung 4.19: Ergebnis: Kalman-Filter Fehlalarm-Sequenz

Verarbeitung	Integer Operation	Float Operation	${ m Speicherbedarf}$
Referenzbild-Verarbeitung	3	0	2 Integer
${ m Differenzbild}$ -Verarbeitung	3	0	2 Integer
Einfache Zeitrekursion	5	2	2 Integer
Erweiterte Zeitrekursion	5	2	2 Integer
Kalman-Filter	4	6	1 Integer, 2 Float

Tabelle 4.1: Rechenaufwand und Speicherbedarf pro Pixel der verschiedenen Methoden



Abbildung 4.20: Hintergrund-, Eingangs-, Differenzbild, Histogramme des Eingangs- und Differenzbildes

unter allen Methoden die meisten Ressourcen, obwohl die Ergebnisse nicht überzeugen können. Vergleicht man den Aufwand mit den Ergebnissen, bietet sich die erweiterte Zeitrekursion für die weitere Verarbeitung an.

4.2.5 Adaptive Schwellwertbestimmung

Die Ergebnisse der vorgestellten Methoden sind, wie die Simulationen gezeigt haben, stark von den Schwellwerten abhängig. Bei konstantem Schwellwert können bei unterschiedlichen Eingangs-Sequenzen die Ergebnisse stark variieren. Daher stellt sich die Frage, ob die Schwellwerte der Algorithmen nicht adaptiv eingestellt werden können. In [1] wurden die in der Literatur [73] bekannten Verfahren zur adaptiven Schwellwertberechnung untersucht. Hierzu wurden die bekannten Verfahren an die zeitrekursiven Methoden angepaßt und getestet.

Die Schwellwert-Operation dient zur Binarisierung eines Bildes, d.h. also zur Trennung zweier Regionen, die im besten Fall von zwei getrennten Objekten, wie z.B. Vorder- und Hintergrund, stammen. Die binäre Maske enthält somit Informationen über die Objekte, wie z.B. Größe, Form, Position, Kompaktheit und Schwerpunkt.

Bei den Verfahren der pixelorientierten Vorverarbeitung wird die binäre Maske durch eine Schwellwert-Operation des Differenzbildes erzeugt. Die Qualität dieser Operation hängt einerseits von der Schätzung des Hindergrundbildes ab und andererseits entscheidend vom Schwellwert. Durch eine falsch gewählte Schwelle gehen die Informationen über die lokalisierten Objekte verloren. Diese Fehler wirken sich bei der zeitrekursiven Verarbeitung auf alle folgenden Masken aus.



Abbildung 4.21: Test-Sequenz für adaptive Schwellwert-Verfahren

Die Verfahren zur adaptiven Bestimmung eines Schwellwertes werden in lokale und globale Verfahren unterteilt. Bei den lokalen Verfahren wird eine Schwellwert-Matrix bestimmt, die in jedem Bildpunkt vom Grauwert i(p, t) und von den lokalen Eigenschaften des Bildpunktes p(x, y) abhängig ist. Globale Verfahren bestimmen den Schwellwert ausschließlich aus den Grauwerten i(p, t). Generell sind die lokalen Verfahren rechenaufwendiger als die globalen. Da die vorgestellten Algorithmen in den Abschnitten 4.2.1 bis 4.2.4 nur globale Schwellwerte benutzen, wurde die Untersuchung der adaptiven Schwellwertbestimmung auf die globalen Verfahren beschränkt. Wie Simulationsergebnisse zeigen werden, erzielen schon die globalen Verfahren sehr gute Ergebnisse.

Die meisten bekannten Methoden zur Berechnung eines globalen Schwellwertes analysieren das Histogramm des von der Kamera gelieferten unbearbeiteten Grauwertbildes. Bei der zeitrekursiven Verarbeitung wird aber ein Schwellwert für das Differenzbild benötigt, das sich vom Histogramm des Eingangsbildes grundsätzlich unterscheidet.

In Abb. 4.20 sind die Histogramme des Eingangsbildes und des Differenzbildes, das mit Hilfe des Hintergrundbildes bestimmt wurde, dargestellt. Das Histogramm des Differenzbildes besteht hauptsächlich aus kleinen Werten, da der Hintergrund den größten Teil bildet. Aus diesem Grund eignen sich die meisten herkömmlichen Methoden der Bildverarbeitung nicht, wie z.B. [20, 43, 70], was zu einer drastischen Reduzierung der zur Verfügung stehenden Anzahl an Methoden führt.

Um die verschiedenen Verfahren qualitativ und quantitativ beurteilen zu können, wurde eine Test-Sequenz (siehe Abb. 4.21) aufgenommen, die für alle adaptiven Verfahren benutzt wurde. In dieser Sequenz läuft eine Person vor einem Schrank von links nach rechts. Einige Frames aus dieser Sequenz sind in Abb. 4.21 dargestellt. Diese Test-Sequenz zeichnet sich dadurch aus, daß sie sehr empfindlich auf Änderungen des Schwellwertes reagiert. Für die Untersuchung der adaptiven Methoden wurde diese mit der erweiterten Zeitrekursion kombiniert, um so einen Vergleich zwischen konstantem und adaptiven Schwellwert ziehen zu können. Zum Vergleich wurde die Test-Sequenz mit dem Kalman-Filter aus Abschnitt 4.2.4 mit einem konstantem Schwellwert von T = 30 bearbeitet. In Abb. 4.22 ist die Maske M(t) als Simulationsergebniss dargestellt. In Frame 18 ist hinter der Person eine Region gekennzeichnet, die fälschlicherweise als eine Veränderung detektiert wird. Auch in Frame 30 und Frame 33 sind Fehler am Rücken der Person zu erkennen, die durch den Schatten der Person selbst erzeugt werden.



Abbildung 4.22: Simulationsergebnis der Test-Sequenz mit dem Kalman-Filter

4.2.5.1 Methode nach Otsu

In [62] wird ein Verfahren vorgestellt, das einen optimalen Schwellwert aus dem Histogramm eines Grauwertbildes bestimmt. Zur Bestimmung des globalen Schwellwertes werden keine a-priori Informationen benötigt. Geht man von der Menge der Grauwerte $\{0, 1, 2, \ldots, L - 1\}$ aus, werden diese – nach Otsu – durch den Schwellwert T in zwei Klassen unterteilt. In diesem Fall sind es die Klassen Vordergrund und Hintergrund. Mit der Häufigkeit der Bildpunkte n_i , die den Grauwert i besitzen, und der Anzahl N aller Bildpunkte eines Frames, kann die normierte Häufigkeitsverteilung h_i der Grauwerte bestimmt werden.

$$h_i = \frac{n_i}{N} \tag{4.38}$$

Die Auftrittswahrscheinlichkeit h_H , der Mittelwert μ_H und die Varianz σ_H der Klasse Hintergrund sind in Gleichung (4.39) definiert. Es ist wichtig an dieser Stelle zwischen der Klasse Hintergrund und dem Hintergrundbild B(t) zu unterscheiden, denn der Hintergrund ist nur ein Teil des Differenzbildes D(t) und wird über den Schwellwert vom Vordergrund getrennt. Mit den Verfahren aus den Abschnitten 4.2.1 bis 4.2.4 wird das Hintergrundbild an den aktuell ermittelten Hintergrund angepaßt, d.h. das Hintergrund und Hintergrundbild sich unterscheiden.

$$h_{H} := \sum_{i=0}^{T-1} h_{i} = h_{H}(T)$$

$$\mu_{H} := \sum_{i=0}^{T-1} i \cdot \frac{h_{i}}{h_{H}} = \frac{\mu(T)}{h_{H}(T)}$$

$$\sigma_{H}^{2} := \sum_{i=0}^{T-1} (i - \mu_{H})^{2} \cdot \frac{h_{i}}{h_{H}(T)}.$$
(4.39)

Für den Vordergrund, der ebenfalls aus dem Differenzbild berechnet wird, gelten die gleichen Beziehungen.

$$h_{V} := \sum_{i=T}^{L-1} h_{i} = h_{V}(T) = 1 - h_{H}(T)$$

$$\mu_{V} := \sum_{i=T}^{L-1} i \cdot \frac{h_{i}}{h_{V}} = \frac{\mu(L) - \mu(T)}{1 - h_{H}(T)} \text{ mit } \mu(L) = \sum_{i=0}^{L-1} i \cdot h_{i}$$

$$\sigma_{V}^{2} := \sum_{i=T}^{L-1} (i - \mu_{V})^{2} \cdot \frac{h_{i}}{1 - h_{H}(T)}$$
(4.40)

Für jeden Schwellwert T gilt:

$$h_V(T) \cdot \mu_V(T) + h_H(T) \cdot \mu_H(T) = \mu(L).$$
 (4.41)

Um die *Güte* des Schwellwertes quantitativ beurteilen zu können, hat Otsu die drei Kriterien $\lambda(T)$, $\kappa(T)$ und $\eta(T)$ eingeführt.

$$\lambda(T) := \frac{\sigma_B^2(T)}{\sigma_W^2(T)}$$

$$\kappa(T) := \frac{\sigma^2(L)}{\sigma_W^2(T)}$$

$$\eta(T) := \frac{\sigma_B^2(T)}{\sigma^2(L)}$$
(4.42)

Die Kriterien benutzen dabei die Varianz der Grauwerte in den Klassen (Within)

$$\sigma_W^2(T) := h_H(T) \cdot \sigma_H^2(T) + h_V(T) \cdot \sigma_V^2(T), \qquad (4.43)$$

die Varianz der Grauwerte zwischen den Klassen (Between)

$$\sigma_B^2(T) := h_H(T) \cdot (\mu_H(T) - \mu(L))^2 + h_V(T) \cdot (\mu_V(T) - \mu(L))^2$$

= $h_H(T)h_V(T)) \cdot (\mu_V(T) - \mu_H(T))^2$ (4.44)

und die Varianz aller Grauwerte eines Frames $\sigma^2(L)$. Der Schwellwert T ist genau dann optimal, wenn einer der drei Kriterien aus 4.42 maximal wird, wodurch die beste Trennung der Klassen erzielt wird. Alle drei Kriterien sind untereinander äquivalent und können mittels

$$\sigma^2(L) = \sigma_B^2(T) + \sigma_W^2(T) \tag{4.45}$$

in die Formen

$$\kappa(T) = \lambda(T) + 1$$

$$\eta(T) = \lambda(T)(\lambda(T) + 1)$$
(4.46)

gebracht werden. Die Maxima der drei Kriterien werden somit bei ein und demselben Schwellwert T erreicht. Daher genügt es, das Maximum eines Kriteriums zu bestimmen.



Abbildung 4.23: Simulationsergebnis der Test-Sequenz mit adaptivem Schwellwert nach der Methode von Otsu



Abbildung 4.24: Schwellwertverlauf: Methode nach Otsu

Da in Gleichung (4.44) – im Vergleich zu Gleichung (4.43) – nur Momente erster Ordnung benutzt werden, bietet es sich an, das Kriterium $\eta(T)$ zur Bestimmung des optimalen Schwellwertes zu benutzen. Da $\sigma^2(L)$ keine Funktion von T ist und somit in jedem Frame konstant ist, genügt es das Maximum von $\sigma_B^2(T)$ zu bestimmen. Der Schwellwert ist also genau dann optimal, wenn

$$\sigma_B^2(T) = \frac{(h_V(T)\mu(L) - \mu(T))^2}{h_V(T) \cdot (1 - h_V(T))}$$
(4.47)

maximal ist.

Die Test-Sequenz mit dem adaptiven Schwellwert ist beispielhaft in Abb. 4.23 mit drei Frames dargestellt. Zum Vergleich sind in Abb. 4.22 die Simulationsergebnisse, die mit dem Kalman-Filter und einem festem Schwellwert T = 30 berechnet wurden, abgebildet. In allen Frames ist bei der adaptiven Berechnung des Schwellwertes eine leichte Verbesserung der Ergebnisse zu beobachten.

Die Berechnung der adaptiven Schwellwerte für die Differenzbilder zeigte, daß bei der adaptiven Bestimmung des Schwellwertes eine untere Schranke festgelegt werden mußte. Dies wird deutlich, wenn man sich den Verlauf des Schwellwertes in Abb. 4.24 betrachtet. Es treten Situationen auf, in denen der Algorithmus so niedrige Schwellwerte berechnet, die für Trennung des Vordergrundes vom Hintergrund aus dem Differenzbild nicht geeignet sind. Der Verlauf des Schwellwertes liegt bis Frame 12 auf einer festgelegten unteren Schranke bei T = 10. Zu diesem Zeitpunkt betritt die Person den von der Kamera beobachteten Bereich. Der Schwellwert steigt bis auf T = 54, wodurch die leicht besseren Ergebnisse in den Frames 18 und 30 zu erklären sind. Sobald die Person wieder aus der beobachteten Szene heraustritt, sinkt der Schwellwert ab. Der Anstieg des Schwellwertes ab Frame 35 wird durch die Person erzeugt, da sie sich umdreht und wieder in die Szene läuft.

Wie die Test-Sequenz gezeigt hat, eignet sich die Methode von Otsu sehr gut für die adaptive Schwellwertbestimmung. Vor der Anwendung dieser Methode muß sorgfältig untersucht werden, ob die Ergebnisse den Rechenaufwand rechtfertigen können. Ist es möglich Beleuchtungsänderungen im aufgenommenen Bild zu vermeiden, wie es bei der NIR-Bildaufnahme der Fall ist, ist eine adaptive Berechnung des Schwellwertes nicht notwendig.

4.2.5.2 Iterative Schwellwertbestimmung

Bei dieser heuristischen Methode werden – ähnlich wie bei der von Otsu – Vorder- und Hintergrund durch den Schwellwert T in zwei Klassen unterteilt. Die Unterscheidungsmerkmale dieser Klassen sind in diesem Fall die Mittelwerte $\mu_V(T), \mu_H(T)$ und die Standardabweichungen $\sigma_V(T), \sigma_H(T)$. In den Frames in denen keine Veränderungen vorkommen, werden $\mu_H(T)$ und $\sigma_H(T)$ wie auch $\mu_V(T)$ und $\sigma_V(T)$ sehr klein sein. Treten Änderungen in den Eingangsbildern auf, sollten sich nur der Mittelwert und die Standardabweichung des Vordergrundes ändern. $\mu_H(T)$ und $\sigma_H(T)$ sollten in der Regel konstant sein, so daß der Schwellwert T nur von den Eigenschaften des Vordergrundes abgeleitet werden kann. Es genügt also den Mittelwert und die Standardabweichung des Differenzbildes zu bestimmen, um einen Schwellwert zu bestimmen.

$$T = \begin{cases} \mu(L) + \sigma(L) & \text{für } \mu(L) + \sigma(L) > T_{min} \\ T_{min} & \text{sonst.} \end{cases}$$
(4.48)

Die Simulationsergebnisse in den Abbildungen 4.25 und 4.26 der iterativen Methode sind verglichen mit der Methode von Otsu etwas schlechter. Die Unterschiede sind aber marginal und bei Betrachtung der Effektivität, in dem man den Rechenaufwand berücksichtigt, sind beide Methoden gleich gut geeignet. Der Schwellwertverlauf ist aufgrund der Segmentierungsergebnisse der vorigen Methode ähnlich. Der Verlauf ist nur um einen Offset von ca 10 - 15 zu niedrigeren Werten versetzt. Für die adaptiven Methoden wurde der

Verarbeitung	Integer Operation	Float Operation	${ m Speicherbedarf}$
Methode nach Otsu	L	T + L + 6	2 Float
Iterative Schwellwertbestimmung	L	4L + 2	2 Integer

Tabelle 4.2: Rechenaufwand und Speicherbedarf pro Schwellwert der adaptiven Schwellwertmethoden



Abbildung 4.25: Maske: Iterative Schwellwertbestimmung



Abbildung 4.26: Schwellwertverlauf: Iterative Schwellwertbestimmung

Rechenaufwand und Speicherbedarf wie bei den pixelorientierten Methoden in Tabelle 4.2 zusammengefaßt. Die Angaben beziehen sich auf die Berechnung eines Schwellwertes. Bei der Iterativen Methode kann nur ein Schwellwert berechnet werden, so daß die Angaben auf ein Frame bezogen richtig sind. Bei der Methode nach Otsu handelt es sich ja um die Suche eines Maximums, so daß für jeden Schwellwert die angegebenen Operationen und Speicher vorgehalten werden muß. Da die Schwellwerte bei Differenzbildern meistens niedrige Werte annehmen, kann durch eine geschickte Suche und Wahl des Startpunktes, die Anzahl der zu berechnenden Schwellwerte klein gehalten werden. Aber schon nach vier berechneten Schwellwerten ist der Rechenaufwand der Methode nach Otsu höher als bei der iterativen Methode.

4.3 Regionenorientierte Bildauswertung

Der regionenorientierten Bildauswertung steht neben dem Eingangsbild die binäre Maske der pixelorientierten Verarbeitung zur Verfügung. Da die Ergebnisse der pixelorientierten Verarbeitung ausreichende Merkmale zur Auswertung enthalten und die Anforderungen an die Rechenleistung niedrig gehalten werden sollen, wurde zuerst die binäre Maske zur Bestimmung des Zustandes benutzt.

Der optische Innenraumschutz hat die Aufgabe, wie in Abschnitt 2.2 erläutert wurde, unbefugtes Eindringen in das Fahrzeug zu verhindern. Die pixelorientierte Verarbeitung in Kombination mit der NIR-Bildaufnahme (Abschnitt 3.6) stellt der regionenorientierten Auswertung ein binäres Bild zur Verfügung, in der alle Änderungen markiert sind, die durch Bewegung von Objekten erzeugt wurden. Eine erste Realisierung kann die Messung der veränderten Flächen sein, die direkt in der Maske durchgeführt werden kann (wie in [67] gezeigt). Der vordere bzw. hintere Innenraum wurde dabei in zwei Seiten geteilt, um nicht einen Alarmzustand zu melden, der durch die summierten Veränderungen auf beiden Seiten verursacht wird.

In Abb. 4.27 ist für die Diebstahl-Sequenz aus Abschnitt 4.2 (siehe Abb. 4.1) die Anzahl der Pixel, die als Veränderung markiert wurden, für Fahrer- und Beifahrerseite über den Frames aufgetragen. Als Vorverarbeitung wurde die erweiterte Zeitrekursion aus Abschnitt 4.2.3.2 mit denselben Parametern wie in Abb. 4.14 verwendet. Man kann eindeutig erkennen, daß der Eingriff auf der Beifahrer-Seite stattfindet. Die Hand der Person greift bei Frame 15 in den Innenraum und verläßt diesen bei Frame 35. Eine erste einfache Auswertung könnte einen Alarm melden, wenn mehr als 1500 Pixel auf Fahrer- oder Beifahrerseite als verändert markiert werden.

Die gleichen Verläufe für die Fehlalarm-Sequenz sind in Abb. 4.28 abgebildet (siehe Abschnitt 4.2 und Abbildungen 4.15 und 4.2). In dieser Sequenz dringt der Kopf der Person zwischen den Frames 84 bis 86 in den Innenraum ein. Der OIS darf bei dieser Sequenz frühestens ab Frame 84 einen Alarmzustand melden. Die einfache Realisierung des OIS, die nur die Anzahl der als verändert markierten Pixel zählt, würde bei einer Alarm-Schwelle von 1500 Pixeln ab Frame 19 einen Fehlalarm melden.

Die fehlerhafte Interpretation entsteht durch den großen Feldwinkel des Objektivs. Dadurch werden auch Bereiche außerhalb des Fahrzeug erfaßt. Dies ist zwar für die Überwachung der Bereiche um das Fahrzeug von Vorteil, doch die Auswertung muß diesen Außen-



Abbildung 4.27: Pixelanzahl: Diebstahl-Sequenz



Abbildung 4.28: Pixelanzahl: Fehlalarm-Sequenz



Abbildung 4.29: Pixelanzahl Innenraum: Diebstahl- und Fehlalarm-Sequenz

raum vom Innenraum trennen. Mit Hilfe den in Abschnitt 3.7 vorgestellten Kalibrationsmasken (Abb. 3.53) ist es einfach möglich, das Eingangsbild in Außen- und Innenraum zu unterscheiden. Damit kann ein Voralarm realisiert werden, wenn in der Umgebung im Außenraum des Fahrzeugs Bewegungen stattfinden. Trennt man also den Außenraum und zieht die Anzahl der Pixel ab, die zum Außenraum gehören, ergeben sich neue Verläufe in Abb. 4.29, die nur die Anzahl der Pixel enthalten, die im Innenraum verändert wurden. Anhand dieser neuen Verläufe kann man für die Diebstahl- und die Fehlalarm-Sequenz eine neue Alarm-Schwelle definieren. In Abb. 4.29 wurde nur die Beifahrerseite dargestellt, da auf der Fahrerseite – wie schon die Abbildungen 4.27 und 4.28 gezeigt haben – keine großflächigen Änderungen stattgefunden haben.

Würde man den Alarm-Schwellwert auf 700 Pixel festlegen würden in beiden Sequenzen zum richtigen Zeitpunkt der Alarmzustand erkannt. Bemerkenswert ist, daß man in Abb. 4.29 bei der Diebstahl-Sequenz bei Frame 26 sehen kann, das das Objekt vom Beifahrersitz bewegt wird. Durch die Wegnahme des Objektes steigt die Anzahl der veränderten Pixel sprunghaft von 1100 Pixel auf 1400 Pixel an. Der Wert von ca. 500 Pixel ab Frame 35 entspricht ungefähr der Größe des Objektes und sinkt langsam ab, da das Objekt in den folgenden Frames fehlt.

Die Verläufe der Pixelanzahl, die in Abb. 4.30 zu sehen sind, lassen erkennen, daß eine Vorstufe zum tatsächlichen Alarm aus der Anzahl der Pixel im Außenraum abgeleitet werden kann. In der Fehlalarm-Sequenz erreicht die Person das Fahrzeug zwischen Frame 17 und 18. In der Diebstahl-Sequenz nähert sich die Person bis Frame 10 dem Fahrzeug und entfernt sich ab Frame 38. Die Vorstufe könnte zur Einsparung des Leistungsverbrauchs, der bei Diebstahlwarnanlagen immer kritisch ist, eingesetzt werden, indem ausschließlich der Außenraum beobachtet wird, bis eine Voralarm-Schwelle im Außenraum überschritten wird. Danach würde der gesamte Innen- und Außenraum beobachtet werden.

In der binären Maske sind viele einzelne Pixel enthalten, die als Rauschen betrachtet



Abbildung 4.30: Pixelanzahl Außenraum: Diebstahl- und Fehlalarm-Sequenz

werden können und nicht zum Objekt gehören. Mit den elementaren morphologischen Operatoren [42] Erosion und Dilatation kann ein Opening und Closing durchgeführt werden, mit dem gezielt kleine Pixelansammlungen entfernt werden können, ohne die Größe der eigentlichen Objekte zu verändern. Die verbleibenden Regionen werden für die weitere Verarbeitung in einer Merkmalsliste aufgenommen. Diese Merkmale werden Formmerkmale [42] bzw. Shape Feature [40] genannt. Die Formmerkmale, die für den OIS eine wichtige Rolle spielen, sind:

- Grösse: Die Anzahl der Pixel einer Region.
- Schwerpunkt
- Extrempunkte: Das sind die Extremwerte der Bildpunkte in x- und y-Richtung.

Die Extremwerte der Bildpunkte einer Region sind diejenigen Randpunkte, die in x- und y-Richtung minimale bzw. maximale Koordinatenwerte annehmen. Sie werden im folgenden zur Abkürzung auch Extrempunkte genannt.

Richtet man sich bei der Größe eines Objektes an die Richtlinien der Zulassungsbehörden und Versicherungen (Abschnitt 2.2) können alle Regionen die kleiner sind vernachlässigt werden. Berücksichtigt man zusätzlich, daß bei geschlossenen Fahrzeugen eine Person nur von Außen eindringen kann, können auch die Regionen vernachlässigt werden, die nicht mit dem Außenbereich verbunden sind. Für den OIS sind nur die Regionen interessant, die eine vorgeschriebene Größe überschreiten und aus dem Außenbereich kommen. Die Bilder in 4.31 und 4.32 wurden unter Berücksichtigung dieser Punkte berechnet.

Die Objekte, die die Kriterien erfüllen wurden in eine *Bounding Box* eingerahmt. Die Bewegung der Extrempunkte des Objektes ist in Abb. 4.34 dargestellt. Hinzu kommt die Größe des jeweiligen Objekts, die für Innen- und Außenraum in Abb. 4.33 angegeben



Abbildung 4.31: Formmerkmale: Diebstahl-Sequenz



Abbildung 4.32: Formmerkmale: Fehlalarm-Sequenz



Abbildung 4.33: Größe der Objekte in der Diebstahl-Sequenz nach Außen- und Innenraum getrennt

wurde. Aus dem Verlauf der Größe des Objekts im Außenraum kann der Zustand eines Voralarms objektbezogen abgeleitet werden. Dieser Zustand tritt in der Diebstahl-Sequenz bei Frame 5 ein und hält bis Frame 46 an. Eine Schwelle bei 400 Pixel bietet einen ausreichenden Schutz vor zu vielen Voralarmen. Der eigentliche Eingriff des Objekts wird bei Frame 17 gemeldet, obwohl der Eingriff schon bei Frame 15 erfolgt. Diese Verzögerung wird durch die Mindestgröße eines Objektes verursacht, d.h. die Größe des Objektes im Innenraum zwischen den Frames 15 bis 17 liegt unter der Mindestgröße von 100 Pixel. Aus dem Zusammenhang, daß das Objekt bei Frame 17 die Mindestgröße überschreitet, den Innenraum verletzt und im Außenraum einen Voralarm ausgelöst hat, kann bei Frame 17 der Zustand des Alarms gemeldet werden.

Anhand der Extrempunkte eines Objektes kann geschlossen werden wann und wie tief der Innenraum von diesem Objekt verletzt wird. Abb. 4.34 zeigt, daß die Person bis zu 75 Pixel tief in den Innenraum eindringt, während es mit seinen Minima außerhalb des Fahrzeug bleibt. In dieser Abbildung sind die Randpunkte von allen Objekten und allen Frames eingezeichnet worden. Für jedes Objekt werden zwei Punkte eingezeichnet: Der Minimumpunkt in x- und y-Richtung und der Maximumpunkt. Der Minimumpunkt eines Objektes, der in Bildkoordinaten angeben wird, ist der Punkt eines Objektes, der die kleinsten x- und y-Koordinaten besitzt. Beim Maximalpunkt sind es die größten x- und y-Koordinaten. Der Innenraum des Fahrzeugs beginnt bei y = 110 Pixel und x = 16 Pixel. Die Längen werden im Gegensatz zu Tabelle 3.1 absichtlich in Pixel angegeben, da durch das Fischaugen-Objektiv die Verzeichnungen so groß sind, daß eine Längenangabe in Meter ohne Berücksichtigung der Verzeichnungen nicht sinnvoll ist. Die Berücksichtigung der Verzeichnungen schlasten viel zu Rechenaufwendig und zweitens nicht notwendig, wie die weitere Auswertung noch zeigen wird. Die Koordinaten-Angaben sind Bildkoordinaten.



Abbildung 4.34: Extrempunkte der Objekte in der Diebstahl-Sequenz

Interessant wird es für die Fehlalarm-Sequenz, die nur eine sehr kurze und geringfügige Verletzung des Innenraums enthält. Der Verlauf der Objektgrößen Sequenz wurde für Außen- und Innenraum in Abb. 4.35 getrennt aufgezeichnet. Anhand der Größe der Objekte im Außenraum, kann der Zustand Voralarm von Frame 14 bis zum Ende der Sequenz detektiert werden. Im Innenraum wird bei Frame 18 die Mindestgröße überschritten. Ein Alarm muß aber nicht unbedingt ausgelöst werden, da die Person nicht tief in das Fahrzeug hinein schaut (siehe Abb. 4.36).

Für den Innenraumschutz ist die Information des geometrischen Schwerpunktes eines Objektes nicht unbedingt entscheidend. Trotzdem wurden Untersuchungen bezüglich des Schwerpunktes unternommen, die u.a. für die optische Belegungszustandserkennung oder andere Anwendungen im Fahrzeug (siehe Abschnitt 6.1) wichtig sind. Bei der Betrachtung des Schwerpunktes ist einmal die aktuelle Lage und anderseits der Weg, den der Schwerpunkt in einer Sequenz zurücklegt, entscheidend. In Abb. 4.37 ist eine typische Sequenz abgebildet, in der eine Person in ein Fahrzeug auf der Beifahrerseite einsteigt. Die Schwerpunkte der als Objekte erkannten Regionen wurden in die jeweiligen Frames mit einem weißen Kreuz gekennzeichnet. Das Öffnen der Beifahrertür erzeugt so große Änderungen, daß ein Objekt an dieser Stelle detektiert wird (Frame 23 - 38). Bei Frame 42 ist die Person zu erkennen und wird als separates Objekt detektiert. Der Schwerpunkt auf dem Beifahrersitz wird durch die starken Schwankungen der Beleuchtung erzeugt. Nachdem die Person im Fahrzeug sitzt, wird nur sie als Objekt segmentiert und der Schwerpunkt korrekt im Bereich des Rumpfes ermittelt.

Die Abb. 4.38 zeigt alle geometrischen Schwerpunkte der als Objekte detektierten Regionen im ersten Frame der jeweiligen Sequenz übereinander eingezeichnet. Hier sieht man deutliche Unterschiede zwischen der Fehlalarm-Sequenz und den beiden anderen Sequenzen. Das Bild der Fehlalarm-Sequenz wurde abgedunkelt, damit die Schwerpunkte besser sichtbar werden. In keinem der Frames der Fehlalarm-Sequenz dringt der Schwerpunkt



Abbildung 4.35: Größe der Objekte in der Fehlalarm-Sequenz nach Aussen- und Innenraum getrennt



Abbildung 4.36: Extrempunkte der Objekte in der Fehlalarm-Sequenz



Abbildung 4.37: Einstiegssequenz mit geometrischen Schwerpunkten



Abbildung 4.38: Schwerpunkte der Diebstahl-, Fehlalarm- und Einstiegssequenz

in den Innenraum des Fahrzeugs ein. Wie man in dem Bild der Diebstahl-Sequenz sieht dringt der Schwerpunkt nicht so tief in den Innenraum, daß eine robuste Alarm-Erkennung realisiert werden könnte. Die Auswertung des Merkmals Schwerpunkt eignet sich eher für andere Anwendungen als den OIS.

Wie die morphologische Auswertung gezeigt hat, ist es möglich, robust einen Zustand zu bestimmen. Die Notwendigkeit dieser Schritte liegt in den Störungen der Vorverarbeitung durch Beleuchtungsänderungen. Hierzu wurde aber in Abschnitt 3.6 die NIR-Bildaufnahme vorgestellt, die im Zustand des Voralarms eingeschaltet werden kann. Die Eingangsbilder, die von der NIR-Bildaufnahme geliefert werden, sind, wie gezeigt wurde, von Beleuchtungsänderungen weitgehend unabhängig.

Der Rechenaufwand für die regionenorienterte Auswertung ist aufgrund der vielen möglichen Zustände schwer zu ermitteln. Als erstes muß der Aufwand für die Lokalisierung einzelner Regionen in der binären Maske berechnet werden. Dieser ist aber je nach Recheneinheit sehr unterschiedlich und hängt auch von der Anzahl und sogar der Form der zu erwartenden Objekte ab. Die Lokalisierung der Objekte muß für den Außenraum und den Gesamtraum separat durchgeführt werden. Danach werden die Formmerkmale für jede Region im Innen- und Gesamtraum bestimmt, die die Mindestgröße überschreitet. Diese Operationen sind stark von der Größe des Objektes abhängig. Die Bestimmung eines Alarmzustandes erfordert im Vergleich zu den vorhergehenden Berechnungen erheblich weniger Rechenaufwand, so daß auf diese Angabe verzichtet wurde.



Abbildung 4.39: Optischer Innenraumschutz

4.4 Zusammenfassung

In diesem Kapitel wurden Algorithmen vorgestellt, die durch die Kombination von Merkmalsextraktion und Auswertung einen optischen Innenraumschutz darstellen. Das Konzept, wie es in Abb. 4.39 dargestellt ist, sieht für die Bildaufnahme zwei Varianten vor. Bei ausreichender Helligkeit ist es nicht unbedingt notwendig die NIR-Bildaufnahme einzuschalten, es reicht aus die einfache Bildakquisition zu verwenden. Dies bietet die Möglichkeit Energie einzusparen, die wiederum für die NIR-Beleuchtung bei niedrigen Helligkeiten zur Verfügung steht. Zur Detektion von Veränderungen hat sich gezeigt, daß die erweiterte Zeitrekursion die besten Ergebnisse im Verhältnis zum Rechenaufwand leistet. Die adaptive Schwellwertbestimmung ist höchstens optional, da das Verhältnis zwischen Rechenaufwand und Ergebnisverbesserung – besonders bei der NIR-Bildaufnahme – nicht ausreichend ist. Die regionenorientierte Auswertung setzt vergleichsweise wenig Merkmale ein, um einen Einbruch oder Diebstahl zu erkennen. Die Simulationen zeigen aber keinen Bedarf an komplexen Verfahren, insbesondere in Hinblick auf den Energieverbrauch ist es sinnvoller mit Hilfe der NIR-Bildaufnahme die Merkmale robuster aufzunehmen.

Die Simulationen haben gezeigt, daß eine Änderung des Grauwertes nicht zwangsläufig durch eine Bewegung eines Objekts verursacht werden muß (siehe [76]). In den gezeigten Sequenzen kommt es bei den Aufnahmen ohne NIR-Beleuchtung zu fehlerhaften Detektionsergebnissen durch Beleuchtungsänderungen. Mit der regionenorientierten Analyse ist es aber möglich, die fehlerhaften Regionen durch einfache Operationen zu unterdrücken. Hier wurde gezeigt, daß eine robuste Auswertung der binären Masken vollkommen ausreicht. Durch die alleinige Auswertung der binären Masken konnten die Anforderungen an die Rechenleistung niedrig gehalten werden. Ein Schritt weiter geht die Implementierung der pixelorientierten Verarbeitung in den Bildsensor bzw. in jeden Bildpunkt. In [15] wurde eine Untersuchung durchgeführt, ob die einfache Zeitrekursion (siehe [65], Abschnitt 4.2.3.1) auf einem CMOS-Bildsensor integrierbar ist. Damit stehen am Aus-



Abbildung 4.40: Schema einer pixelorientierten On-Chip Verarbeitung

gang des Bildsensors zwei Bilder (Grauwertbild und binäre Maske) zur Verfügung. Es ist weiterhin möglich einfachste regionenorientierte Auswertungen auf dem Bildsensor zu realisieren. Ein großer Vorteil liegt in der generellen Verwendbarkeit dieser Algorithmen, da viele Bildverarbeitungsanwendungen diese Vorverarbeitung benutzen können. Die Abbildung Abb. 4.40 zeigt ein Konzept zur Integration von pixelorientierter Bildverarbeitung auf einen Bildsensor.

Neben der guten Integrierbarkeit können die pixelorientierten Verfahren auch sehr effektiv programmiert werden. Da die Auslese des Sensors die Pixelwerte mit einer gewissen Frequenz am Ausgang der Sensor-Matrix zur Verfügung stellt, kann die pixelorientierte Verarbeitung schon nach dem ersten Pixelwert beginnen und ist nach dem letzten Pixelwert mit der Verarbeitung des aktuellen Bildes fertig. Somit verringert sich die Bearbeitungszeit für jedes Frame auf die Bearbeitungszeit eines Bildpunktes.

Die Simulationen haben gezeigt, daß der optische Innenraumschutz (OIS), wie er in diesem Kapitel vorgeschlagen wurde, bei offenen Fenstern funktioniert. Dies legt die Vermutung nahe, daß der OIS auch in offenen Fahrzeugen eingesetzt werden kann. Weiterhin ist der OIS gegenüber kleinen Objekten unanfällig, da die Auslösung eines Alarms nicht allein von der Größe eines Objekts abhängig gemacht wird.

Durch die Benutzung der NIR-Bildaufnahme werden die Artefakte, die durch Beleuchtungsänderungen hervorgerufen werden, nahezu vollständig unterdrückt. Auf den generellen Einsatz wurde aus Gründen der Energieeinsparung verzichtet. Es wurde versucht allein durch die pixel- und regionorientierte Verarbeitung nur Objekte zu segmentieren. Die NIR-Bildaufnahme könnte in Situationen wie z.B. häufigen Voralarmen eingeschaltet werden, um eine noch sichere Aussage treffen zu können. Die Fälle dürften nicht zu oft auftreten, um die Grenzwerte bezüglich des Energieverbrauchs von FAS einzuhalten.