

# **Lokalisation und Verfolgung von Personen in Echtzeit unter Verwendung kooperierender 2D/3D-Kameras**

genehmigte  
DISSERTATION  
zur Erlangung des Grades eines Doktors der Ingenieurwissenschaften

vorgelegt von  
Dipl.-Phys. Omar Edmond Loepprich

eingereicht bei der Naturwissenschaftlich-Technischen Fakultät  
der Universität Siegen  
Siegen 2013

1. Gutachter: Prof. Dr.-Ing. Otmar Loffeld
2. Gutachter: Prof. Dr.-Ing. Madjid Fathi-Torbaghan

Tag der mündlichen Prüfung: 24-01-2014.

# **EINLEITUNG UND VORWORT**

Die grundsätzliche Zielsetzung von intelligenten Systemen zur Überwachung von Innenräumen besteht in der automatisierten Interpretation von Aktionen in dem zu betrachtenden Gebiet. Eine Prämisse hierfür ist die Fähigkeit der Detektion, Lokalisation, Verfolgung und Klassifikation von Objekten im entsprechenden Zielbereich.

Entscheidend für die Konzeption eines Überwachungssystems ist weiterhin das primäre Einsatzgebiet. Während für kleine Überwachungsbereiche und unter der Annahme nur weniger gleichzeitig vorhandener Objekte eventuell die Verwendung eines einzelnen Sensorsystems ausreicht, bedarf es im Fall größerer Bereiche, mit mehreren gleichzeitig vorhandenen und zu verfolgenden Objekten, in der Regel einer verteilten und miteinander kooperierenden Sensorik.

Ein in letzter Zeit verstärkt verfolgter Ansatz zur Steigerung der Leistung der Prozesse von Detektion, Lokalisation und Klassifikation liegt dabei in der Verwendung einer Kombination von 2D- und 3D-Informationen, welche, je nach Anforderung, auf Pixel-, Merkmal- oder Entscheidungsebene miteinander fusioniert werden. Die Kombination unterschiedlicher Systeme zu 2D/3D-Messsystemen ist jedoch keineswegs trivial. Zunächst bedarf es im Falle der Betrachtung beweglicher Objekte einer zeitlichen Synchronisation der anfallenden Daten. Im Hinblick auf eine effektive Fusion ist weiterhin eine Registrierung der Daten der unterschiedlichen Quellen notwendig. Tritt zusätzlich die Notwendigkeit einer koordinierten Verwendung mehrerer Systeme auf, so kann das aufgrund der sich ergebenden Komplexität in Bezug auf die Handhabung der informationsliefernden Einheiten in einer inakzeptablen Operation des Gesamtsystems münden.

Die am Zentrum für Sensorsysteme (ZESS) entwickelte MultiCam, ein zurzeit einmaliges Sensorsystem zur kombinierten Akquisition von CMOS-basierten Intensitäts- (2D) und Time-of-Flight (ToF) basierten Distanzdaten (3D) einer gesamten Szene, erlaubt aufgrund ihres monokularen Aufbaus eine drastische Vereinfachung der Registrierung der 2D- und 3D-Informationen. Weiterhin ermöglicht die interne Logik, untergebracht in einem dedizierten FPGA, eine akkurate Synchronisation der unterschiedlichen Informationsströme. Durch die Reduktion der Komplexität der Prozesse der Registrierung sowie Synchronisierung der 2D- und 3D-Daten bietet sich die Möglichkeit der Kombination einer Vielzahl von simultan operierenden MultiCams zu einem Überwachungssystem an, mit im Vergleich zu konventionellen 2D/3D-Systemen drastisch vereinfachter Handhabung.

Bedingt durch den Einsatz der ebenfalls am ZESS entwickelten Beleuchtungseinheiten, erlauben die im Rahmen dieser Arbeit verwendeten MultiCams die Überwachung eines Distanzbereichs bis ca. 9 m bei einer Bildwiederholrate von 20 Bildern pro Sekunde. Dieses entspricht einem Entfernungsbereich, der zurzeit von ToF-Systemen vergleichbarer Art nur unter Einsatz erhöhter Integrationszeiten und somit reduzierter Bildwiederholraten abgedeckt werden kann. Weiterhin ergibt sich durch die Verwendung einer Linse mit einer Brennweite von 6mm die Möglichkeit der Überwachung von Volumina, in denen sich mehrere Personen simultan aufhalten und miteinander interagieren können. In Verbindung mit der vorhandenen Zeitauflösung kann damit der Bewegungsdynamik der Objekte effektiv Rechnung getragen werden. Aufgrund der genannten Eigenschaften eignet sich die MultiCam somit ideal zur Detektion, Lokalisation, Verfolgung und Klassifikation von Objekten in Innenräumen.

Durch die Verwendung der Informationen von mehreren, simultan operierenden MultiCams mit verschiedenen Ausrichtungen und einem sich teilweise überlappenden Sichtbereich, lässt sich weiterhin eine direkte Erweiterung des Überwachungsbereichs sowie eine Reduktion durch gegenseitige Objektverdeckungen induzierten Uneindeutigkeiten erreichen und ermöglicht die Erzeugung von Objekttrajektorien mit einem erhöhten Informationsgehalt.

## **Einleitung und Vorwort**

Dem Problem der anfallenden Datenmenge, die durch die Verwendung mehrerer simultan operierender MultiCams entsteht, wird durch den Einsatz eines agentenbasierten verteilten Systems begegnet.

Informationsliefernde Agenten, bestehend aus einer MultiCam mit zugehöriger Prozessierungseinheit, übernehmen die Erzeugung von lokalen Informationen bzw. Features bzgl. der aus ihrer Sicht vorhandenen Objekte und senden diese in Form eines Statusvektors an einen zur Fusion aller Informationen zuständigen Agenten. Durch die drastische Reduktion der zu versendenden Informationsmenge wird gewährleistet, dass eine Vielzahl von gleichzeitig operierenden Agenten möglich ist, ohne eine Beschränkung durch die TCP/IP-basierte Kommunikationsinfrastruktur befürchten zu müssen. Weiterhin können aufgrund der gegenseitigen Unabhängigkeit der informationsliefernden Agenten die Daten der vorhandenen MultiCams parallel bearbeitet werden, was eine einfache Skalierung des Gesamtsystems erlaubt.

Ziel dieser Arbeit ist die Diskussion eines solchen agentenbasierten, verteilten Systems zur Detektion, Lokalisation, Verfolgung und Klassifikation multipler Objekte unter Verwendung kooperierender Sensorik auf Basis der MultiCam.

Im ersten Teil dieser Arbeit erfolgt zunächst die Diskussion der Sensorik sowie der Topologie auf Grundlage eines Agentensystems. Beleuchtet werden dabei die Hauptfehlerquellen sowie die notwendige Vorprozessierung der Daten der MultiCam. Zusätzlich dazu werden die Bedingungen angegeben, welche im Rahmen einer gleichzeitigen Operation mehrerer MultiCams mit überlappendem Sichtbereich erfüllt werden müssen. Anschließend erfolgt eine Beschreibung des Agentensystems, der Konfiguration der Agenten sowie des zugrunde liegenden Modells der Datenprozessierung. Abgeschlossen wird der erste Teil durch die Vorstellung des implementierten Operationszyklus des Gesamtsystems.

Der zweite Teil dieser Arbeit dient der Präsentation der Hauptpunkte. Diskutiert werden zunächst die implementierten Verfahren zur Detektion, Lokalisation, Verfolgung sowie der beschleunigten Klassifikation von Objekten auf Basis der unterschiedlichen Modalitäten der MultiCam unter Verwendung des im ersten Teil der Arbeit vorgestellten Datenprozessmodells. Weiterhin werden die Vorteile, welche sich aus der Erhöhung des Informationsgehalts durch die Fusion der Informationen zweier zeitlich synchronisierter, simultan operierender MultiCams ergeben, beleuchtet. Anschließend dazu erfolgt eine Betrachtung der Operationsgeschwindigkeit des Gesamtsystems sowie eine abschließende Diskussion des Potenzials, der momentanen Beschränkung sowie der Anwendbarkeit und Erweiterbarkeit des Ansatzes.

# ABSTRACT

The fundamental objective of intelligent systems targeted towards surveillance of indoor areas lies in the automated interpretation of actions within the monitored space. One premise therefore is the ability to detect, localize, track and classify objects within the surveyed area

In terms of system design, it's also crucial to incorporate the targeted application area. For the case of surveillance of just relatively small volumes and with the additional assumption that only very few objects are simultaneously present within the detection area, data from a single sensor might be sufficient in order to infer adequate information. Increasing the area to be monitored and also allowing multiple objects to dynamically interact in the space to be surveyed, usually leads to the necessity to apply multiple distributed sensor systems which, in order to generate useful information, have to work together in a cooperative manner.

One in recent times increasingly adopted approach in order to increase the performance of detection, localization and classification is to utilize a combination of devices delivering 2D- and 3D-data. Depending on the requirement, fusion of the information is then done on pixel-, feature- or decision-level. The combination of different systems in a 2D/3D-measurement device is not trivial, though. In case of taking moving objects into account, there's a necessity to provide an adequate temporal synchronisation mechanism between the participating devices. With regard to an effective fusion process, the process of registration of the acquired data needs also to be applied. Furthermore, by combining multiple of such 2D/3D-devices into one all encompassing system, the additional need for a coordinated management emerges. Having to cope with all of the aforementioned points might well lead to a system complexity which is intricate to handle and which in turn might result in an unacceptable overall system performance.

The at the Center for Sensorsystems (ZESS) developed MultiCam, an at the moment unique sensor system usable to acquire a combination of CMOS-based intensity (2D) and Time-of-Flight (ToF) based distance information (3D) of a complete scene. Due to the monocular set-up, a drastic simplification of the process of image registration of the 2D- and 3D-images is attained. Furthermore, the internal logic, situated in an embedded FPGA, permits an accurate temporal synchronisation of the different information streams. Due to the reduction of the overall complexity of the processes of registration and temporal synchronisation of the 2D- and 3D-data, it is feasible to combine a multitude of simultaneously operating MultiCams to a surveillance system with, in comparison with conventional 2D/3D-systems, a drastic simplification in terms of handling.

The additional application of the also at the ZESS developed dedicated MultiCam illumination units enables to supervise an area up to a distance of maximal 9m at a camera frame rate of 20fps. With similar ToF devices, the same distance range can only be achieved by an increase of the sensor integration times and which in turn directly leads to a reduced frame rate. Furthermore, as a result of the utilization of a lens with a focal length of 6mm, it is possible to monitor volumes in which multiple persons can simultaneously be present and also interact with each other. Due to the available temporal resolution, one can account for the dynamic behaviour of the objects effectively. As a results of all of the aforementioned properties, the MultiCam is ideally suited for the detection, localization, tracking and classification of objects in indoor areas.

Through the combination of multiple simultaneously operating MultiCams with different orientations and partially overlapping field of views, it is possible to directly extend the surveyed area and also reduce the amount of mutual object occlusion. This directly leads to the generation of object trajectories with an increased information content.

In order to respond to the amount of data generated by the simultaneous operation of multiple devices, an agent based system is utilized. Information delivering agents, composed of one MultiCam with dedicated processing unit, are responsible for the direct processing of the acquired

## Abstract

raw sensor data. Based on their view onto the scene, they generate information in form of a local status vector which they send to a dedicated fusion agent. This fusion agent is in turn responsible for the combination of all of the local information received into one global status vector. Because of the drastic reduction of the information amount necessary to be sent each frame and also because of TCP/IP based communication infrastructure, a simultaneous operation of a multitude of agents is assured. Due to the mutual independence of the participating agents, parallel processing of the acquired data is inherently possible which in turn results in the possibility to be able to scale the system in an easy manner.

The objective of this thesis is the discussion of an agent based, distributed system for the purpose of detection, localisation, tracking and classification of multiple objects through the usage of cooperating MultiCams.

The first part of this thesis begins with the discussion of the used sensor technology as well as the agent based topology. We will focus on the main error sources and the necessary pre-processing of the MultiCam data. In addition to that, the mandatory prerequisites for the operation of multiple cameras with a partial overlap of the field of views will be elucidated. Following that, we give a description of the agent based system as well as an explanation of the underlying processing model this works relies on. The first part concludes with the specification of the complete operation cycle of the implemented system.

The second part starts by addressing the implemented techniques applied for the detection, localization, tracking and accelerated classification of objects by using the different modalities the MultiCam delivers. This will be done within the scope of the data processing model described in the first part of the thesis. Furthermore, the advantages emerging from the usage of two simultaneously operating MultiCams is highlighted. Connected to this, we will also discuss the speed of the implemented system. The final part of this thesis addresses the potential of the approach, it's momentary confinements as well as possible applications and extensions.

## Danksagung

# DANKSAGUNG

Den ersten Dank richte ich an meinen Erstgutachter und Doktorvater Prof. Dr.-Ing. habil. Ottmar Loffeld für seine wissenschaftliche Betreuung sowie den daraus ergangenen Anregungen. Meinem Koreferenten, Herrn Prof. Dr.-Ing. Madjid Fathi-Torbaghan, möchte ebenfalls für die erbrachte Mühe der Begutachtung meiner Arbeit danken.

Großem Dank bin dem Geschäftsführer des ZESS, Herrn Dr. Klaus Hartmann verpflichtet. Trotz seines immensen Arbeitspensums fand er immer die zur Klärung entscheidender Fragen notwendige Zeit. Seine Hinweise und Vorschläge sowie Anregungen haben sich oft als entscheidend und essenziell erwiesen. Zusätzlich danken möchte ich ihm für die Chance, die er mir als Berufseinsteiger geboten hat. Die in im Rahmen meiner Tätigkeit als wissenschaftlicher Mitarbeiter gemachte Projekterfahrung schätze ich als äußerst wertvoll ein, hat sie doch einen nicht unwesentlichen Teil zu meiner beruflichen Entwicklung beigetragen.

Meinem Freund und Kollegen Herrn Dr.-Ing. Seyed Eghbal Ghobadi danke ich für die vielen Diskussionen, Anregungen und Hilfen, weit über das Berufliche hinaus. Seine aufrichtige und ungekünstelte Art war für mich stets ein Quell der Inspiration.

Herrn Dr.-Ing. Oliver Lottner möchte ich für Anregungen und Hilfen bei der Bewältigung mancher Softwareprobleme sowie den wertvollen Hinweisen in Bezug auf die Verwendung der MultiCam danken.

Herrn Wolf Twelsiek, Herrn Ralf Wurmbach sowie Herrn Dr. Wolfgang Weihs möchte ich für ihre Mühe bei der Entwicklung der Sensorik sowie für die Unterstützung bei der Inbetriebnahme von störrischer Hardware ebenfalls von Herzen danken. Herrn Sven Stark danke ich für seine immerwährende Bereitschaft, mir bei filigranen Arbeiten eine unschätzbare Hilfe gewesen zu sein. Ohne die Hilfe der Kollegen am ZESS wäre die Fertigstellung dieser Arbeit schlicht nicht möglich gewesen.

Weiterhin möchte ich den Sekretärinnen des ZESS, Frau Renate Szabo, Frau Ira Dexling sowie Frau Katharina Haut für ihre stets freundliche Art und ihre Hilfsbereitschaft bei der Bewältigung administrativer Aufgaben meinen herzlichen Dank aussprechen.

Meiner Freundin Saskia M. W. Bloemen möchte ich dafür danken, dass ich an ihrer Seite sein darf.

Meinen Eltern, Maria und Hans Loepprich, bin ich für ihre bedingungslose Liebe und Fürsorge mehr als nur dankbar. Dank ihrer grenzenlose Großzügigkeit und Nachsichtigkeit in so vielen Dingen, ist mir die Möglichkeit der Aufnahme eines Studiums sowie der Verfolgung einer Promotion erst gegeben worden.

Amsterdam, im November 2013

Omar Edmond Loepprich



# INHALTSANGABE

	EINLEITUNG UND VORWORT.....	III
	ABSTRACT.....	V
	DANKSAGUNG.....	VII
	INHALTSANGABE.....	IX
1	EINFÜHRUNG.....	11
1.1	MOTIVATION UND ZIELSETZUNG .....	11
1.2	ANSATZ.....	14
1.3	BEITRAG DER ARBEIT.....	22
1.4	GLIEDERUNG.....	23
2	SENSORIK.....	25
2.1	CW-MODULATIONSINTERFEROMETRIE.....	25
2.1.1	ENTFERNUNGSBESTIMMUNG MITTELS HOMODYNVERFAHREN.....	27
2.2	MULTICAM – MONOKULARES 2D/3D-KAMERASYSTEM.....	29
2.2.1	ZEITLICHE SYNCHRONISATION.....	30
2.2.2	LATERALE REGISTRIERUNG.....	31
2.2.3	VORVERARBEITUNG (PRÄPROZESSIERUNG) DER ENTFERNUNGSDATEN.....	31
2.2.4	SIMULTANE OPERATION MEHRERER MULTICAMS.....	36
2.3	FAZIT.....	38
3	AGENTENPARADIGMA-SYSTEMTOPOLOGIE.....	41
3.1	DAS AGENTENPARADIGMA.....	41
3.1.1	AGENTENDEFINITION.....	41
3.1.2	BDI-AGENTENMODELL.....	42
3.1.3	PRS - PROCEDURAL REASONING SYSTEM.....	45
3.1.4	AGENTENKOMMUNIKATION.....	47
3.1.5	IMPLEMENTIERTES AGENTENSYSTEM.....	48
3.1.6	AGENTENARCHITEKTUR.....	50
3.1.7	KONSTITUTION DER VERWENDETEN AGENTEN.....	52
3.1.8	ZEITLICHE SYNCHRONISATION .....	53
3.2	INFORMATIONSFUSION.....	53
3.2.1	FUSIONSTYPEN.....	54
3.2.2	DATENANGLEICHUNG.....	54
3.2.3	JDL DATENFUSIONSPROZESS-MODELL.....	55
3.2.4	PROZESSABLAUF.....	57
3.3	ZUSAMMENFASSUNG.....	58
4	AGENTENBASIERTE DETEKTION, LOKALISATION UND KLASSIFIKATION VON OBJEKTEN.....	63
4.1	OPERATIONEN AUF JDL-PROZESSEBENE 0.....	63
4.1.1	VERTEILTE DATENAKQUISITION.....	63
4.1.2	PRÄPROZESSIERUNG DER SENSORDATEN.....	64
4.2	OPERATIONEN AUF JDL-PROZESSEBENE 1.....	67
4.2.1	OBJEKTDETEKTION.....	67
4.2.2	IMPLEMENTIERTER ALGORITHMUS ZUR OBJEKTDETEKTION.....	71

## Inhaltsangabe

4.2.3	OBJEKTLOKALISATION.....	77
4.2.4	IMPLEMENTIERTER ALGORITHMUS ZUR OBJEKTLOKALISATION.....	82
4.2.5	ZUSÄTZLICHE BETRACHTUNGEN ZUR OBJEKTLOKALISATION.....	85
4.3	OBJEKTKLASSIFIKATION.....	89
4.3.1	OBJEKTKLASSIFIKATION MITTELS ADABOOST.....	90
4.3.2	OBJEKTKLASSIFIKATION MITTELS HOG.....	92
4.3.3	PERSONENERKENNUNG UNTER VERWENDUNG KOOPERIERENDER FUSION.....	95
4.4	ZUSAMMENFASSUNG.....	102
5	OBJEKT- UND PERSONENVERFOLGUNG.....	105
5.1	PROBABILISTISCHE ESTIMATION.....	107
5.2	LOKALE OBJEKT- UND PERSONENVERFOLGUNG (JDL-PROZESSEBENE 1) ..	108
5.2.1	LOKALES OBJEKTMANAGEMENT.....	108
5.2.2	KALMAN-FILTER.....	110
5.2.3	ESTIMATION DER OBJEKTZUSTÄNDE.....	112
5.2.4	ANALYSE DER DYNAMIK.....	113
5.2.5	POSITIONSWICHTUNG.....	115
5.2.6	ERZEUGUNG DES LOKALEN ZUSTANDSVEKTORS.....	117
5.2.7	ANALYSE DES ZEITBEDARFS ZUR ERZEUGUNG LOKALER ZUSTANDSVEKTOREN.....	118
5.3	Globale Objekt- und Personenverfolgung (JDL-Prozessebene 2) ..	118
5.3.1	DATENANGLEICHUNG.....	119
5.3.2	GLOBALES OBJEKTMANAGEMENT.....	120
5.3.3	CONDENSATION-ALGORITHMUS.....	121
5.3.4	ESTIMATION DER OBJEKTZUSTÄNDE.....	122
5.3.5	ANALYSE DER FILTERDYNAMIK.....	123
5.4	STEIGERUNG DES INFORMATIONSGEHALTS.....	127
5.4.1	ERWEITERUNG DES DETEKTIONSBEREICHS.....	127
5.4.2	REDUKTION GEGENSEITIGER OBJEKTVERDECKUNGEN.....	130
5.5	ANALYSE DES ZEITBEDARFS ZUR ERZEUGUNG GLOBALER INFORMATIONEN ..	133
5.5.1	GESCHWINDIGKEIT DER DATENFUSION.....	133
5.5.2	GESCHWINDIGKEIT DES GESAMTSYSTEMS.....	134
6	SCHLUSSBETRACHTUNG.....	137
6.1	POTENZIAL DER SENSORIK.....	137
6.2	BETRACHTUNG DER ZIELSETZUNG UND ERRUNGENSCHAFTEN.....	138
6.3	EINSATZMÖGLICHKEITEN.....	140
6.4	MOMENTANE SYSTEMBESCHRÄNKUNGEN.....	141
6.5	ERWEITERUNGSMÖGLICHKEITEN UND AUSBLICK.....	142
A	LOKALES OBJEKTMANAGEMENT.....	143
B	GLOBALES OBJEKTMANAGEMENT.....	149
C	BAYES'SCHE ESTIMATION.....	153
D	CONDENSATION-ALGORITHMUS.....	157
E	QUATERNIONEN.....	161
F	TABELLEN.....	163
G	LITERATURVERZEICHNIS.....	167

# 1 EINFÜHRUNG

Milliarden Jahre der Evolution haben auf der Erde zu einer schier unglaublichen Anzahl von intelligent handelnden biologischen Organismen geführt. Die intrinsische Fähigkeit der effizienten Wahrnehmung ihrer Umwelt ermöglicht es ihnen dabei, eine Einschätzung komplexer Situationen vorzunehmen und entsprechend zu handeln.

Im Vergleich zu biologischen Organismen bieten die momentan zur Verfügung stehenden technischen Systeme zur Betrachtung der Umgebung sowie anschließender Schlussfolgerung eine weitaus geringere Leistungsfähigkeit. Die Anforderungen sind aufgrund der Vielzahl von auftretenden Parametern enorm, was dazu führt, dass eine entsprechende Handhabung sich als äußerst komplex gestaltet. Es ist daher davon auszugehen, dass es trotz der in den letzten Jahren stattgefundenen Fortschritte auf Gebieten wie der Sensor- und Prozesstechnik sowie der Algorithmik voraussichtlich noch einer langen Zeitspanne bedarf, bis sie mit der Effektivität biologischer Systeme konkurrieren können.

Diesen Umstand wohl erkennend, liegt das Ziel dieser Arbeit in der Vorstellung und Diskussion eines technischen Systems zur Lokalisation und Verfolgung von Objekten bzw. Personen in Innenräumen. Im Rahmen einer übergeordneten, intelligenten Raumüberwachung kann dessen Einsatz der Erzeugung grundlegender Informationen dienen.

## 1.1 MOTIVATION UND ZIELSETZUNG

Der Bereich der intelligenten Innenraumüberwachung hat, nicht zuletzt durch die globalen Ereignisse der letzten Jahre, ein verstärktes Forschungsinteresse erfahren. Allein im Rahmen des 6. und 7. EU-Forschungsrahmenprogramms sind diverse Projekte [1]-[7] etabliert worden, welche als Zielsetzung die Entwicklung intelligenter Überwachungssysteme beinhalten. Gemäß der durch Valera und Velastin in [8] durchgeführten Analyse besteht deren Aufgabe in einer echtzeitfähigen Beobachtung flüchtiger sowie permanenter Objekte in einem definierten Gebiet, mit der Zielsetzung, einer automatisierten Interpretation der in einer Szene beobachtbaren Aktionen. Die Art der Interpretation ist dabei abhängig von der zugrunde liegenden Applikation und kann von simpler Objektpräsenz bis hin zu einer kompletten Situationsanalyse unterschiedlicher Komplexität reichen und führt zu einer entsprechend umfangreichen sowie unterschiedlichen Algorithmik.

Im Falle einer intelligenten Überwachung sind die Prozesse der Detektion, Lokalisation und Verfolgung von Objekten von herausragender Bedeutung [8], [18]. Eine schnelle und zuverlässige Bestimmung der Positionen von Zielobjekten eröffnet dabei weitreichende Möglichkeiten, wie die Partitionierung eines Zielgebietes in Bereiche unterschiedlicher Bedeutung und einer darauf operierenden Präsenzanalyse oder der Untersuchung von durch Zielobjekte erzeugten Trajektorien, um nur zwei Beispiele zu nennen. Weiterhin erlaubt eine echtzeitfähige Operation eine direkte Interaktion, z. B. durch Gesten oder spezifischen Bewegungsmuster. Speziell die Möglichkeit einer auf Basis von Bewegungsmustern stattfindenden Computer-Mensch-Kommunikation eröffnet dabei neue Ansätze<sup>1</sup> und ist ein Feld aktiver Forschung. Beachtet werden muss jedoch, dass eine bewegungsbasierte Szeneninterpretation bzw. Kommunikation gewisse Anforderungen an die eingesetzte Sensorik stellt. Es muss sichergestellt sein, dass diese der Dynamik der betrachteten Objekte Rechnung tragen und impliziert die Forderung nach einer entsprechend feinen Zeitauflösung bzw. ausreichend hohen Bildwiederholraten.

Im Allgemeinen wird die Topologie eines Überwachungssystems durch das zugrunde liegende Anwendungsgebiet definiert. Während für relativ einfache Aufgaben, wie der simplen Objektpräsenz in einem eng beschränkten Gebiet, z. B. einem Geldautomaten oder Autoinnenraum, unter Umständen bereits eine einfache, isolierte Sensorik ausreichende Ergebnisse liefert, trifft das

---

<sup>1</sup>Beispielhaft hierfür sind die durch Bewegungsmuster steuerbaren Spiele und Anwendungen [204], [205].

## 1 Einführung

---

für Anwendungen, in denen multiple Objekte wahrgenommen, verfolgt und deren Bewegung interpretiert werden müssen, nur begrenzt zu. Aufgrund des limitierten Detektionsbereichs der Sensorik bedarf es der Hinzunahme zusätzlicher informationsliefernder Einheiten. Durch die Erhöhung der Anzahl kooperierender Systeme ergibt sich unter der Annahme teilweiser überlappender Sichtbereiche naturbedingt eine Erweiterung des Detektionsbereichs, wobei durch den Grad der Überlappung die Redundanz der Informationen gesteuert werden kann. Aufgrund der erhöhten Informationsmenge in den sich überdeckenden Bereichen ergibt sich zusätzlich das Potenzial einer gesteigerten bzw. robusteren Wahrnehmungsleistung.

Ein weiterer Nutzen, der sich durch die Verwendung verteilter Sensorik mit überlappendem Sichtbereich ergibt, ist die Reduktion der Ambiguität. Insbesondere bietet die Analyse synchronisierter, aus unterschiedlichen Blickwinkeln erzeugter Positionsinformationen die Möglichkeit einer verbesserten, d. h. über einen längeren Zeitraum stattfindenden, simultanen Verfolgung multipler Objekte bzw. Personen.

Im Hinblick auf die verwendete Sensorik sind zurzeit konventionelle Kamerasysteme (CCD<sup>2</sup>- oder CMOS<sup>3</sup>-basiert) dominierend. Nicht zuletzt aufgrund der einfachen Verfügbarkeit, der geringen Baugröße, der unterschiedlichen Leistungsfähigkeit und Auflösung, der Vielzahl von vorhandenen Schnittstellen sowie des Preises, sind sie als informationsliefernde Einheiten attraktiv. Durch Kombination zweier Einheiten zu einem Stereo-Vision-System lassen sich unter Verwendung stereoskopischer Verfahren [18] hoch aufgelöste Tiefeninformationen gewinnen, wobei diverse Fehlerquellen (z. B. Parallaxefehler, Linsenverzerrungen, zeitlicher Versatz der Teilbilder), zu beachten sind. Aufgrund der relativ hohen Datenrate ist die Prozessierung der Daten rechenintensiv und führt besonders im Rahmen einer echtzeitfähigen Operation bei gleichzeitiger hoher Zeitauflösung zu enormen Anforderungen an die Hardware. Im Falle der Beobachtung dynamischer Objekte bedarf es zusätzlich einer strengen zeitlichen Synchronisation der Daten bzw. Bilder. Ein zusätzliches Manko solcher Stereo-Vision-Systeme ist dabei die grundsätzliche Abhängigkeit von Texturinformationen, welche zur Berechnung von Tiefeninformationen benötigt werden, was zu einer starken Abhängigkeit von dem Umgebungslicht führt und somit ihren Einsatz begrenzt.

Bedingt durch das aktive Messprinzip ist der Einsatz von auf Pulslaufzeit basierenden LIDAR<sup>4</sup>-Systemen (Laser-Range-Finder) auch unter fordernden bzw. kritischen Lichtverhältnissen durchführbar. Des Weiteren erlauben sie die Gewinnung von relativ genauen Tiefeninformationen über einem Distanzbereich von bis zu mehreren Hundert Metern. Aufgrund des Messprinzips muss das Detektionsgebiet jedoch einzeln abgetastet werden, was im Falle einer in horizontalen und vertikalen Richtung ausgedehnten Betrachtung der Szene zu einer relativ geringen Bildwiederholfrequenz führt. Weiterhin sind die Abmessungen und der nicht zu unterschätzende Geräuschpegel während der Operation sowie der relativ hohe Preis für schnelle Systeme Eigenschaften, welche den Einsatz von Laser-Range-Findern in intelligenten Systemen zur Überwachung von eventuell bewohnten Innenräumen als nicht optimal erscheinen lassen.

Die relativ neue Sensorklasse der auf Pulslaufzeitverfahren basierten ToF<sup>5</sup>-Kameras erlaubt eine simultane Akquisition von Tiefeninformationen der komplett betrachteten Szene, wobei Bildwiederholraten erreicht werden, die eine Lokalisation und Verfolgung von sich bewegenden Objekten erlaubt. Analog zum Laser-Range-Finder benutzen sie dabei eine Beleuchtung. Um auf Basis von unterschiedlichen Laufzeiten Distanzinformation zu gewinnen, benötigen sie jedoch hierfür eine komplette Ausleuchtung des zu betrachtenden Gebietes. Bedingt durch die limitierte Emissionsleistung der verwendeten Lichtquellen ist die Einsetzbarkeit dieser Systeme im Falle der Notwendigkeit hoher Bildwiederholraten zurzeit auf einige wenige Meter beschränkt<sup>6</sup>. Des Weiteren sind ToF-Kameras aufgrund ihrer Empfindlichkeit gegenüber dem Sonnenlicht für Outdoor-Anwendungen, jedenfalls zum momentanen Zeitpunkt, nur eingeschränkt geeignet. Durch

---

<sup>2</sup>CCD - *Charge Coupled Device*.

<sup>3</sup>CMOS - *Complementary Metal Oxide Semiconductor*.

<sup>4</sup>LIDAR - *LIght Detection And Ranging*.

<sup>5</sup>ToF - *Time of Flight*.

<sup>6</sup>Im Wesentlichen sind die mit diesen ToF-Systemen betrachteten Distanzen auf einen Bereich bis etwa 5m beschränkt, was nicht zuletzt der fehlenden Beleuchtungsintensität geschuldet ist [27].

---

## 1 Einführung

die zusätzliche Sonnenbestrahlung im Wellenlängenbereich der Beleuchtung<sup>7</sup> und der sich damit einstellenden Überlagerung ergeben sich Distanzwerte mit geringer Zuverlässigkeit, welche nur eingeschränkt verwendbar sind [62], [188]. Neben der geringen Reichweite stellt die niedrige laterale Auflösung eine weitere Beschränkung dar. Während neuere ToF-Systeme maximale Auflösungen von lediglich einigen Tausend Pixel bieten<sup>8</sup>, existieren konventionelle Kameras mit Auflösungen im Megapixel-Bereich.

Eine im Rahmen des Projektes POMSe<sup>9</sup> [217] durchgeführte Bewertung unterschiedlicher Sensorprinzipien in Hinblick auf ihre Eignung zur Personendetektion unter Beachtung verschiedener menschlicher Signaturen (Abb. 1) ergab, dass speziell in Bezug auf eine Lokalisation und Klassifikation von Objekten im Nahbereich von bis zu 10m eine Kombination aus konventionellen und ToF-Kameras als besonders geeignet erschien [218].

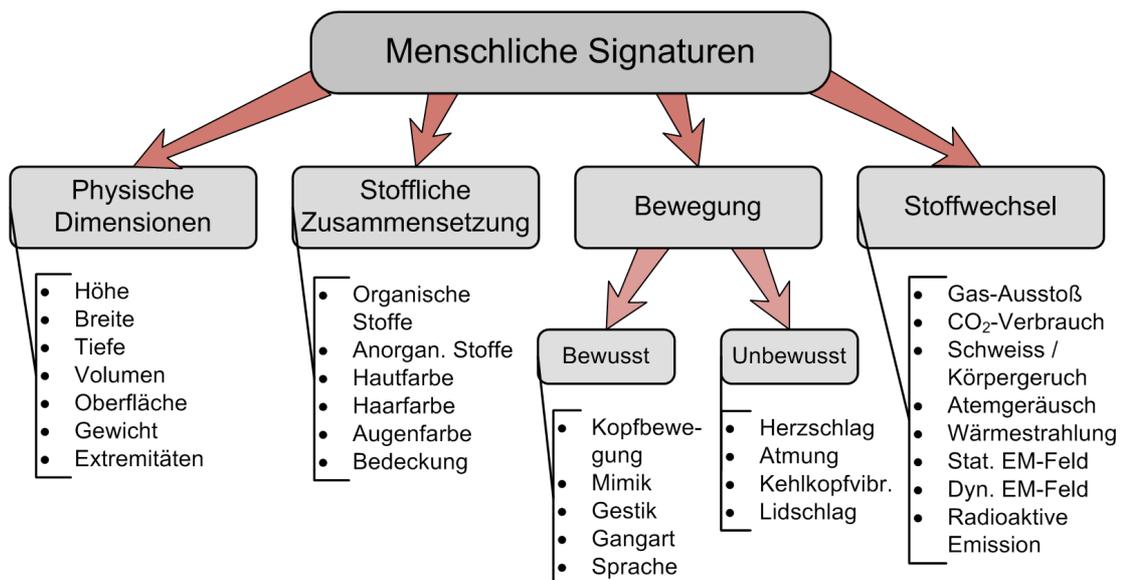


Abbildung 1: Klassifizierung menschlicher Signaturen.

Nicht überraschend ist daher der in den letzten Jahren verstärkt verfolgte Ansatz der Zusammenfügung von Systemen zur simultanen Akquisition von Intensitäts- und Tiefeninformationen zu 2D/3D-Vision-Systemen. Eingesetzt worden sind Kombinationen aus Laser-Range-Finder und hoch aufgelösten 2D-Kameras [76], [77], [80], [81], Laser-Range-Finder und Stereo-Vision-Systemen [12]-[14], ToF-Kameras und Stereo-Vision-Systemen [9]-[11] sowie ToF-Kameras mit hoch aufgelösten 2D-Kameras [63]-[66], [179]. Der für diese Arbeit interessante Fall eines aus ToF- und hoch aufgelöster 2D-Kamera zusammengesetzten binokularen 2D/3D-Systems wird im Hinblick auf mögliche Fehlerquellen in [62] ausführlich diskutiert. Eine Diskussion der grundsätzlichen Problematik, welche sich durch die Konsolidierung obig genannter Sensorsysteme zu einem 2D/3D-System ergibt, findet sich in [188].

Aufgrund der in [62] und [188] getroffenen Aussagen kann geschlussfolgert werden, dass eine Zusammenführung unterschiedlicher Sensorsysteme zur Erhöhung der Informationsdichte durch Fusion der Daten auf Pixelebene bzw. der Informationen auf Merkmal- oder Entscheidungsebene nicht trivial ist. Im Allgemeinen bedarf es gewisser Voraussetzungen, um eine erfolgreiche Konsolidierung der Teilinformationen der partizipierenden Systeme zu ermöglichen. Neben einer notwendigen Registrierung der verschiedenen Daten bzw. Bilder ist eine zeitliche Synchronisation für die Betrachtung dynamischer Objekte von essenzieller Bedeutung. Komplexe Prozesse zur Datenangleichung führen insbesondere in Umgebungen mit fordernden Rahmenbedingungen<sup>10</sup>

<sup>7</sup>Die Wellenlänge der verwendeten Beleuchtung liegt im nahen Infrarotbereich um 870 nm.

<sup>8</sup>Z. B. PMD-3k-S (64x48 Pixel) [222], PMD-41k-S (204x204 Pixel) [67], SwissRanger 4000 (176x144 Pixel) [78].

<sup>9</sup>POMSe: Detektion und Lokalisation von Personen und Objekten unter Einsatz mobiler Sensorik. Eine kompakte Beschreibung des Projekts findet sich in Abschnitt 3.3.

<sup>10</sup>Z. B. Echtzeitfähigkeit, hohe Zeitauflösung, Skalierbarkeit sowie robuste Operation.

# 1 Einführung

---

dazu, dass der Einsatz einer Vielzahl zusammengesetzter 2D/3D-Sensorsysteme nicht oder nur eingeschränkt möglich ist.

Die im Rahmen des Dyn-3D Projektes am ZESS entwickelte MultiCam ermöglicht aufgrund ihrer Konzeption eine effektive Akquisition von Intensitäts- und Tiefendaten. Die immense Vereinfachung der Registrierung sowie Synchronisation der Informationen<sup>11</sup> der verschiedenen Modalitäten erlaubt eine rechentechnisch effektive Fusion der Informationen auf Pixel-, Merkmal- und Entscheidungsebene [188], [62] und hat den Einsatz dieses zur Zeit einmaligen Systems im Hinblick auf eine Lokalisation und Verfolgung von multiplen Personen und Objekten in Innenräumen motiviert. Bedingt durch die unkomplizierte Verfügbarkeit einer hohen Informationsdichte besitzen Sensorsysteme wie die MultiCam das Potenzial, die vorhandenen Ansätze zur Implementierung intelligenter Räume zu revolutionieren.

Die Zielsetzung dieser Arbeit ist demnach die räumliche Lokalisation und Verfolgung multipler Objekte in Innenräumen sowie die Klassifikation dieser in Personen- und Objekt-Klasse unter Verwendung mehrerer im Raum verteilter, kooperierender MultiCams. Im Hinblick auf die Verwendung im Umfeld einer intelligenten Raumüberwachung sind dabei folgende Forderungen beachtet worden:

1. **Echtzeitfähige Operation:** Sensorinformationen werden zur Laufzeit direkt bearbeitet. Es findet keine Prozessierung der Informationen im Nachhinein (offline) statt. Die zeitliche Auflösung des Systems ist ausreichend, um eine Verfolgung von sich bewegenden Objekten bzw. Personen in Innenräumen zu gewährleisten.
2. **Einfache Skalierbarkeit:** Möglichkeit einer unbeschränkten Hinzunahme zusätzlicher informationsliefernder Einheiten ohne aufwendige Rekonfiguration des Gesamtsystems.
3. **Graceful Degradation:** Grundfunktionalität wie Lokalisation und Verfolgung von Objekten auch unter fordernden Lichtverhältnissen, was insbesondere das Fehlen von Umgebungslicht mit einschließt.
4. **Funktionale Erweiterbarkeit:** Hinzunahme von zusätzlicher bzw. veränderter Algorithmik und Sensorik ohne aufwendige Veränderungen aller am Prozess beteiligten Komponenten oder gar der Restrukturierung des kompletten Systems.

## 1.2 ANSATZ

Basierend auf der Sensorik der MultiCam bietet der verfolgte Ansatz die Möglichkeit einer Lokalisation von Objekten auch unter geringem oder fehlendem Umgebungslicht, was im Vergleich zu konventionellen Kameras einen massiven Vorteil darstellt. Konzeptionsbedingt ist die MultiCam in der Lage, kombinierte 2D/3D-Bilder mit einer Bildwiederholrate von bis zu 80 fps<sup>12</sup> zu liefern und eignet sich somit für eine Verfolgung dynamischer Objekte in Innenräumen [188]. Aufgrund des betrachteten Distanzintervalls von 1.5 m bis 9 m sowie eines Objektivs mit einer Brennweite von 6 mm und einem Öffnungswinkel<sup>13</sup> von horizontal 56° und vertikal 44° ist der durch eine MultiCam betrachtete Raumbereich ausreichend groß, um bereits mit wenigen Einheiten unter Verwendung sich partiell überlappender Sichtbereiche eine Detektion und Verfolgung mehrerer gleichzeitig vorhandener Objekte von Personengröße zu gewährleisten. Durch die Wahl der Modulationsfrequenz der Beleuchtung von ca. 15 MHz<sup>14</sup> ergibt sich eine Erweiterung des Eindeutigkeitsbereichs der Distanzdaten auf etwa 11.25 m und ermöglicht eine Lokalisation von Objekten in dem betrachteten Distanzintervall ohne die Anwendung von Phase-Unwrapping-Techniken [15].

---

<sup>11</sup>Detaillierte Angaben zur MultiCam erfolgen in Kapitel 2.

<sup>12</sup>fps - *frames per second*.

<sup>13</sup>Chipgröße des verwendeten ToF-Sensors (PMD 3k-S): 1/2“ [222].

<sup>14</sup>Die gleichzeitige Verwendung mehrerer, sich gegenseitig wahrnehmender MultiCams erfordert, dass für jede Einheit, zwecks Vermeidung gegenseitiger Beeinflussungen, eine eindeutige Modulationsfrequenz gewählt werden muss. Eine erweiterte Diskussion dieses Punktes erfolgt ebenfalls in Kapitel 2.

---

# 1 Einführung

Der für die Akquisition eines 2D/3D-Bildes notwendige Zeitraum  $t^{Akq}$  ist naturgemäß abhängig von der vorhandenen Intensität der Beleuchtung, der zu überwachenden maximalen Distanz sowie des vorhandenen Umgebungslichts. Aufgrund der Verwendung von ebenfalls am ZESS entwickelten Beleuchtungseinheiten mit ausreichender Bestrahlungsleistung kann für den betrachteten Distanzbereich die strikte Obergrenze der Aufnahmezeit kompletter 2D/3D-Bilder auf  $t_{Max}^{Akq} = 50$  ms gesetzt werden<sup>15</sup>, wobei dieses variierende Lichtverhältnisse mit einschließt. Die Festlegung dieser Grenze ist insbesondere für eine echtzeitfähige Operation von Belang, weil durch sie eine von der Sensorik limitierte zeitliche Auflösung des Systems definiert werden kann, welche die Verwendbarkeit im Hinblick auf eine Objektverfolgung bestimmt.

Ein zusätzlicher die Operationsgeschwindigkeit beeinflussender Faktor ist die zur Bearbeitung der Daten eines Aufnahmezeitpunkts  $t_n$  notwendige Zeitspanne  $t_n^{Proc}$ ,  $n \in \mathbb{N}$ . Anstrebenswert ist die Erfüllung der Forderung  $t_n^{Proc} < t_{Max}^{Akq} \forall n$ , da damit gewährleistet werden kann, dass die erreichbare Zeitauflösung lediglich durch die eingesetzte Hardware limitiert wird. Vorausgesetzt wird hierbei, dass die Akquisition sowie Prozessierung der Daten parallelisierbar ist. Eine solche Ausführung ist unter der Annahme der gegenseitigen Unabhängigkeit der zu verschiedenen Zeitpunkten akquirierten Daten jedoch in einfacher Weise möglich, indem die Prozessierung der Daten auf einem Datensatz des vorangegangenen Zeitpunktes verlagert wird (Abb. 2)<sup>16</sup>.

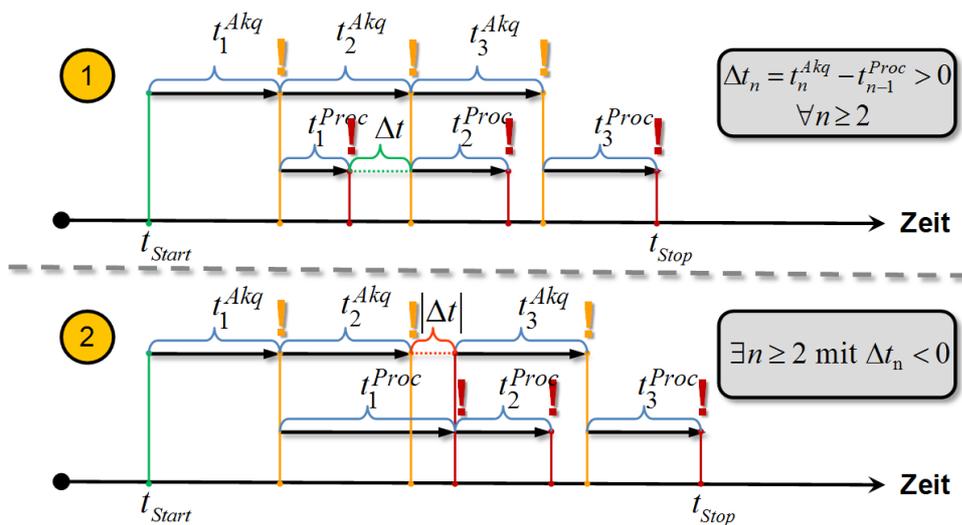


Abbildung 2: Zeitablauf der Akquisition und Prozessierung der Informationen.

Angenommen wurde hierbei der Fall, dass jede Bildakquisition den maximalen Zeitbedarf beansprucht ( $t_n^{Akq} = t_{Max}^{Akq}$ ). Die mit Ausrufezeichen markierten Stellen entsprechen den Zeitpunkten, an denen der Akquisitions- bzw. Bearbeitungsprozess seine Terminierung signalisiert und ermöglicht dem jeweils anderen die Ausführung. Der unerwünschte Effekt von Bearbeitungsperioden mit  $t_{n-1}^{Proc} > t_{max}^{Akq} \rightarrow \Delta t_n < 0$  wird anhand des unteren Beispiels ersichtlich. Aufgrund der gegenseitigen Abhängigkeit von Akquisition und Prozessierung wird in diesem Falle die temporale Auflösung zusätzlich von der zugrunde liegenden Algorithmik mitbestimmt, eventuell sogar dominiert.

Die Prozesse der Datenakquisition bzw. der Datenbearbeitung eines Zeitpunkts  $t_n$  werden dabei in Abhängigkeit des Vorhandenseins eines entsprechenden Signals des jeweilig parallelen

<sup>15</sup>Diese Zeitperiode schließt sowohl sämtliche Integrationszeiten der Sensoren als auch den Transport der Daten von der MultiCam zu dem die Informationen verarbeitenden Rechner mit ein.

<sup>16</sup>Die im Rahmen dieser Arbeit verwendeten Terme  $\forall$  (Allquantor) und  $\exists$  (Existenzquantor) entsprechen Operatoren der Prädikatenlogik.

# 1 Einführung

Prozessstranges gestartet, was das Auseinanderdriften beider Prozesse verhindert. Bedingt durch die zeitliche Auflösung der MultiCam hat diese Verlagerung keine wesentlichen Auswirkungen auf die grundsätzliche Systemleistung. Die erzeugten Informationen besitzen lediglich einen durch  $t_{Max}^{Akq}$  vorgegebenen Zeitversatz.

Für den Fall der Verwendung einer verteilter Sensorik ist der Einfluss der Hinzunahme zusätzlicher Daten liefernder Systeme auf  $t_n^{Proc}$  entscheidend. Eine schlechte Skalierung der zur Informationsgewinnung notwendigen Berechnungsdauer führt dazu, dass bereits eine geringe Anzahl informationsliefernder Systeme die Zeitauflösung des Gesamtsystems verringert und die Fähigkeit der Verfolgung dynamischer Objekte reduziert. Die Wahl der zugrunde liegenden Topologie zur Akquisition und Verarbeitung der Daten kann unter Umständen der entscheidende Faktor sein, welcher einen sinngemäßen Einsatz eines Gesamtsystems ermöglicht bzw. verhindert.

Im zentralisierten Ansatz werden sämtliche von den Sensoren akquirierten Daten direkt an eine zentrale Instanz gesendet, welche für den Empfang und die Weiterverarbeitung zuständig ist. Die Notwendigkeit einer zusätzlichen Infrastruktur zwecks Kommunikation und Synchronisation entfällt aufgrund der direkten Verbindung zwischen Sensorik und Datensinke<sup>17</sup>. Der Vorteil dieses Ansatzes ist darin zu sehen, dass durch die schnelle Implementierbarkeit die Möglichkeit besteht, Ergebnisse relativ rasch produzieren zu können. Beispiele des zentralisierten Ansatzes zur innenräumlichen Personendetektion unter Verwendung multipler Sensorik und überlappender Sichtbereiche finden sich u. a. in [17], [19]-[21], [23] sowie [133]. Erwähnenswert ist das im Rahmen des 7. EU-Forschungsrahmenprogramms durchgeführte SAFROS-Projekt [7], welches als Zielsetzung die Erhöhung der Patientensicherheit im Kontext einer roboterassistierten Operation verfolgt. Zur Überwachung des Patienten werden diesbezüglich multiple, simultan operierende ToF-Kameras<sup>18</sup> mit gegenseitig partiell überlappendem Sichtbereich eingesetzt, welche ihre Daten zentral an einen dedizierten Rechner weiterreichen und welchem die Aufgabe zukommt, eine entsprechende Situations- bzw. Gefahrenanalyse durchzuführen. Aufgrund der Limitierung<sup>19</sup> der eingesetzten Sensorik kann jedoch nur eine Untermenge der vorhandenen Kameras dazu benutzt werden, Daten der betrachteten Szene simultan zu akquirieren.

Die Nachteile, welche sich durch eine zentralisierte Topologie ergeben, sind jedoch nicht unwesentlich:

1. **Beschränkte Skalierbarkeit:** Durch die direkte Verbindung zwischen Datenquellen und der zentralen Datensinke ergibt sich zunächst die Notwendigkeit, eine gewisse Anzahl der entsprechenden Schnittstellen zur Verfügung stellen zu müssen. Auch durch Verwendung einer skalierbaren Verbindungsstruktur, z. B. ein auf TCP/IP, USB oder FireWire basiertes lokales Netz, stellt keine hinreichende Lösung dar, da im Falle von Sensorsystemen mit hoher Datenrate eine Limitierung der verwendbaren Einheiten aufgrund der maximalen Bandbreite bereits von Anfang an festgelegt ist.
2. **Nichttriviale Erweiterbarkeit:** Eine im Nachhinein durchzuführende funktionale Erweiterung kann sich als kompliziert bis nicht durchführbar erweisen. Änderungen müssen an dem kompletten System durchgeführt werden, was in der Regel zu einem vollständigen Ausfall der Operation führt.
3. **Hohe Rechenleistung:** Die Verarbeitung aller pro Zeiteinheit anfallenden Daten auf einer zentralen Einheit führt zu einer hohen Anforderung an die bereitzustellende Rechenleistung und ist ein weiterer limitierender Faktor im Bezug auf die Anzahl simultan verwendbarer Einheiten. Insbesondere ist die zeitnahe Behandlung der Daten mehrerer parallel arbeitenden Sensoren mit hoher Datendichte (z. B. 2D-Sensoren) eine Hürde im Hinblick auf eine akzeptable Bearbeitungsgeschwindigkeit und führt leicht zu  $t_n^{Proc} \gg t_{Max}^{Akq}$ .

<sup>17</sup>Datensinke bezeichnet hierbei eine Einheit, welche für die Aufnahme der von der Sensorik generierten Information zuständig ist.

<sup>18</sup>Die eingesetzten ToF-Kameras entstammen der momentanen Produktlinie von PMDtec [67].

<sup>19</sup>Die simultane Operation mehrerer ToF-Kameras des verwendeten Typs erfordert eine für jede Kamera unterschiedliche Modulationsfrequenz der Beleuchtung. Konzeptionsbedingt bieten die eingesetzten Modelle jeweils nur drei verschiedene Einstellungen an, sodass für eine simultane Operation mit sich gegenseitig überlappendem Sichtbereich ebenfalls maximal drei Kameras eingesetzt werden können.

## 1 Einführung

---

Eine alternative Topologie ergibt sich durch die Verwendung einer Client/Server-Infrastruktur. Die Grundidee ist, die Berechnung auf Basis der Sensorrohdaten auf die Server zu verlagern, welche dann im Anschluss ihre Informationen in kompakterer Form dem zentralen Client zur Verfügung stellen. Bedingt durch die systeminhärente Parallelität ergibt sich eine erhebliche Reduktion der zur Berechnung von Informationen notwendigen Zeit. Vorausgesetzt wird hierbei, dass die verwendeten Server über eine ausreichende Intelligenz verfügen, um die anfallenden Rohdaten sinngemäß zu bearbeiten. Der immense Vorteil, neben der bereits erwähnten Verringerung notwendiger Rechenleistung, liegt in der hohen Skalierbarkeit. Zusätzliche informationsliefernde Einheiten lassen sich durch die Hinzunahme von weiteren Servern in das System integrieren. Beispiele einer Client/Server-Topologie finden sich u. a. in [16], [18], [22] sowie [24]-[26].

Im Vergleich zum zentralistischen Ansatz bedarf es in dieser Topologievariante der Verwendung einer zugrunde liegenden Kommunikationsinfrastruktur. Als schnellste Lösung bietet sich zunächst die Möglichkeit einer proprietären Implementierung an. Das kann aber dazu führen, dass im Falle einer Konsolidierung mit fremden Systemen die verwendete Implementierung mittels einer Schnittstelle angepasst oder sogar neu implementiert werden muss. Einen Ausweg bietet der Einsatz einer standardisierten Infrastruktur wie SOAP oder CORBA [101]-[103], welche eine Interoperabilität zwischen unterschiedlichen Systemen ermöglicht. Gemäß [97] sind herkömmliche Client/Server-Ansätze jedoch im Bezug auf eine Erweiterung der Funktionalität nicht hinreichend flexibel, da sich die Interoperabilität lediglich auf die der zugrunde liegenden Applikation angepassten Infrastrukturebene beschränkt. Eine erweiterte Bereitstellung sowie Pflege der Dienste, auf denen Client und Server aufbauen, ist nicht trivial erfüllbar, da in der Regel die Basisfunktionalität samt notwendiger Erweiterungen erneut implementiert werden müssen.

Eine weitere Topologie ergibt sich durch Anwendung des Agentenparadigmas (Kapitel 3). Die Besonderheit dieses Ansatzes liegt in der Möglichkeit, die unterschiedlichen Funktionalitäten eines Systems durch entsprechende Agenten darzustellen, wodurch sich gewisse Vorteile ergeben:

1. **Hohe Skalierbarkeit:** Verteilte Systeme, die in der Regel eine systeminhärente Komplexität besitzen, lassen sich mit Hilfe von Agenten als Basiselemente relativ leicht modellieren bzw. beschreiben [84].
2. **Parallele Prozessierung:** Als eigenständig arbeitende Einheiten bieten Agenten die Möglichkeit einer parallelen Ausführung bzw. Abarbeitung der ihnen zur Verfügung gestellten Informationen.
3. **Einfache Erweiterbarkeit:** Implementiert als offene, kommunikationsbasierte sowie skalierbare modulare Systeme, bieten sie die Möglichkeit des flexiblen Einsatzes in unterschiedlichen Bereichen<sup>20</sup>. Beispiele einer Gebietsüberwachung unter Einsatz von Agenten finden sich in [27]-[35].

Bedingt durch die Verwendung des Agentenparadigmas kann der Prozess der kooperativen Lokalisation und Verfolgung durch die Definition unterschiedlicher Agenten als Basiselemente in einfacher Weise beschrieben werden. Der im Rahmen dieser Arbeit verfolgte Ansatz basiert auf der Verwendung zweier unterschiedlicher Agententypen. Während informationsliefernde Agenten die Aufgabe übernehmen, Informationen bezüglich ihrer Umwelt zu erzeugen, übernimmt der Fusionsagent die Zusammenführung sämtlicher vorhandener Informationen und fügt diese zu einer globalen Sicht zusammen. Zusätzlich ist Letzterer auch für die synchronisierte Akquisition der Informationen durch die am Prozess beteiligten informationsliefernden Agenten zuständig.

Das grundlegende Prinzip des Ansatzes liefert Abb. 3, wobei zwecks besserer Darlegbarkeit der Sachverhalte im Folgenden informationsliefernde Agenten als Multisensorknoten (MSK) und der fusionierende Agent als Fusionsknoten (FSK) bezeichnet werden.

---

<sup>20</sup>Z. B. Prozessmanagement [105], Luftraumüberwachung [106], Verarbeitung von Satellitendaten [107], E-Kommerz [108] sowie Computerspiele [109].

---

# 1 Einführung

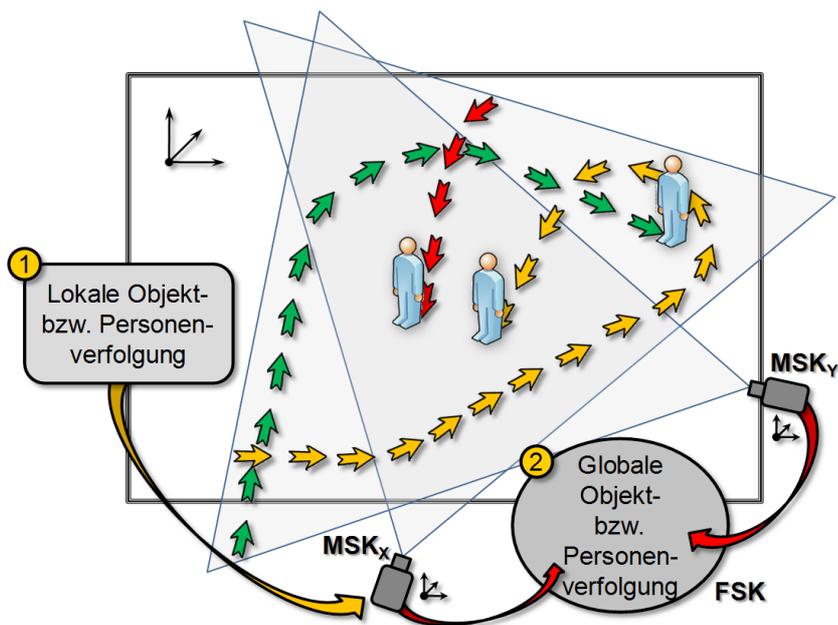


Abbildung 3: Prinzipieller Ansatz der verteilten Systemoperation.

Nach einer synchronisierten Akquisition der Bildinformationen durch die MSK folgt die lokale Objekt- bzw. Personenverfolgung im zugehörigen lokalen Koordinatensystem. Diese Informationen werden im Anschluss in Form eines lokalen Statusvektors an den FSK gesendet, welcher die Fusion der Informationen übernimmt und einen globalen Systemzustand auf Grundlage der verschiedenen Perspektiven in einem globalen Koordinatensystem berechnet.

Begonnen wird mit der durch den FSK synchronisierten Akquisition der Informationen durch die an unterschiedlichen Stellen im Raum positionierten MSK, welche diesbezüglich mit jeweils einer MultiCam ausgestattet sind. Angenommen wird dabei, dass sich durch die Ausrichtungen der MSK-MultiCams teilweise überlappende Sichtbereiche ergeben. Anschließend findet auf jedem MSK die parallele Prozessierung der jeweilig akquirierten Sensorinformationen statt. Aufgrund der unterschiedlichen Positionierung erfolgt die Behandlung der Informationen eines MSK auf Grundlage seines eigenen Koordinatensystems und wird entsprechend als lokale Operation interpretiert. Nach Beendigung der lokalen Prozessierung sendet jeder MSK seine Informationen in Form eines lokalen Zustandsvektors an den Fusionsknoten, welcher die Zusammenführung der vorhandenen Informationen auf Basis eines globalen Koordinatensystems übernimmt.

Eine getroffene Annahme dabei ist, dass die Beziehung zwischen dem globalen Koordinatensystem des FSK und den lokalen Koordinatensystemen der MSK bekannt ist<sup>21</sup> und keiner zeitlichen Änderung unterliegt. Eine detaillierte Beschreibung der Systemkomponenten, der Systemoperation sowie der Kommunikationsinfrastruktur erfolgt in Kapitel 3.

Einer der wesentlichen Vorteile von ToF-Systemen im Vergleich zu konventionellen CMOS- bzw. CCD-Kameras ist ihre bereits angesprochene operative Unabhängigkeit von dem Umgebungslicht. Lokalisation und Verfolgung auf Basis von Tiefeninformationen lassen sich somit auch unter extremen Lichtverhältnissen<sup>22</sup> durchführen und ermöglichen einen Einsatz in Bereichen, in denen auf konventionelle Kameratechnik aufbauende Algorithmen nicht oder nur sehr eingeschränkt einsetzbar sind.

<sup>21</sup>Die Bestimmung der Beziehung zwischen den unterschiedlichen Koordinatensystemen kann im Verlauf der Systemaufstellung erfolgen.

<sup>22</sup>Z. B. lichtschwache Umgebungen oder Bereiche mit stark variierender Intensität des Umgebungslichts.

## 1 Einführung

Während die Lokalisation und Verfolgung von Objekten auf Grundlage der durch den PMD-Chip der MultiCam erzeugten Tiefeninformationen gemäß eines Top-Down-Modells erfolgt (Abschnitt 4.2), basiert die Personenerkennung auf der Verwendung zweier, auf den 2D-Daten der MultiCam operierender Bottom-Up-Methoden (Abschnitt 4.3) unter Verwendung von auf Tiefeninformationen beruhendem Vorwissen (Abb. 4). Der verfolgte Operationsansatz zur lokalen Prozessierung der Informationen fußt somit auf einer kooperativen Fusion der 2D- und 3D-Daten der MultiCam (Abschnitt 3.2).

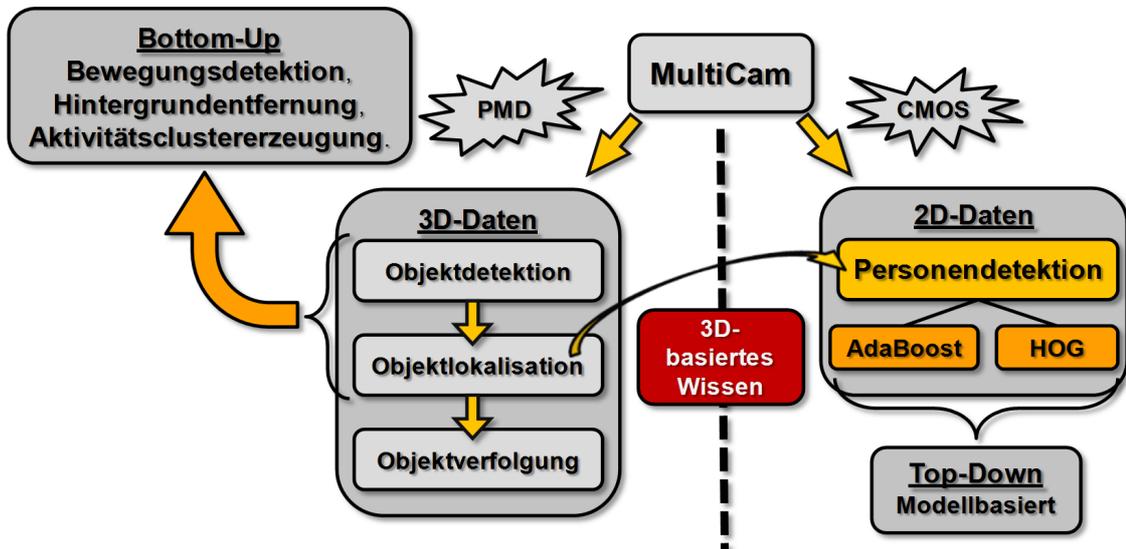


Abbildung 4: Ansatz zur Personen- und Objektverfolgung.

Motiviert wird der Ansatz dabei durch folgende Überlegungen:

1. Während die Lokalisation und Verfolgung von Objekten auf Basis der Distanzinformationen über den gesamten betrachteten Distanzbereich selbst unter extremen Lichtverhältnissen möglich ist, ist ihre Klassifikation aufgrund ihrer unterschiedlichen Repräsentation in Abhängigkeit ihrer Entfernung zur MultiCam sowie der relativ geringen lateralen Auflösung des verwendeten PMD-Chips nur eingeschränkt möglich (Abschnitt 4.2.4), [62], [188].
2. Durch den hohen Informationsgehalt der 2D-Daten bietet sich die Möglichkeit einer Klassifikation der lokalisierten Objekte auf dem kompletten Distanzbereich. Wie Untersuchungen der Geschwindigkeiten der verwendeten modellbasierten Top-Down-Methoden jedoch zeigen (Abschnitt 4.3), eignen sich diese Verfahren nur bedingt für eine echtzeitfähige Operation. Die Bearbeitungszeiten der Klassifikationsalgorithmen auf den kompletten 2D-Daten der MultiCam führen zwangsläufig zu dem unerwünschten Fall  $t_n^{Proc} \gg t_{Max}^{Akq}$ . Durch kooperative Fusion (Abschnitt 3.2) von 2D- und 3D-Daten unter Verwendung der durch die im Rahmen der Objektlokalisierung erzeugten Information kann eine Reduktion der für eine Klassifikation von Objekten notwendigen Datenmenge erreicht werden. Dieses führt dazu, dass die zur Berechnung notwendige Zeitspanne der eingesetzten Algorithmen wesentlich verringert wird (Abschnitt 4.3.3) und somit ihren Einsatz im Hinblick auf die Forderung  $t_n^{Proc} < t_{Max}^{Akq}$  ermöglicht.
3. Die Separation von Lokalisation und Verfolgung von Objekten auf Basis der PMD-Daten zum Einen und der Klassifikation derer auf Grundlage der hoch aufgelösten 2D-Daten zum Anderen erlaubt eine Restfunktionalität selbst unter Umgebungslichtverhältnissen, welche den Einsatz von visuellen Sensoren verhindern und entspricht der Forderung einer anmutigen Degradierung (*Graceful Degradation*).

# 1 Einführung

---

Als Grundlage der Implementierung und Analyse der Algorithmen wurden synchronisierte Videos mit einem zeitlichen Versatz der Bilder von im Mittel 80 ms<sup>23</sup> verwendet, welche mithilfe der veränderbaren Funktionalität des zugrunde liegenden Agentensystems (Abschnitt 3.1.5) erstellt worden sind<sup>24</sup>. Die Integrationszeit des PMD-Sensors von 8 ms wurde durch mehrere Messungen zu unterschiedlichen Zeitpunkten bestimmt und bot ein ausgewogenes Verhältnis zwischen erreichbarer Zeitauflösung und betrachtbarer Distanz.

Um eine Überbelichtung des 2D-Sensors zu verhindern, wurde seine Integrationszeit auf 7 ms gesetzt. Bedingt durch das noch zu diskutierende Verfahren der Tiefenbilderzeugung, welches auf der Akquisition von vier Phasenbildern basiert (Abschnitt 2.1.1), des zur Aufnahme von Intensitätsdaten notwendigen Zeitraums, sowie der zum Transport der Daten von der Kamera zum Rechner via USB 2.0 benötigten Zeit (ca. 12 ms), ergab sich die Festlegung von  $t_{Max}^{Akq}$  zu 50 ms, mit einer Zeitspanne von ca. 5 ms als Puffer. Die Diskrepanz zwischen zeitlicher Auflösung der Sensorik und Aufnahmegeschwindigkeit lässt sich damit begründen, dass zusätzlich zur Aufnahme eine anschließende Analyse, Visualisation und Speicherung der Daten durchgeführt wurde. Die Zusammensetzung eines Videobildes findet sich in Tab. 7, Anhang F. Die Gründe, welche den Einsatz von synchronisierten Videos motiviert haben, sind vielfältig: Zunächst bedarf es einer definierten, reproduzierbaren Basis, auf der eine Überprüfung der Algorithmik stattfinden kann. Des Weiteren bietet sie die Möglichkeit einer Verifikation der Ergebnisse. Zusätzlich erlauben sie eine Unabhängigkeit bzgl. des Aufbaus, da eine exakte Beibehaltung einer definierten Anordnung bestehend aus mehreren MSK über einen längeren Zeitraum nur in seltenen Fällen möglich ist<sup>25</sup>.

Wichtig zu erwähnen ist, dass bedingt durch die prototypische Natur der MultiCam die Anzahl der zur Akquisition der Daten verfügbaren MultiCam auf zwei Einheiten beschränkt gewesen ist. Der Grund für die Wahl einer seitlichen Sicht auf die Objekte im Detektionsbereich liegt zum Einen in der Maximierung des Überlappungsbereichs und zum Anderen an der für die klassifizierenden, intensitätsdatenbasierten 2D-Algorithmen notwendigen Perspektive auf die Objekte.

Aufnahmen der zur Videoakquisition verwendeten Umgebung sowie der Anordnung der beiden MSK finden sich in Abb. 6. Eine schematische Darstellung des Aufbaus zusammen mit den Positionen und Sichtbereichen<sup>26</sup> der Sensorknoten in einem globalen Koordinatensystem  $K^{Glob}$  enthält Abb. 5. Während für die Positionen der MSK die Mitte der Vorderfront der Linse der jeweiligen MultiCam angenommen wurde, ergaben sich die zugehörigen Referenzpunkte als die vom mittleren Pixel des hoch aufgelösten Intensitätsbildes der MultiCam abgedeckten Raumpunkte. Anzumerken ist noch, dass die Definition des globalen Koordinatensystems willkürlich ist und zu Beginn einer Systeminstallation einmalig erfolgen muss (Abb. 6).

Die betrachtete Umgebung stellt für die verwendete Sensorik eine Herausforderung in zweifacher Hinsicht dar.

1. Zunächst ergibt sich aufgrund der örtlichen Gegebenheiten eine inhomogene Verteilung des Umgebungslichts, welche durch die sich ändernde Richtung der Sonneneinstrahlung noch verstärkt wird. Für die auf Daten visueller Sensorik operierenden Algorithmen entsprechen diese Umstände einem schwierigen Umfeld.
2. Aufgrund der spiegelnden Eigenschaft des Untergrunds und der damit verbundenen Reflexion des eingestrahlt Lichts ergibt sich auch für den PMD-Sensor eine anspruchsvolle Arbeitsumgebung.

Speziell die zweite Anmerkung impliziert, dass die einzusetzende Algorithmik eine gewisse Robustheit gegenüber fordernden Rahmenbedingungen aufweisen muss.

---

<sup>23</sup>Eine genaue Angabe des Versatzes ist aufgrund des nichtdeterministischen Verhaltens der eingesetzten Betriebssysteme (Windows 7 und Windows XP) nicht möglich.

<sup>24</sup>Dazu sendete der FSK nach Aufnahmestart Signale in einem festgelegten Intervall von 80 ms an die beteiligten MSK, welche wiederum, nach Beendigung der Akquisition, die akquirierten Daten in entsprechende Dateien abspeicherten.

<sup>25</sup>Die verwendeten MultiCams wurden im ZESS durch mehrere Personen zu unterschiedlichen Zwecken und mit unterschiedlichen Beleuchtungskonfigurationen benutzt.

<sup>26</sup>Als Richtwert wurde hierbei der horizontale Öffnungswinkel der Optik von 56° verwendet.

---

# 1 Einführung

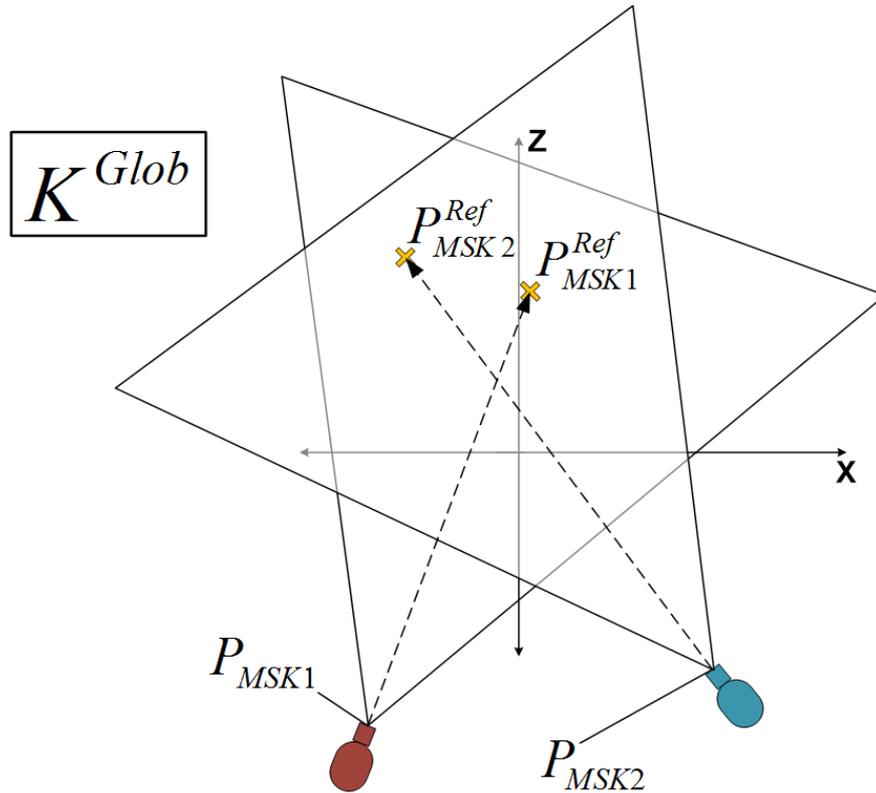


Abbildung 5: Positionen der MSK und der zugehörigen Referenzpunkte in der  $x-z$ -Ebene des globalen Koordinatensystems  $K^{Glob}$ .

Positionen:  $P_{MSK1} = (-229, 170, -429)$ ,  $P_{MSK2} = (298, 170, -334)$ ,  $P_{MSK1}^{Ref} = (22, 50, 258)$ ,  $P_{MSK2}^{Ref} = (-182, 10, 309)$ . Alle Angaben in cm.

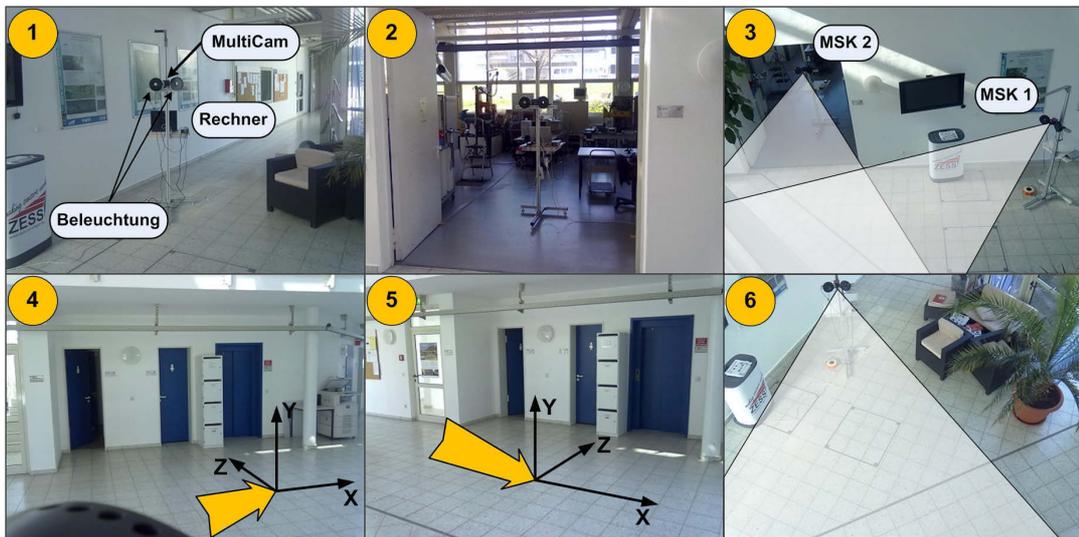


Abbildung 6: Verwendete Umgebung und Anordnung der beiden MSK.

Bild 1: Erste MultiCam mit dediziertem Laptop (MSK1) Bild 2: MSK2. Bild 3: Übersicht des aufgebauten Systems mit geschätzten Sichtbereichen sowie deren Überlappung. Bild 4: Ungefähre Sichtbereich der MultiCam von MSK1 (manuelle Aufnahme mittels Digitalkamera). Bild 5: Ungefähre Sichtbereich der MultiCam von MSK2. Bild 6: Geschätzter Sichtbereich von MSK1. Der in den Bildern 4 und 5 markierte Punkt repräsentiert den Ursprung des globalen Koordinatensystems.

### 1.3 BEITRAG DER ARBEIT

Als Beitrag dieser Arbeit ist die Lokalisation und Verfolgung von Objekten und Personen mittels eines verteilten, skalierbaren, funktional erweiterbaren und in Echtzeit arbeitenden, agentenbasierten Systems unter Verwendung einer bislang einzigartigen Kombination aus 2D- und 3D-Sensoren (MultiCam) zu nennen. Der durch jeweils eine MultiCam betrachtete Abdeckungsbereich ist dabei, bezogen auf seine Abmessungen und unter Beibehaltung einer für die Verfolgung von dynamischen Objekten ausreichenden zeitlichen Auflösung, zurzeit einmalig.

Aufgrund der Neuartigkeit der eingesetzten Sensorik sind im Vorfeld der Systementwicklung zunächst notwendige Werkzeuge zur Akquisition<sup>27</sup>, Evaluation sowie Darstellung der Daten der MultiCam<sup>28</sup> entwickelt und als grundlegende Werkzeuge im Rahmen der Implementierung der Algorithmik benutzt worden.

Das primär auf den PMD-Daten basierte Verfahren zur Hintergrundsubtraktion bietet die Möglichkeit der Adaption an sich verändernde Umgebungen<sup>29</sup>. Die Verlagerung des Gewichts auf die Informationen des PMD-Chips zur Erzeugung eines Hintergrundmodells ermöglicht eine vom Umgebungslicht weitgehend unabhängige Operation. Die optionale Hinzunahme der Informationen des 2D-Sensors erhöht dabei die Performanz in Hinblick auf die Detektion von Bewegungen. Aufgrund der geringen benötigten Rechenleistung ist das konzipierte Verfahren für den Einsatz in Systemen, welche eine hohe zeitliche Auflösung erfordern, geeignet.

Der im Rahmen dieser Arbeit entwickelte Lokalisationsalgorithmus basiert auf der Erzeugung sogenannter Aktivitätscluster und bietet die Möglichkeit der simultanen Lokalisierung multipler Objekte. Aufbauend auf den Ergebnissen der Hintergrundsubtraktion sowie den präprozessierten Daten der eingesetzten Sensorik sind Lokalisationen bis zu einer Distanz von 9 m bei relativ hohen Bildwiederholraten von 20 fps und variablen, entfernungsbedingten Repräsentationen der Objekte, welche eine der wesentlichen Herausforderungen darstellt, möglich. Aufgrund der im Vergleich zum 2D-Sensor niedrigen Auflösung ist der benötigte Rechenaufwand zur simultanen Lokalisation mehrerer Objekte gering und erlaubt ebenfalls eine echtzeitfähige Operation. Gleiches gilt für die lokal auf einem MSK durchgeführte Objektverfolgung.

Durch die Ausnutzung der durch die Objektlokalisierung erzeugten Informationen wird unter Verwendung kooperativer Fusion eine beschleunigte 2D-Daten-basierte Personendetektion ermöglicht. Unter Verwendung der 3D-Daten-basierten Positionsinformationen der Objekte wird eine Reduktion des Suchgebiets im zugehörigen 2D-Bild erreicht, was dazu führt, dass die zur Klassifizierung durch die Top-Down-Algorithmen notwendige Bearbeitungszeit verringert und deren Einsatz in Systemen, welche eine hohe zeitliche Auflösung benötigen, möglich wird.

Das implementierte, agentenbasierte Multisensorknotenkonzept erlaubt einen effektiven Einsatz einer Mehrzahl kooperierender MSK. Der im Rahmen dieser Arbeit verfolgte Ansatz der Fusion von lediglich Zustandsinformationen der MSK auf einem zentralen Fusionsknoten unter Verwendung von Partikelfiltern ist sowohl vom benötigten Rechenaufwand als auch von der benötigten Kommunikationsbandbreite her moderat und ermöglicht die Konsolidierung einer Vielzahl von simultan durch unterschiedliche Sensorknoten erzeugte Informationen. Aufgrund der geringen Zeitspanne zur Erzeugung sowohl der lokalen, MSK-basierten Informationen, als auch der zentral fusionierten Ergebnisse, ist der verfolgte Ansatz ebenfalls für die Lokalisation und Verfolgung multipler Objekte in Echtzeit geeignet.

Aufgrund des modularen Aufbaus ist das agentenbasierte Rahmenwerk weiterhin nicht auf einen Sensortyp beschränkt. Durch die Verwendung entsprechender Module zur Akquisition, Verarbeitung und Interpretation können unterschiedlich konfigurierte MSK kooperativ verwendet werden. Ein mehrfacher Einsatz des Rahmenwerks, unter Einsatz einer im Vergleich zu dieser Arbeit verschiedenen Sensorik, ist im Rahmen des vom BSI in Auftrag gegebenen Projektes P0mSe [218], [219] erfolgreich durchgeführt worden.

---

<sup>27</sup>Dazu zählen ein MultiCam-Treiber (Windows) sowie eine MultiCam-API (*Application Programming Interface*, Windows).

<sup>28</sup>Entwickelt wurde diesbezüglich ein grafisches Benutzerinterface (GUI – *Graphical User Interface*, Windows).

<sup>29</sup>Verwendet wird die Methode des laufenden Mittelwerts sowie des GMM (*Gaussian Mixture Model*).

---

### 1.4 GLIEDERUNG

Kapitel 2 dient der Vorstellung der verwendeten Sensorik. Der anfänglichen Diskussion der zugrunde liegenden Methodik zur Distanzbestimmung folgt eine Beschreibung der MultiCam. Beleuchtet werden die einzelnen Komponenten des Aufbaus sowie die implementierte Funktionalität. Zusätzlich werden die zur Verwendung der Distanzinformationen notwendigen Korrekturen der Rohdaten des Sensorsystems diskutiert sowie die zur simultanen Operation mehrerer MultiCams unter Verwendung eines überlappenden Sichtbereichs notwendigen Bedingungen besprochen.

Kapitel 3 widmet sich der Diskussion der Multisensorknoten sowie der Datenfusion. Im ersten Teil erfolgt zunächst die Beschreibung der agentenbasierten Architektur sowie die Beschreibung der entsprechenden Implementierung. Der zweite Teil dient der Vorstellung des verfolgten Fusionsansatzes sowie des verwendeten Fusionsprozessmodells. Basierend auf dessen Komponenten wird der Informationsfluss des implementierten Systems vorgestellt.

Kapitel 4 behandelt die auf einem einzelnen Multisensorknoten eingesetzte Algorithmik zur Lokalisation und Klassifikation von Objekten. Die im Rahmen dieser Arbeit verwendete Form zur Beschreibung estimationstheoretischer Vorgänge basiert dabei auf der in [180] verwendeten Notation. Begonnen wird mit der Vorstellung der zur Lokalisation von Objekten entwickelten Algorithmen. Diskutiert wird die auf den 2D/3D-Daten-basierte Bewegungsdetektion, die Subtraktion des Hintergrunds sowie die Erzeugung der Repräsentationen verfolgbarer Objekte auf Basis von Aktivitätsclustern.

Der letzte Teil von Kapitel 4 widmet sich der Personendetektion unter Verwendung zweier auf Intensitätsdaten operierender State-of-the-Art-Algorithmen. Durch die Verfügbarkeit von Tiefeninformationen bietet sich die Möglichkeit der Definition von Distanzbereichen, in denen die verwendeten Algorithmen eine optimale Performanz auf den 2D-Daten versprechen. Zusätzlich wird gezeigt, wie unter Einsatz kooperativer Datenfusion der 2D- und 3D-Daten mittels Verwendung der durch die Lokalisation erzeugten Informationen eine beschleunigte Ausführung der klassifizierenden Algorithmen erreicht wird.

Kapitel 5 dient der Diskussion der Objekt- bzw. Personenverfolgung. Der erste Teil enthält die Beschreibung und Analyse der lokal auf einem MSK stattfindenden Verfolgung von Objekten und umfasst ebenfalls die Vorstellung der Methodik der lokalen Objektverwaltung.

Im zweiten Teil von Kapitel 5 erfolgt die Erläuterung der kooperativen Objekt- bzw. Personenverfolgung auf einem FSK. Nach kurzer Diskussion der Verwaltung globaler Objekte auf Basis der Informationen der am Prozess beteiligten MSK schließt sich die Beschreibung der globalen Objektverfolgung an. Im Anschluss dazu werden sowohl die durch Fusion der MSK-Informationen erreichten Steigerungen der Lokalisations- und Verfolgungsfähigkeit als auch die Geschwindigkeit des Gesamtsystems analysiert.

Kapitel 6 dient der Diskussion der Ergebnisse, wobei die Vorteile des im Rahmen dieser Arbeit verfolgten Ansatzes beleuchtet werden. Beendet wird dieses Kapitel mit einer Betrachtung problematischer Aspekte sowie möglicher Erweiterungen.



## 2 SENSORIK

Das Hauptelement der Datenakquisition ist die am Zentrum für Sensorsysteme ZESS entwickelte MultiCam, ein monokulares 2D/3D-Kamerasystem basierend auf einer Kombination aus CMOS-Farbchip der Firma Micron vom Typ MT9V403C12 mit VGA Auflösung (640x480 Pixel) [79] sowie einem entfernungsmessenden Sensor der Firma PMD-Tec vom Typ PMD-3k-S [67] mit einer Auflösung von 64x48 Pixel. Der Beschreibung des Funktionsprinzips zur Erzeugung von Distanzdaten sowie dem Aufbau der MultiCam sind die folgenden Abschnitte gewidmet.

In Abschnitt 2.1 wird zunächst die Methodik der Distanzdatengewinnung diskutiert.

Abschnitt 2.2 dient der Beschreibung des prinzipiellen Aufbaus der MultiCam. Zusätzlich wird auf den simultanen Betrieb mehrerer MultiCam-Systeme eingegangen, wobei die notwendigen Bedingungen für eine störungsfreie Datenakquisition präzisiert werden.

Auf eine Erläuterung der Funktionalität konventioneller CCD- oder CMOS-Chips wird im Rahmen dieser Arbeit verzichtet. Eine diesbezügliche Diskussion, sowie eine Übersicht über optisch entfernungsmessender Verfahren, kann z. B. gefunden werden in [37], [40], [42], [44], [45], [53], [55], [57], [61] sowie [71].

### 2.1 CW-MODULATIONSINTERFEROMETRIE

Grundsätzlich basiert die von der MultiCam durchgeführte Entfernungsmessung auf dem Prinzip der Modulationsinterferometrie unter Verwendung einer permanent modulierten Beleuchtung<sup>30</sup>. Das emittierte inkohärente Licht entstammt mit einer Wellenlänge von  $\lambda=870\text{ nm}$  aus dem Spektrum des nahen Infrarotbereichs.

Die gewählte Modulationsfrequenz  $f_{Mod}$  ist dabei um Größenordnungen geringer als die Frequenz des verwendeten Lichts. Im Falle des Einsatzes einer Modulationsfrequenz aus dem Hochfrequenzbereich<sup>31</sup> ergeben sich Interferogramme mit Wellenlängen ( $\lambda_{Mod}=c/f_{Mod}$ ) im Meterbereich. Das grundlegende Prinzip der Modulationsinterferometrie liefert Abb. 7.

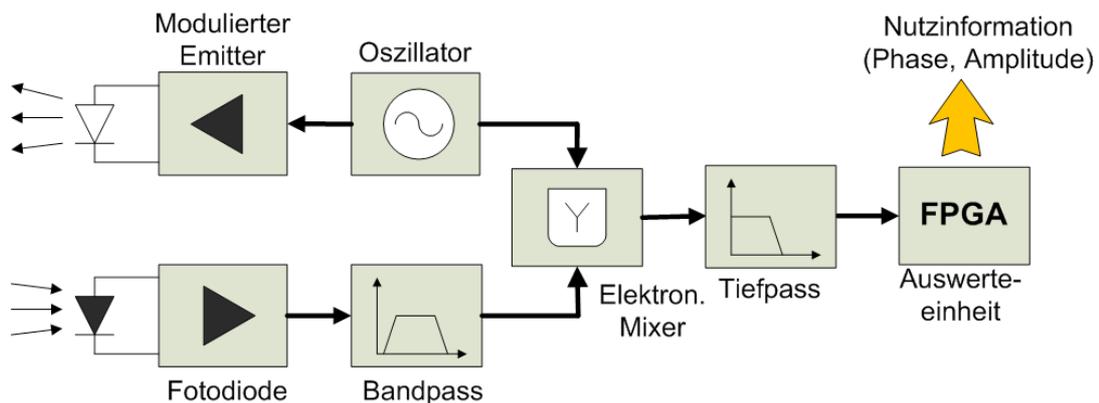


Abbildung 7: Basisprinzip der Modulationsinterferometrie.

Ein über einen Oszillator angesteuerter Emitter sendet mittels des angelegten Modulationssignals amplitudenmoduliertes, inkohärentes Infrarotlicht, wobei im Folgenden angenommen wird, dass die Amplitudenform einer harmonischen Funktion entspricht<sup>32</sup>. Das amplitudenmodulierte Signal wird in Abhängigkeit der betrachteten Szene reflektiert und von einem Rezeptor empfangen,

<sup>30</sup>CWM - Continuous-Wave-Modulation.

<sup>31</sup>Man spricht in diesem Falle auch von optischer Radiofrequenz-Interferometrie (ORFI).

<sup>32</sup>Die Wahl verschiedener Modulationssignale wird in [36] diskutiert.

## 2 Sensorik

welcher die Tiefeninformation in Form einer Laufzeit und Modulationsfrequenz proportionalen Phasenverschiebung  $\Delta \varphi$  enthält.

Um  $\Delta \varphi$  zu bestimmen, wird zunächst das empfangene Signal nach einer Bandpassfilterung zusammen mit dem originalen Modulationssignal als Referenz einem elektronischen Mischer zugeführt. Unter Verwendung von Letzterem sowie eines nachgeschalteten Tiefpasses wird eine Korrelation auf Basis des sich ergebenden Interferogramms zwischen emittiertem und detektiertem Signal ermittelt. Der Zusammenhang zwischen der Korrelation und der gesuchten Entfernung in Form der Phasenverschiebung  $\Delta \varphi$  wird im Folgenden mathematisch wiedergegeben, wobei die Vorgehensweise sich an der in [36] Gewählten anlehnt.

Angenommen wird zunächst, dass sich die empfangene Strahlungsleistung auf einem Rezeptor, wie einem PMD-Pixel, aus reflektiertem Licht sowie unmoduliertem Hintergrund

$$\Phi(t, \Delta \varphi) = M_0 + M_R \sin(\omega_{Mod} t + \Delta \varphi) \quad (1)$$

Variable	Definition	Einheit
$M_0$	Gleichanteil (unmoduliert)	W
$M_R$	Reflexionsamplitude	W
$\omega_{Mod} = 2\pi f_{Mod}$	Modulationskreisfrequenz	rad/s
$\Delta \varphi$	Entfernungsabhängige Phasenverschiebung	rad
$c$	Lichtgeschwindigkeit	m/s

zusammensetzt [36]. Der Zusammenhang zwischen der in (1) auftretenden Phasenverzögerung  $\Delta \varphi$  und der zugehörigen Entfernung  $z$  ist dabei durch

$$z = \frac{c \Delta \varphi}{4\pi f_{Mod}} \quad (2)$$

gegeben. Aufgrund von  $\Delta \varphi \in [0, 2\pi)$  besagt (2), dass die Länge des Eindeutigkeitsintervalls allein durch die Frequenz des Modulationssignals  $f_{Mod}$  bestimmt wird.

Zur Bestimmung der in (1) vorhandenen Entfernungsinformation wird das in [59] diskutierte Lock-In Verfahren eingesetzt. Die notwendigen Demodulationssignale werden dabei als symmetrische Gegentaktsignale mit gleicher Kreisfrequenz

$$U_{+,-}(t, \psi) = U_0 \pm U \sin(\omega_{Mod} t + \psi) \quad (3)$$

gewählt, mit  $U_0$  als dem Gleichanteil,  $U$  der Demodulationsamplitude und  $\psi$  als einstellbare Phasenverschiebung zwischen Modulation und Demodulation. Der in Abb. 7 dargestellte Mischprozess ergibt sich mathematisch durch die Integration des Produktes der Demodulationssignale (3) mit der die Strahlungsleistung definierenden Gleichung (1) über ein ganzzahliges Vielfaches der Periodendauer der Modulationsfrequenz  $T_{Mod} = 1/f_{Mod}$  [36]:

$$\begin{aligned}
N_{+,-}(\Delta\varphi, \psi) &= K \int_0^{n \cdot T_{Mod}} U_{+,-}(t, \psi) \Phi(t, \Delta\varphi) dt, \\
&= K \int_0^{n \cdot T_{Mod}} (U_0 \pm U \sin(\omega_{Mod} t + \psi))(M_0 + M_R \sin(\omega_{Mod} t + \Delta\varphi)) dt, \\
&= K U_0 M_0 n T_{Mod} \pm K U M_R \int_0^{n \cdot T_{Mod}} \sin(\omega_{Mod} t + \psi) \sin(\omega_{Mod} t + \Delta\varphi) dt \\
&\quad \pm K M_0 U \underbrace{\int_0^{n \cdot T_{Mod}} \sin(\omega_{Mod} t + \psi) dt}_{=0} + K U_0 M_R \underbrace{\int_0^{n \cdot T_{Mod}} \sin(\omega_{Mod} t + \Delta\varphi) dt}_{=0}, \\
&= \tilde{K} \pm K U M_R \left( \cos(\psi) \cos(\Delta\varphi) \int_0^{n \cdot T_{Mod}} \sin^2(\omega_{Mod} t) dt \right. \\
&\quad \left. + \sin(\psi) \sin(\Delta\varphi) \int_0^{n \cdot T_{Mod}} \cos^2(\omega_{Mod} t) dt \right) \\
&\quad \pm K U M_R \cos(\psi) \sin(\Delta\varphi) \underbrace{\int_0^{n \cdot T_{Mod}} \cos(\omega_{Mod} t) \sin(\omega_{Mod} t) dt}_{=0} \\
&\quad \pm K U M_R \cos(\Delta\varphi) \sin(\psi) \underbrace{\int_0^{n \cdot T_{Mod}} \cos(\omega_{Mod} t) \sin(\omega_{Mod} t) dt}_{=0}, \\
&= \tilde{K} \pm 0.5 U M_R n T_{Mod} (\cos(\psi) \cos(\Delta\varphi) + \sin(\psi) \sin(\Delta\varphi)), \\
&= \tilde{K} \pm \tilde{G} \cos(\Delta\varphi - \psi).
\end{aligned} \tag{4}$$

Mit  $\tilde{K} = K U_0 M_0 n T_{Mod}$  und  $\tilde{G} = 0.5 U M_R n T_{Mod}$ . Die Konstante  $K$  ist dabei materialabhängig und beinhaltet die entsprechende Reflektivität. Aufgrund der Komposition der Korrelationssignale  $N_{+,-}(\Delta\varphi, \psi)$  ergibt sich eine Abhängigkeit von vier Parametern:

1. Phasenverschiebung  $\psi$ .
2. Variable Phasenverzögerung  $\Delta\varphi = \Delta\varphi(z, f_{Mod})$ .
3. Modulierter Anteil des Signals  $\tilde{G} = \tilde{G}(U, M_R, f_{Mod}, n)$ .
4. Unmodulierter Hintergrundanteil  $\tilde{K} = \tilde{K}(K, U_0, M_0, f_{Mod}, n)$ .

### 2.1.1 ENTFERNUNGSBESTIMMUNG MITTELS HOMODYNVERFAHREN

Die Bestimmung der Entfernung bzw. der Phasenverzögerung  $\Delta\varphi$  unter Verwendung der durch (4) berechneten Korrelationssignale  $N_{+,-}(\Delta\varphi, \psi)$  basiert auf dem Homodynverfahren [36]-[38]. Die diesbezüglich notwendigen Informationen ergeben sich unter Verwendung zweier definierter Phasenverschiebungen  $\psi_1 = 0^\circ$ ,  $\psi_2 = 90^\circ$ :

$$N_{+,-}^{\psi_1}(\Delta\varphi, \psi_1) = \tilde{K} \pm \tilde{G} \cos(\Delta\varphi - \psi_1), \tag{5}$$

$$N_{+,-}^{\psi_2}(\Delta\varphi, \psi_2) = \tilde{K} \pm \tilde{G} \cos(\Delta\varphi - \psi_2). \tag{6}$$

## 2 Sensorik

---

Eine anschließende Differenzbildung zwecks Eliminierung des Gleichanteils liefert:

$$\begin{aligned} N_+^{\psi_1} - N_-^{\psi_1} &= 2\tilde{G} \cos(\Delta\varphi), \\ N_+^{\psi_2} - N_-^{\psi_2} &= 2\tilde{G} \sin(\Delta\varphi). \end{aligned} \quad (7)$$

Durch Bildung von

$$\Delta\varphi = \arctan\left(\frac{N_+^{\psi_2} - N_-^{\psi_2}}{N_+^{\psi_1} - N_-^{\psi_1}}\right) \quad (8)$$

ergibt sich die Phasenverschiebung  $\Delta\varphi$ , welche unter Verwendung von (2) in eine Distanz umgewandelt werden kann. Um eine Abbildung von (8) auf das gesamte Intervall von  $2\pi$  zu erhalten, müssen dabei die Vorzeichen der Komponenten der Argumente der  $\arctan$ -Funktion beachtet werden.

Eine Erweiterung des obigen 2-Phasen-Algorithmus ergibt sich durch die Verwendung der zusätzlichen Phasen  $\psi_3=180^\circ$ ,  $\psi_4=270^\circ$ . Die Bestimmung der Phasenverschiebung  $\Delta\varphi$  geschieht dann mittels

$$\Delta\varphi = \arctan\left(\frac{(N_+^{\psi_2} - N_-^{\psi_2}) - (N_+^{\psi_4} - N_-^{\psi_4})}{(N_+^{\psi_1} - N_-^{\psi_1}) - (N_+^{\psi_3} - N_-^{\psi_3})}\right), \text{ wobei } \psi_n = \frac{2\pi}{4}(n-1), n=1,2,3,4. \quad (9)$$

Durch

$$A_M = \frac{\sqrt{((N_+^{\psi_2} - N_-^{\psi_2}) - (N_+^{\psi_4} - N_-^{\psi_4}))^2 + ((N_+^{\psi_1} - N_-^{\psi_1}) - (N_+^{\psi_3} - N_-^{\psi_3}))^2}}{2} \quad (10)$$

wird die Modulationsamplitude und mittels

$$I = \frac{(N_+^{\psi_1} - N_-^{\psi_1}) + (N_+^{\psi_2} - N_-^{\psi_2}) + (N_+^{\psi_3} - N_-^{\psi_3}) + (N_+^{\psi_4} - N_-^{\psi_4})}{4} \quad (11)$$

die Intensität berechnet. Der Vorteil, der sich durch die Verwendung des 4-Phasen-Algorithmus ergibt, ist, dass sich dadurch, im Vergleich zum 2-Phasen-Algorithmus, etwaige Unsymmetrien im Aufbau des PMD-Chips kompensieren lassen [36], [40], [42], [62]. Dieser Umstand hat den Einsatz dieses Verfahrens im Rahmen dieser Arbeit motiviert.

Die in [36] hergeleitete Standardabweichung der Phasenverzögerung

$$\sigma_{\Delta\varphi} = \sqrt{\frac{\tilde{K}}{2G^2}} \quad (12)$$

## 2 Sensorik

sowie die daraus resultierende Standardabweichung der Entfernungsbestimmung,

$$\sigma_z = \frac{\partial z}{\partial \Delta \varphi} \sigma_{\Delta \varphi} = \frac{c}{4\pi f_{Mod}} \sqrt{\frac{\tilde{K}}{2\tilde{G}^2}}, \quad (13)$$

lässt erkennen, dass sich die Präzision der Entfernungsbestimmung sowohl durch Minimierung von  $\tilde{K}$  als auch durch Maximierung<sup>33</sup> von  $\tilde{G}$  erhöhen lässt.

### 2.2 MULTICAM – MONOKULARES 2D/3D-KAMERASYSTEM

Der grundsätzliche Aufbau der MultiCam besteht aus drei Hauptkomponenten (Abb. 8):

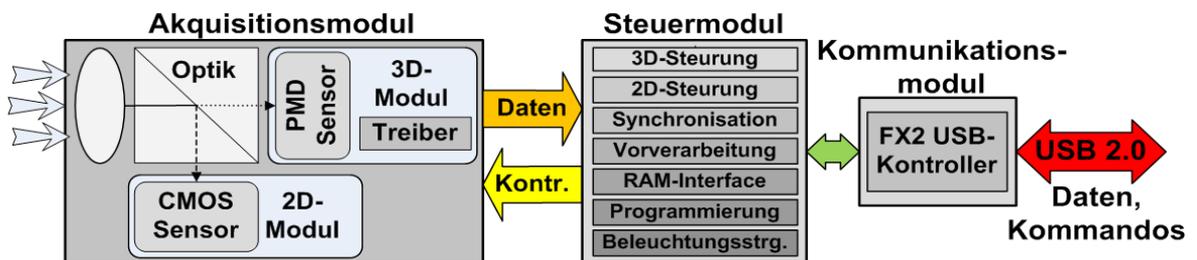


Abbildung 8: Hauptkomponenten der MultiCam.

Das Akquisitionsmodul besitzt ein monokulares optisches Design, bestehend aus einer Kombination eines farbigen Bilddaten liefernden CMOS-Chips, eines PMD-Sensors sowie eines dichroitischen Splitters, welcher die Aufgabe der Separation von sichtbarem und nahem Infrarotlicht übernimmt (Abb. 9). Der auf dem PMD-Chip implementierte SBI-Schaltkreis<sup>34</sup> dient der Erhöhung der Sensordynamik und erlaubt dessen Einsatz auch unter starkem Umgebungslicht [67], [72].

Das Steuermodul besteht aus einer FPGA-basierten, programmierbaren Logik und bietet die Möglichkeit einer Vorverarbeitung der Sensordaten bereits in der Kamera selbst. Zusätzlich übernimmt es die Aufgabe der Ansteuerung der Sensoren und ermöglicht die Festlegung von Beleuchtungszeit, zeitlicher Synchronisation der 2D/3D-Bilder sowie der Modulationsfrequenz.

Das Kommunikationsmodul übernimmt den Transport der Daten zu sowie der Kommandos von einer angeschlossenen Datensenke (Rechner). In der benutzten Version besitzt die Kamera ein USB 2.0-Interface mit einer effektiv gemessenen Übertragungsgeschwindigkeit von ca. 30 MB/s.

Der monokulare optische Aufbau der MultiCam, dargestellt in Abb. 9, wird durch die Verwendung eines dichroitischen Filters ermöglicht, welcher diesbezüglich zwischen den Hypotenusen zweier Glasprismen eingebettet wird. Während Lichtwellen aus dem nahen Infrarotbereich zwischen ca. 760nm und 1000nm direkt auf den PMD-Chip projiziert werden, wird das Licht aus dem sichtbaren Bereich auf den CMOS-Sensor reflektiert. Eine Analyse der Eigenschaften des verwendeten dichroitischen Filters findet sich in [62].

Der am Eingang der Optik vorhandene Infrarot-Filter dient zur Abschirmung der Hintergrundstrahlung im nahen Infrarotbereich, welche insbesondere durch die Sonne hervorgerufen wird.

<sup>33</sup>Z. B. durch Verwendung einer stärkeren Beleuchtung.

<sup>34</sup>SBI - *Suppression of Background Illumination* [222].

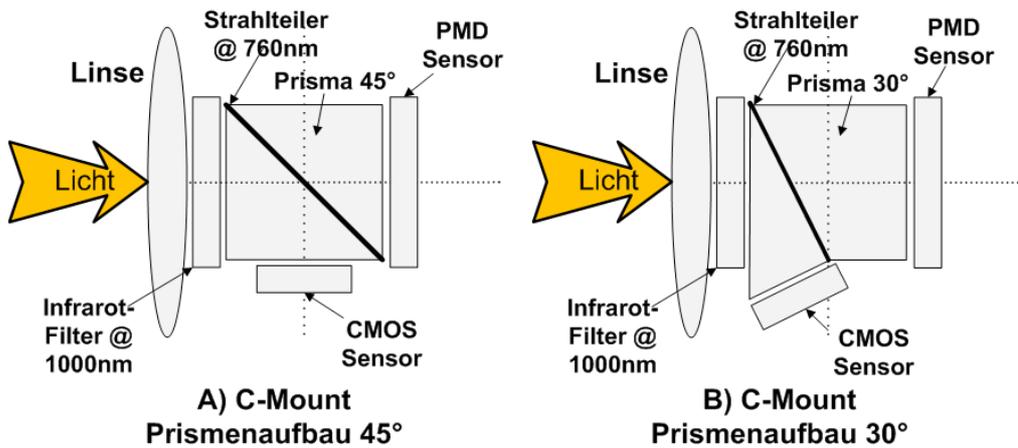


Abbildung 9: Optischer Aufbau der MultiCam.

Links: Ursprünglicher Aufbau. Rechts: Modifizierte Version.

Für beide Konfigurationen aus Abb. 9 gilt, dass sich durch eine exakte Positionierung der Sensoren eine pixelgenaue Abdeckung des Sichtbereichs beider Chips erreichen lässt. Ein Vorteil, welcher sich durch die Verwendung eines Aufbaus gemäß des Schemas B aus Abb. 9 ergibt, ist, dass in diesem Fall die Notwendigkeit einer softwaremäßigen Spiegelung des Bildes des CMOS-Sensors entfällt.

### 2.2.1 ZEITLICHE SYNCHRONISATION

Im Rahmen einer Beobachtung bzw. Analyse von dynamischen Umgebungen mittels multipler Sensoren stellt die zeitliche Synchronisation der anfallenden Informationen bzw. Bilder eine fundamentale Notwendigkeit dar. Speziell für den Fall binokularer bzw. multiokularer Aufbauten mit eventuell verschiedenen Bildwiederholraten und Anbindungen zur Extraktion von Tiefeninformationen ergibt sich die Problematik einer zuverlässigen Zeitsynchronisation [63].

Die Verwendung der MultiCam erleichtert die Synchronisation der Bilder der unterschiedlichen Modalitäten erheblich. Aufgrund der Flexibilität des FPGA ist es möglich, eine konstante Zeitsynchronisation zwischen 2D- und 3D-Bilder zu gewährleisten, wobei zusätzlich die Möglichkeit gegeben ist, das 2D-Bild zu einem beliebigen der vier PMD-Phasenbilder zu synchronisieren (Abb. 10).

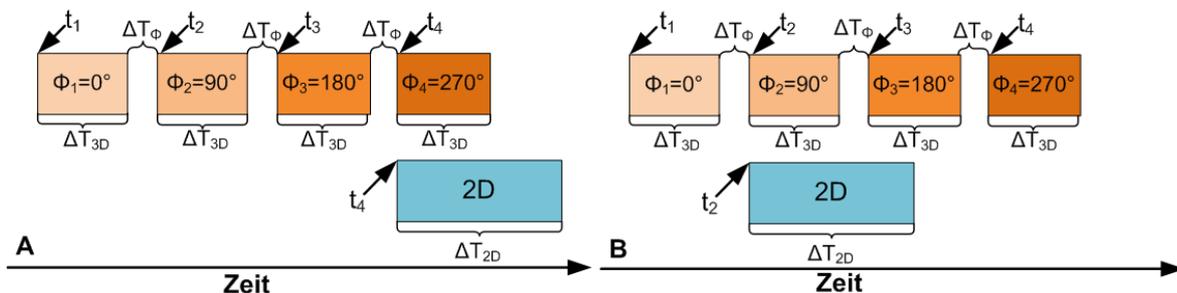


Abbildung 10: Implementierte 2D/3D-Bildsynchronisation.

Der Vorteil der konstanten zeitlichen Abstände zwischen jeweils zwei Phasenbildern liegt in der Minimierung von sogenannten *Ghosting*-Artefakten. Ein Effekt, welcher besonders im Falle unterschiedlicher zeitlicher Abstände zwischen den einzelnen Phasenbildern bei der Betrachtung beweglicher Objekte auftritt. Eine Diskussion diesbezüglich findet sich ebenfalls in [62].

### 2.2.2 LATERALE REGISTRIERUNG

Ein weiterer Vorteil, der sich durch die Benutzung der MultiCam ergibt, ist die im Vergleich zu multiokularen Systemen immense Vereinfachung der lateralen Registrierung der 2D-Intensitäts- und 3D-Distanzbildern [63]-[66]. Eine detaillierte Erläuterung der Registrierung von Farb- und Entfernungsbildern unter Benutzung der MultiCam sowie der Besonderheiten in Bezug auf die gewählten Sensoren findet sich ebenfalls in [62]. Die zentrale Aussage lautet, dass aufgrund der Kommensurabilität<sup>35</sup> der eingesetzten Sensoren sowie deren internen Anordnung eine laterale Registrierung der Sensordaten bereits durch eine simple Translation möglich ist. Faktisch bedeutet dieses, dass durch eine Verschiebung der 2D-Daten eine hinreichend genaue Abdeckung der Distanz- und Intensitätsbilder der unterschiedlichen Sensoren gegeben ist<sup>36</sup>.

### 2.2.3 VORVERARBEITUNG (PRÄPROZESSIERUNG) DER ENTFERNUNGSDATEN

Im Allgemeinen bedürfen die von dem 3D-Sensor direkt akquirierten Daten einer Vorverarbeitung. Gründe hierfür sind u. a. das Rauschen des PMD-Sensors, Bewegungsartefakte, *Flying*-Pixel sowie die Uneindeutigkeit der Entfernung im Falle von Messungen außerhalb des durch die Modulationsfrequenz definierten Eindeutigkeitsbereichs (2). Diesbezügliche Diskussionen sowie eventuell anwendbare Gegenmaßnahmen zur Reduzierung bzw. Eliminierung von entsprechend induzierten Fehlern finden sich in [36], [37], [40], [42], [62], [67]-[70], [75], [188], [224].

Eine systematische Fehlerquelle stellt die verwendete Beleuchtung selbst dar. Zunächst führen die unterschiedlichen Signallaufzeiten zum PMD-Chip sowie zu der Beleuchtung zu einer festen Phasenverschiebung bzw. einem Distanzoffset. Weiterhin ergibt sich, bedingt durch die Deformation des Modulationssignals aufgrund nichtidealer Bauteile, eine von der Distanz, der Modulationsfrequenz sowie der Integrationszeit abhängige, nichtlineare Veränderung (Abb. 11) der Entfernungsdaten [40], [42], [62], [223]-[225]. Dieser Effekt wird in der spezifischen Literatur u. a. als *wiggling* oder *circular error* bezeichnet [226].

Durch die Varianz der verwendeten Bauteile, der unterschiedlichen Betriebsdauer sowie Abstrahlintensitäten ergibt sich eine individuelle Beleuchtungscharakteristik, was letztendlich direkten Einfluss auf die Erzeugung der Entfernungsdaten hat und dazu führt, dass die Bestimmung der charakteristischen Abweichungen von den realen Distanzen für jede MultiCam-Beleuchtungskombination durchgeführt werden muss (Abb. 12).

Der in [74] verwendete Ansatz zur Bestimmung einer charakteristischen Abweichfunktion besteht in einer Annäherung der Abweichung der Distanzdaten des PMD-Sensors von den realen Werten durch ein Polynom. Als verwendeter gemessener Entfernungswert wird dabei jeweils der Mittelwert der Entfernungen der vier zentralen Pixel verwendet. Dem Effekt der Verschiebung der Distanzdaten durch Verwendung unterschiedlicher Integrationszeiten wird dadurch begegnet, dass für jede Integrationszeit ein entsprechendes Polynom aus den akquirierten Distanzdaten erzeugt wird.

Eine Annäherung der charakteristischen Abweichfunktion durch die Verwendung einer Look-Up Tabelle (LUT) diskutiert der Ansatz in [212]. Ähnlich dem Ansatz von [74] werden lediglich die Distanzdaten eines zentralen Pixel verwendet, um die Distanzabweichungen bei unterschiedlichen Integrationszeiten und Entfernungen in die LUT einzutragen.

Die Einbeziehung aller PMD-Pixel in eine globale B-Spline Funktion diskutiert der Ansatz in [69]. Eine Erweiterung dieses Ansatzes zwecks Einbeziehung der Verschiebung der Distanzdaten aufgrund unterschiedlicher Integrationszeiten wird in [213] erläutert<sup>37</sup>.

---

<sup>35</sup>Kommensurabilität bedeutet in diesem Kontext die prinzipielle Art, wie die Sensoren ihre Informationen zur Verfügung stellen. Im Falle der MultiCam liefern beide Chips ihre Daten in Form einer zweidimensionalen Matrix.

<sup>36</sup>Die Wahl der 2D-Daten als Translationsobjekt ist aufgrund der im Vergleich zum PMD-Chip höheren Auflösung des 2D-Sensors getroffen worden.

<sup>37</sup>Eine detailliertere Beschreibung des in [213] vorgestellten Ansatzes findet sich in [227].

## 2 Sensorik

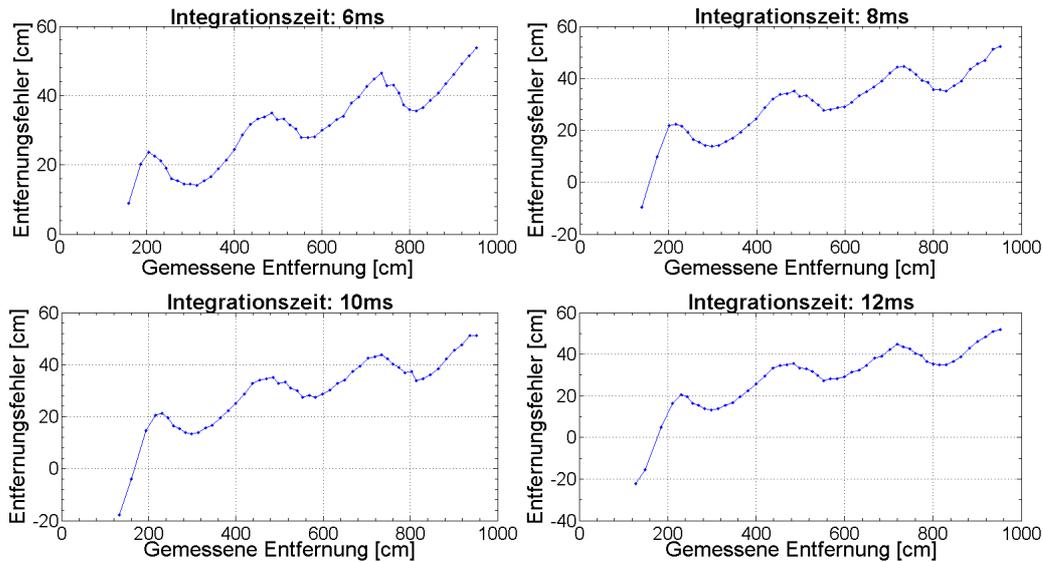


Abbildung 11: Abweichungen der Entfernungsdaten der MultiCam bei unterschiedlichen Integrationszeiten.

Abgebildet sind die Differenzen zwischen realen und gemessenen Entfernungen für vier unterschiedliche Integrationszeiten des PMD-Chips einer MultiCam-Beleuchtungskombination und gleicher Modulationsfrequenz. Neben dem „wiggling“ ist ebenfalls eine von der Integrationszeit abhängige Veränderung der Distanzdaten zu erkennen (Daten zwischen Messpunkten wurden linear interpoliert).

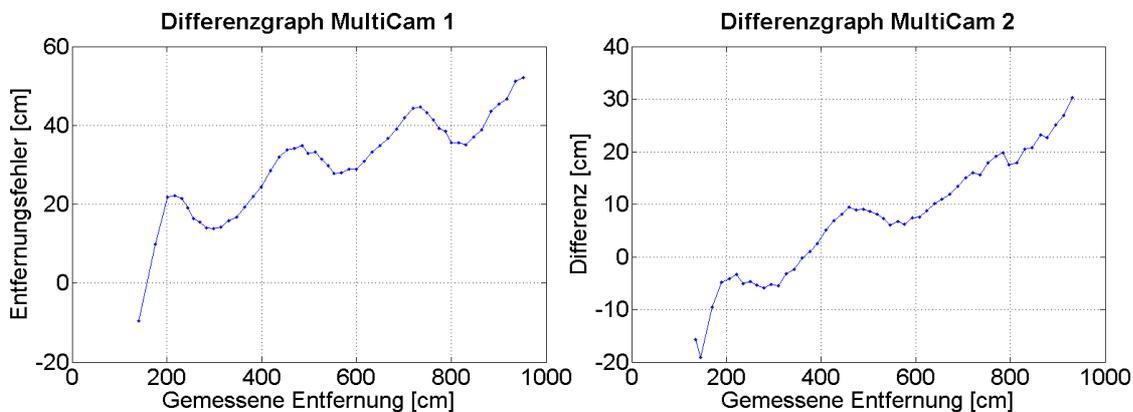


Abbildung 12: Distanzabweichung der verwendeten MultiCam-Beleuchtungskombinationen. Verwendet wurden dabei gleiche Modulationsfrequenzen sowie Integrationszeiten (Daten linear interpoliert).

### 2.2.3.1 CHARAKTERISTISCHE ABWEICHFUNKTION

Der in dieser Arbeit verwendete Ansatz zur Bestimmung einer beleuchtungscharakteristischen Distanzkorrekturfunktion basiert auf den gemittelten Distanzinformatoren von vier mittigen Pixel des PMD-Sensors. Die Daten wurden unter Verwendung eines gleichmäßigen, sich nicht verändernden Hintergrunds akquiriert. Das betrachtete Abstandsintervall entsprach der Entfernung von 1.5m bis 9.0m zum Hintergrund, wobei der reale Abstand zwischen zwei Messpunkten jeweils 15cm betrug. Zur Überprüfung des realen Abstands wurde ein Laser-Distanzgeber *Disto A5* von Leica mit einer Genauigkeit von  $\pm 1.5$  mm eingesetzt<sup>38</sup>. Gefittet wurden die Abweichungen

<sup>38</sup>Anzumerken ist, dass die Distanzinformatoren aus Abb. 11 und 12 auf die gleiche Weise akquiriert worden sind.

## 2 Sensorik

des gemessenen Abstands zur realen Entfernung durch ein Polynom 10.ter Ordnung. Durch die Verfügbarkeit der Abweichfunktion

$$P(z) = \sum_{i=0}^{10} p_i z^i \quad (14)$$

ergibt sich die Möglichkeit einer kontinuierlichen Distanzkorrektur über das betrachtete Distanzintervall (Abb. 13). Im Gegensatz zum Ansatz in [69], [74], [188] ist die Bestimmung eines Offsets nicht zusätzlich durchgeführt worden, da durch die Korrekturfunktion (14) der entsprechende Einfluss bereits einbezogen worden ist.

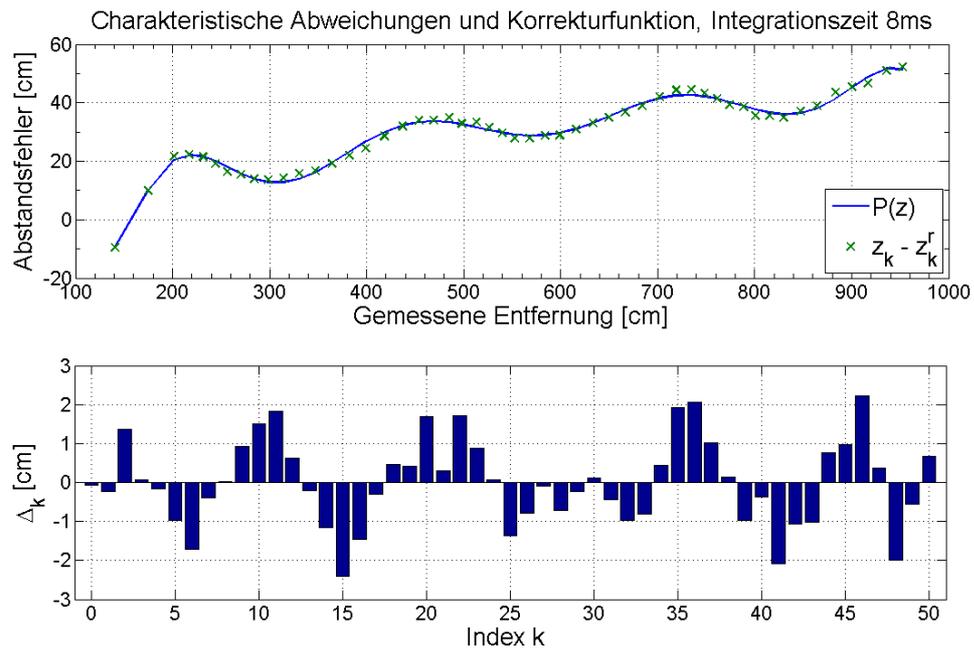


Abbildung 13: Charakteristische Abweichung und Korrekturfunktion.

Oben: Überlagerung der Graphen der Differenzen zwischen gemessenen Distanzen  $z_k$ ,  $k=0, \dots, 50$ ,  $k \in \mathbb{N}$  und realer Entfernung  $z_k^r = 150\text{cm} + k 15\text{cm}$  sowie der berechneten Korrekturfunktion  $P(z)$ ,  $z \in \mathbb{R}$ . Unten: Differenzen zwischen den Abstandsfehlern und der Korrekturfunktion  $\Delta_k = (z_k - z_k^r) - P(z_k)$ .

Die Anwendung der Korrekturfunktion (14) auf die gemessenen Werte ist in (Abb. 14) enthalten. Zu erkennen ist die nun gute Übereinstimmung der korrigierten Messpunkte mit den realen Distanzen.

### 2.2.3.2 RADIALE KORREKTUR

Zusätzlich zur kameraspezifischen Distanzkorrektur durch eine konfigurationsabhängige Korrekturfunktion (14) bedarf es einer von der verwendeten Konfiguration der Kamera unabhängigen, radialen Korrektur. Aufgrund unterschiedlicher Weglängen der reflektierten Strahlen entlang der optischen Achse der MultiCam im Vergleich zu denen aus den Randbereichen des Sichtfeldes, müssen die gemessenen Entfernungsdaten entsprechend korrigiert werden.

## 2 Sensorik

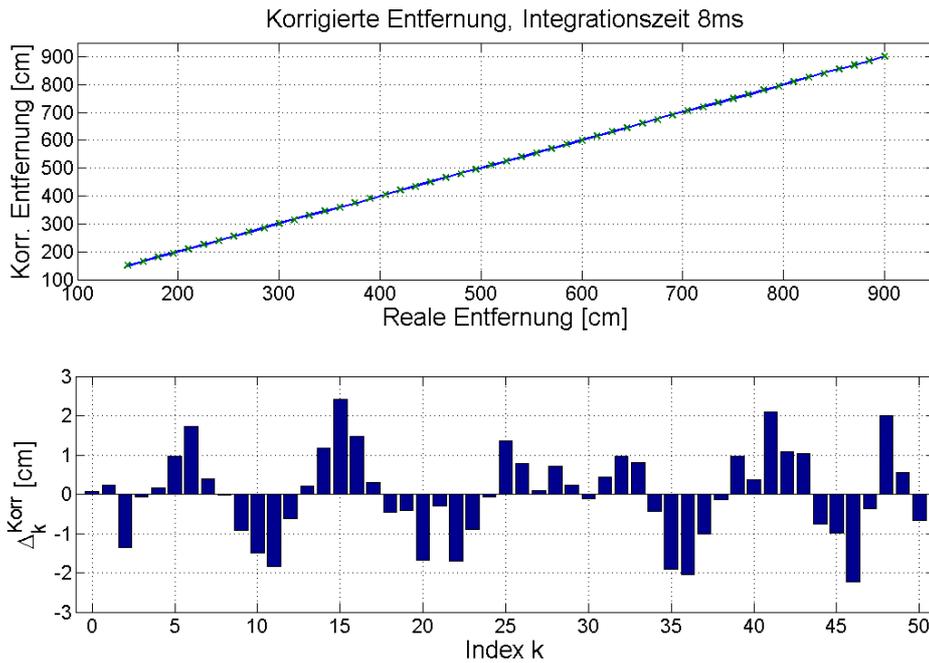


Abbildung 14: Anwendung der Korrekturfunktion auf die gemessenen Entfernungen.

Oben: Reale Entfernungen gegen korrigierte Distanzen (Zwischenwerte linear interpoliert). Unten: Differenzen  $\Delta_k^{Korr} = (z_k - P(z_k)) - z_k^r$  zwischen korrigierten Distanzwerten und realer Entfernung.

Angenommen wird hierbei eine monostatische Anordnung: Die Koordinatensysteme der Beleuchtung  $K^{Illu}$  und der Kamera  $K^{MCam}$  besitzen den gleichen Ursprung, welche als das optische Zentrum der Kamera interpretiert wird (Abb. 15).

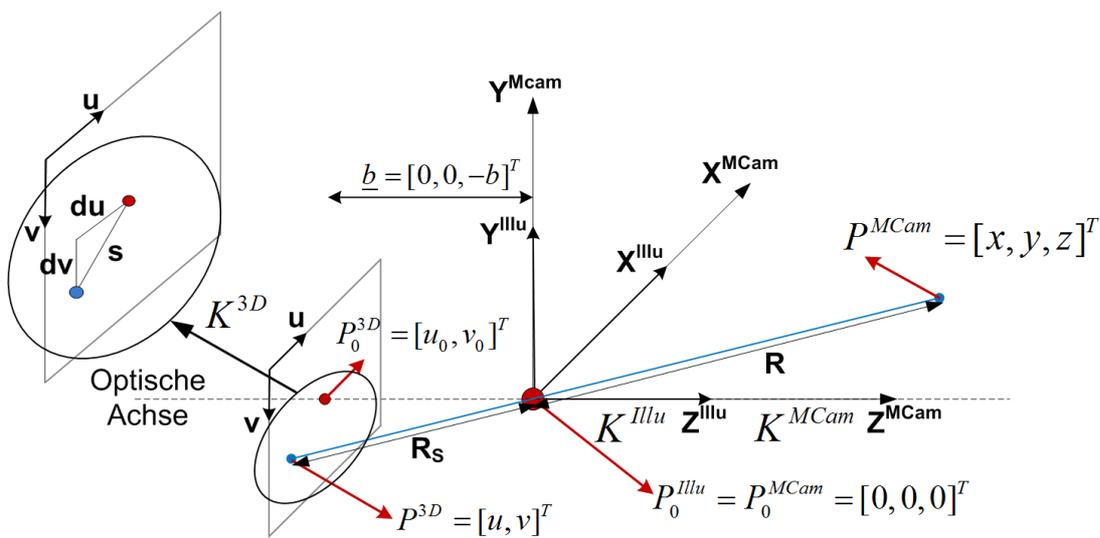


Abbildung 15: Zusammenhang verschiedener Koordinatensysteme.

Informationen des PMD-Chips, welche im  $K^{3D}$ -Koordinatensystem vorhanden sind, müssen unter Verwendung von (16) in das dreidimensionale Koordinatensystem  $K^{MCam}$  der Kamera projiziert werden.

## 2 Sensorik

Weiterhin wird angenommen, dass die optische Achse der MultiCam den PMD-Chip im Bildmittelpunkt  $P_0^{3D}=[u_0, v_0]^T$  trifft, wobei  $u_0, v_0$  Pixelkoordinaten im Koordinatensystem  $K^{3D}$  des PMD-Chip darstellen. Der Vektor  $\underline{b}=[0, 0, -b]^T$  entspricht der Verbindung vom optischen Zentrum der Kamera  $P_0^{MCam}$  zum Bildmittelpunkt  $P_0^{3D}$  in  $K^{MCam}$ . Gemäß [42] und [62] kann eine Vereinfachung vorgenommen werden, indem die Länge von  $\underline{b}$  mit der Brennweite  $f$  der verwendeten Optik gleichgesetzt wird. Gegeben sei der Punkt  $P^{MCam}=[x, y, z]^T$  im Koordinatensystem  $K^{MCam}$ . Seine Position im Koordinatensystem des Sensors entspricht  $P^{3D}=[u, v]^T$ . Die vom PMD-Chip gemessene Entfernung ist dann bestimmt durch

$$R_{PMD}=2R+R_S, \quad R_S=\sqrt{(s_u du)^2+(s_v dv)^2+f^2}. \quad (15)$$

Die Skalierungsfaktoren  $s_u, s_v$  können aus den jeweiligen Sensorspezifikationen in [67] bzw. [73] entnommen werden und betragen 0.1 mm/Pixel für den PMD3k-Chip sowie 0.01 mm/Pixel für den 2D-Sensor. Der Faktor Zwei in (15) rührt von der Tatsache her, dass das Licht zunächst vom Emissionszentrum zum Objekt und von dort wieder zurückreist. Nach [62] lassen sich die Koordinaten aus  $K^{3D}$  in das Kamerasystem  $K^{MCam}$  unter Verwendung des Strahlensatzes transformieren, wobei gleichzeitig eine radiale Korrektur der von der optischen Achse entfernten Strahlen vorgenommen wird:

$$\begin{aligned} x &= z \frac{s_u du}{f}, \\ y &= z \frac{s_v dv}{f}, \\ z &= R \cos \alpha, \quad \tan \alpha = \frac{\sqrt{(s_u du)^2+(s_v dv)^2}}{f} = \frac{s}{f}. \end{aligned} \quad (16)$$

Ein wichtiger und nicht kompensierbarer Effekt, welcher grundsätzlich beachten werden muss, ist die durch die Multipropagation des ausgesandten Lichts in der Kamera selbst sich ergebende Verfälschung von Distanzdaten durch Objekte in geringer Entfernung (Abb. 16).

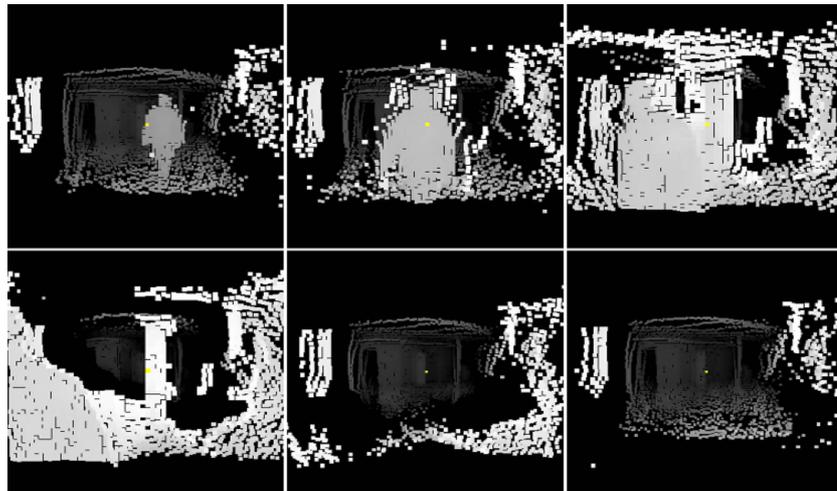


Abbildung 16: Effekt der Multipropagation.

*Verfälschung der Hintergrunddaten durch ein sich der MultiCam näherndes Objekt.*

## 2 Sensorik

Die hohe Reflexionsintensität dieser Objekte wechselwirkt mit den schwächeren Reflexionen weiter entfernter Objekte bzw. des Hintergrunds und führt dazu, dass deren Tiefeninformationen drastisch verändert bzw. verfälscht werden und im weiteren Verlauf nicht oder nur sehr eingeschränkt verwendet können. Dieses impliziert, dass im Falle der Bewegung eines Objekts in unmittelbarer Nähe der MultiCam lediglich dessen Distanzinformationen für eine verlässliche Bearbeitung verwendet werden können. Eine Diskussion der Multipropagation sowie der durch sie induzierten Problematik findet sich in [62], [75].

### 2.2.4 SIMULTANE OPERATION MEHRERER MULTICAMS

Aufgrund der Tatsache, dass die Entfernungsmessung einem aktiven Messverfahren entspricht, stellt sich zusätzlich die Frage nach einer gegenseitigen Beeinflussung im Falle simultan operierender MultiCams, welche im Rahmen dieser Arbeit durch folgende Eigenschaften definiert wird:

1. Alle am Geschehen partizipierenden MultiCams besitzen teilweise überlappende Integrationsintervalle.
2. Die Möglichkeit der Wahrnehmung der Beleuchtung fremder Kameras ist gegeben.

An dieser Stelle ist anzumerken, dass die in diesem Kapitel durchgeführten Untersuchungen lediglich der Bestimmung von Parametern dienen, die eine simultane Operation mehrerer MultiCams mit teilweise überlappendem Sichtbereich und ohne gegenseitige Beeinflussung ermöglichen. Eine detaillierte Analyse der simultanen Operation zweier MultiCams unter Verwendung mehrerer Beleuchtungen sowie Beleuchtungsanordnungen findet sich in [62].

In Abb. 17 sind die Entfernungswerte eines ausgewählten Pixels einer MultiCam über eine Sequenz von jeweils 20 aufeinanderfolgenden Bildern unter Verwendung zweier simultan operierender MultiCams abgebildet. Für die Erzeugung der 9 Bildsequenzen wurde die Modulationsfrequenz einer MultiCam auf 15.2 MHz festgelegt, während die Modulationsfrequenz der Zweiten in 10 kHz-Schritten variiert wurde.

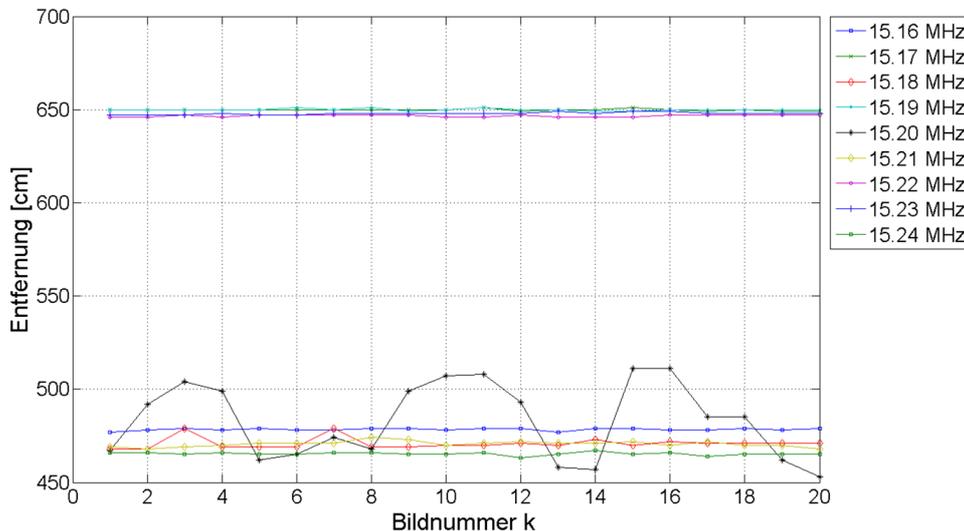


Abbildung 17: Entfernungswerte eines ausgewählten Pixels der MCam1 unter Verwendung diverser Modulationsfrequenzen I.

Analyse durchgeführt auf 9 Videosequenzen mit einer jeweiligen Länge von 20 aufeinanderfolgenden Bildern. Die Punkte der Graphen entsprechen den Distanzwerten eines ausgewählten Pixels bei entsprechender Modulationsamplitude und Bildnummer.

## 2 Sensorik

Gut zu erkennen ist die massive Schwankung der Entfernung im Falle der Verwendung gleicher Modulationsfrequenzen, was auf eine gegenseitige Beeinflussung der Beleuchtungen schließen lässt. Eine deutlichere Aussage liefert die Darstellung der Entfernungsvarianzen (Abb. 18). Der eklatante Unterschied im Falle gleicher Modulationsfrequenzen lässt erkennen, dass eine simultane Operation mehrerer MultiCams unter Verwendung gleicher Modulationsfrequenzen nicht möglich ist.

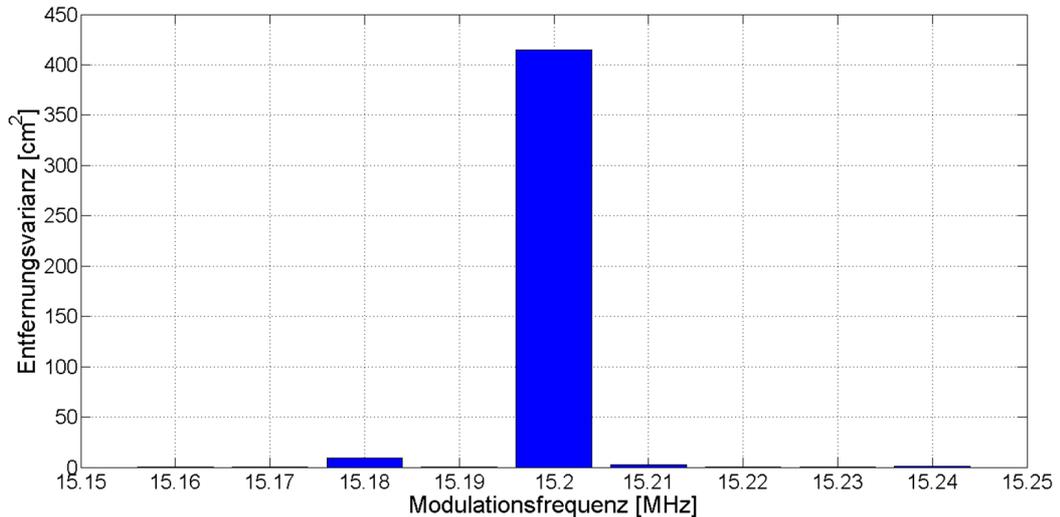


Abbildung 18: Entfernungsvarianzen I.

Die zu Abb. 17 zugehörigen Varianzen der gemessenen Entfernungen des Pixels. Deutlich erkennbar ist die massive Zunahme der Varianz im Falle gleicher Modulationsfrequenzen, was auf eine gegenseitige Beeinflussung der Beleuchtungen hindeutet.

Die Ergebnisse einer zusätzlichen Messreihe, jetzt unter Verwendung größerer Frequenzintervalle von jeweils 50 kHz, sind in den Abb. 19 und 20 dargestellt.

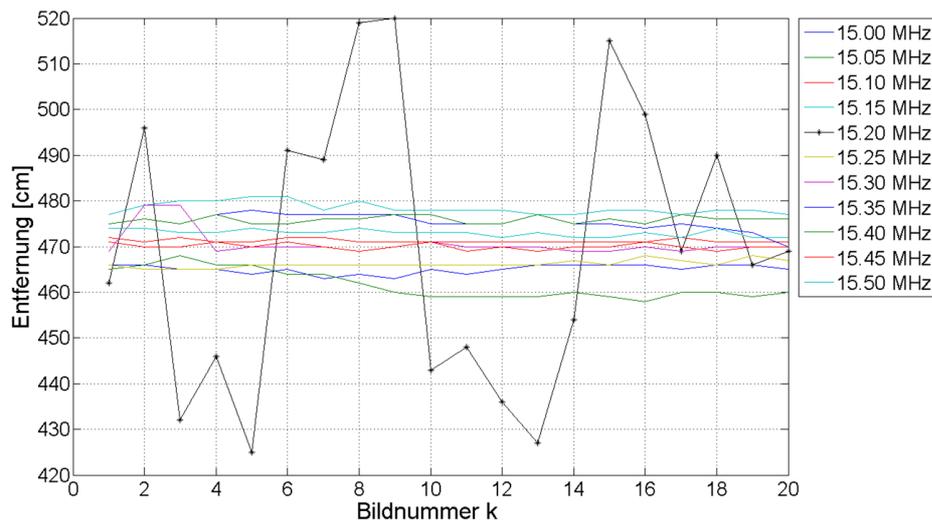


Abbildung 19: Entfernungswerte der MCam1 unter Verwendung diverser Modulationsfrequenzen II.

Analyse durchgeführt auf 11 Videosequenzen mit einer jeweiligen Länge von 20 aufeinanderfolgenden Bildern. Wie auch in Abb. 17 zu sehen ist, entsprechen die Punkte der Graphen den gemessenen Distanzen eines ausgewählten Pixels.

## 2 Sensorik

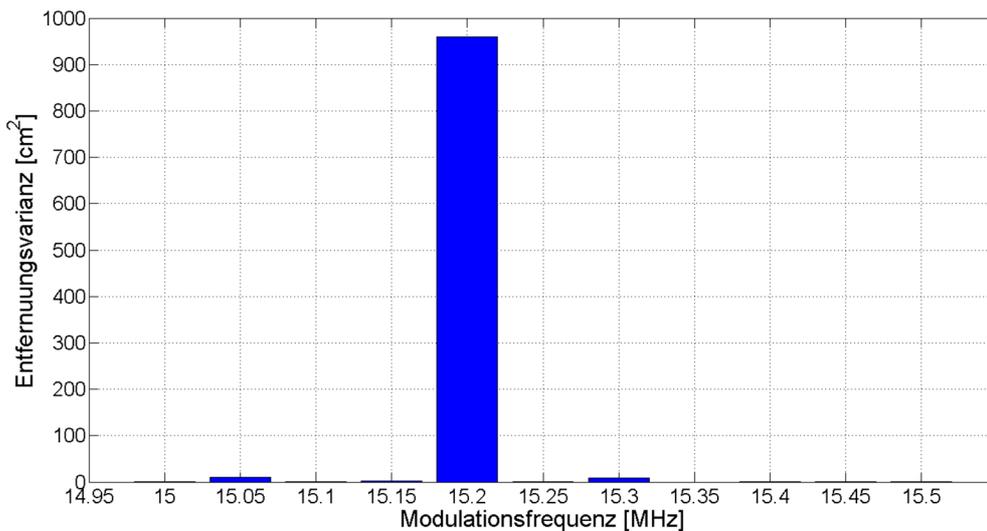


Abbildung 20: Varianzen der Entfernungsdaten II.

Eindeutig erkennbar sind die massiven Schwankungen der Entfernungsdaten im Falle gleicher Modulationsfrequenzen bei 15.2 MHz.

Die Aussage dieser Graphen ist, dass für Unterschiede in der Modulationsfrequenz von  $\geq 50$  kHz, eine vernachlässigbare gegenseitige Beeinflussung durch andere, simultan operierende MultiCams zu erwarten ist.

Eine vergleichbare Aussage liefert eine unter anderen Bedingungen durchgeführte Analyse in [62]. Für die im Rahmen dieser Arbeit eingesetzten MultiCams wurde die Modulationsfrequenz der ersten Kamera auf 15.0 MHz und der zweiten Kamera auf 15.2 MHz gesetzt.

### 2.3 FAZIT

Speziell im Vergleich zu bin- oder multiokularen Systemen wird der Vorteil des gewählten monokularen Aufbaus deutlich:

1. Eine hinreichend genaue Registrierung diverser Datenquellen erfordert einen streng statischen Aufbau, eine Voraussetzung, die nicht immer in simpler Weise zu erfüllen ist. Unbekannte, ungewollte oder versehentlich durchgeführte relative Bewegungen der partizipierenden Systeme untereinander, insbesondere in bewegten Anordnungen, führen zwangsläufig zu einer fehlerhaften Registrierung. Vor allem in Anordnungen, die eine Mehrzahl entfernungsmessender Geräte verwenden, ermöglicht der hier diskutierte monokulare Aufbau eine drastische Reduktion des zur Registrierung der Daten notwendigen Arbeitsaufwandes.
2. Zusätzlich zu der einfach durchführbaren lateralen Registrierung liefert die MultiCam temporal synchronisierte Bilder. Die Forderung an zeitlich synchronisierten Daten im Falle bin- oder multiokularer Systeme ist im Allgemeinen nur unter gewissen Annahmen (z. B. reduzierte Bildwiederholrate) gewährleistet und ist nicht trivial lösbar.
3. Die prinzipielle Unabhängigkeit der Distanzmessung bedingt durch die Anwendung eines aktiven Messverfahrens bietet eine zuverlässige Entfernungsmessung auch im Falle veränderlicher Lichtverhältnisse (Tag, Nacht, etc.) und stellt einen fundamentalen Vorteil gegenüber rein auf opto-elektrischen Sensoren basierenden Systemen dar, da diese bzw. die auf ihren Daten operierenden Algorithmen im Allgemeinen empfindlich gegenüber variierenden Lichtverhältnissen sind.

## 2 Sensorik

---

4. Insbesondere im Falle der geringen oder gar nicht vorhandenen Texturierung der zu lokalisierenden Objekte bietet die MultiCam aufgrund ihrer Unabhängigkeit in Bezug auf Texturinformationen im Vergleich zu optischen Stereokamerasystemen einen operativen Vorteil. Es sollte jedoch beachtet werden, dass die Entfernungsdaten eine gewisse Abhängigkeit von dem Reflexionskoeffizienten eines bestrahlten Objekts besitzen. Eine diesbezügliche Analyse findet in [62] statt.
5. Der Einsatz multipler simultan operierender Geräte mit partiell überdeckendem Sichtbereich ist durch eine entsprechende Variation der Modulationsfrequenzen der einzelnen MultiCams durchführbar. Die notwendige Frequenzdifferenz von einigen Dekaden an kHz ist im Vergleich zur Modulationsfrequenz relativ gering und hat zur Folge, dass die Anzahl der einsetzbaren Geräte mit ähnlichen Modulationsamplituden zwecks gemeinsamer Überwachung eines Innenraumes hinreichend groß gewählt werden kann.



## 3 AGENTENPARADIGMA-SYSTEMTOPOLOGIE

In Abschnitt 1.2 wurden die wesentlichen Vorzüge der Verwendung eines agentenbasierten Ansatzes sowie der Ansatz zur Lokalisation sowie Verfolgung von Personen unter Verwendung verteilter, kooperierender Multisensorknoten kurz diskutiert, wobei der Begriff des Agenten vorweggenommen wurde. Ziel dieses Kapitels ist die Einführung notwendiger Definitionen, der Vorstellung der verwendeten Infrastruktur auf Basis des Agentenparadigmas, der Datenfusion sowie der Diskussion des implementierten Informationsflusses des Systems.

### 3.1 DAS AGENTENPARADIGMA

#### 3.1.1 AGENTENDEFINITION

Zunächst muss konstatiert werden, dass aufgrund der Vielzahl von vorhandenen und unterschiedlichen Definitionen eine Konkretisierung des Agentenbegriffs problematisch ist. Der erste Versuch, einer im Kontext dieser Arbeit verwendbaren Definition, gibt Wooldridge [84] mit:

*An agent is a computer system that is situated in some environment, and that is capable of autonomous action in this environment in order to meet its design objectives.*

Diese Definition ist abhängig von gewissen intuitiven Vorstellungen und bietet streng genommen kein zu einer Festlegung hinreichendes, objektives Kriterium<sup>39</sup>. Zur weiteren Präzisierung des Agentenbegriffs werden dazu in [84] vier zusätzliche graduelle Konstituenten angegeben:

1. **Autonomie:** Die Aktivitäten eines Agenten unterliegen in hohem Maße seiner Kontrolle. Externe Anweisungen haben nur einen bedingten Einfluss.
2. **Reaktivität:** Agenten sind dazu in der Lage, Änderungen in ihrer Umgebung wahrzunehmen, um auf diese in angemessener Form zu reagieren.
3. **Proaktivität:** Agenten können zwecks Erfüllung der Zielsetzung aus eigener Initiative heraus einem zielgerichteten Verhalten nachgehen.
4. **Interaktivität:** Agenten besitzen die Möglichkeit der gemeinsamen Kommunikation.

Eine weitere, sich in Bezug auf die praktische Realität orientierende Definition wird von der *Foundation for Intelligent Physical Agents* (FIPA), einer Vereinigung mit dem Ziel, Standardisierungen im Bereich der Softwareagenten durchzuführen, geliefert [111]:

*An agent is a computational process that implements the autonomous, communicating functionality of an application. Agents communicate using an **agent communication language** (ACL). An agent is the fundamental actor on an **agent platform** (AP) which combines one or more service capabilities, as published in a service description, into a unified and integrated execution model. An agent must have at least one owner and an agent must support at least one notion of identity. This notion of identity is the **agent identifier** (AID) that labels an agent so that it may be distinguished unambiguously within the agent universe.*

---

<sup>39</sup>Nach der von Wooldridge getroffenen Definition lässt sich jedes Programm als ein Agent interpretieren.

### 3 Agentenparadigma-Systemtopologie

---

Der entscheidende Punkt dieser Definition ist die Festlegung, dass ein Agent als ein autonomer, kommunikativer Prozess mit nicht festgelegter Anzahl von Fähigkeiten und eindeutiger Kennung aufgefasst werden kann. Wichtig ist an dieser Stelle anzumerken, dass die kommunikativen Fähigkeiten sowohl die Interaktion mit anderen Agenten, mit Benutzer als auch mit eventuell vorhandener Hardware mit einbezieht. Gemäß obiger Definition wird der Agent als eine eigenständige Entität betrachtet und bietet ein hinreichendes Potenzial zur Spezifizierung, ohne jedoch dabei die Anwendbarkeit auf bestimmte Gebiete einzuschränken. Sie wird daher im Rahmen dieser Arbeit übernommen.

#### 3.1.2 BDI-AGENTENMODELL

Einer der am weitesten verbreiteten Ansätze zur Implementierung von Agenten bzw. Agentensystemen [100], [114] basiert auf dem vom Philosophen Michael Bratmann am Stanford Research Institute Mitte der 1980er Jahre entwickelten *Belief-Desire-Intention* (BDI) Modells zur Modellierung menschlichen Verhaltens und entstammt der Theorie der praktischen Schlussfolgerung [112]. Der konzeptionelle Rahmen des BDI-Modells wurde von Bratman und Kollegen durch die Einführung einer als *Intelligent Resource-bounded Machine Architecture* (IRMA) benannten Agentenarchitektur in [113] beschrieben.

Hauptaspekt dieses Modells ist der Ansatz, einem logisch operierenden System Merkmale zuzuordnen, welche ebenfalls im Rahmen einer Beschreibung des menschlichen Verhaltens zum Einsatz kommen. Das Zusammenspiel, oder besser gesagt, die Konfiguration dieser Merkmale, definiert man als einen mentalen Zustand des Systems. Prinzipiell besteht das BDI-Modell dazu aus drei Komponenten [86], [112], [113]:

1. **Beliefs (Annahmen):** Annahmen stellen die fundamentalen Informationsblöcke eines Agenten dar. Sie repräsentieren die Sicht bzw., sie stellen die Informationen des Agenten bezüglich seiner Umwelt dar und können durchaus ungenau, unvollständig und/oder auch veraltet sein.
2. **Desires (Wünsche):** Sämtliche Situationen oder Zustände, die ein Agent bewältigen bzw. potenziell erreichen möchte, werden als Wünsche interpretiert. Diese können durchaus im gegenseitigen Widerspruch existieren. Aufgrund der Semantik des Ausdrucks ist die Annahme berechtigt, dass das Vorhandensein von Wünschen keineswegs eine konkrete Handlung impliziert. Vielmehr können sie als Optionen eines Agenten verstanden werden, die bei Auswahl zu einer entsprechenden Handlung führen.
3. **Intentions (Intentionen, Zielvorgaben):** Intentionen sind die Ziele, denen sich ein Agent verpflichtet hat. Sie beinhalten das aktive Streben nach Bewältigung einer Situation bzw. der Erreichung eines Zustands. Sie können somit als manifestierte Wünsche interpretiert werden. Als Initiator einer Intention kann dabei eine von außen zugeführte Information bzw. Zielvorgabe dienen, in deren Abarbeitungsprozess gewisse Fakten geschaffen werden, welche die vorhandenen Wünsche in Intentionen umwandeln. Aber auch veränderte Annahmen, die im Rahmen einer Auffrischung der Wahrnehmungen eintreten (z. B. durch neue Sensorinformationen), können zu einem sich manifestierenden Wunsch führen.

##### 3.1.2.1 PRAKTISCHES SCHLUSSFOLGERN

Das dem BDI-Modell zugrunde liegende Paradigma zur Entscheidungsfindung wird als praktisches Schlussfolgern (*practical reasoning*) bezeichnet und ist gemäß [115] definiert als:

*Practical reasoning is a matter of weighting conflicting considerations for and against competing options, where the relevant considerations are provided by what the agent desires / values / cares and what the agent believes.*

---

### 3 Agentenparadigma-Systemtopologie

---

Anhand obiger Definition kann der Prozess des praktischen Schlussfolgerns in zwei separate, aufeinanderfolgende Aktivitäten aufspalten werden:

#### **PROZESS DER ERWÄGUNG (*DELIBERATION*)**

Der Prozess der Erwägung ist die Betrachtung der vorhandenen Informationen bzgl. der Umwelt<sup>40</sup> und der vorhandenen Wünsche und mündet in die Annahme einer Intention und beinhaltet somit die Verfolgung eines Ziels. Eine angenommene Eigenschaft der Intention ist, dass sie persistent ist. Einmal in Erwägung gezogen, wird versucht, das zugrunde liegende Ziel auch zu erreichen, wobei vorausgesetzt wird, dass die dazu notwendigen Bedingungen gültig sind.

Das Fehlen notwendiger Prämissen führt zwangsläufig zu einer Aufgabe des Ziels. In [86] werden verfolgte Intentionen auch als Zulassungsfiler interpretiert, da sie den möglichen Handlungsspielraum des Agenten einschränken.

#### **PROZESS DER ZIELGERICHTETEN SCHLUSSFOLGERUNG (*MEANS-END REASONING*)**

Der aus dem Gebiet der künstlichen Intelligenz stammende Begriff der Planung [116] beschreibt den Prozess einer Ablauffindung zwecks Erreichens eines vorgegebenen Zieles, welches durch die Annahme einer expliziten Intention vermöge des Prozesses der Erwägung getriggert wird.

In der Regel besteht ein Planungssystem aus:

1. Einem definierten Zustand (Ziel), welcher von einem Agenten zu erreichen ist.
2. Der Information des Agenten bzgl. seiner Umwelt.
3. Die dem Agenten zur Verfügung stehenden Aktionen bzw. Hilfsmittel zur Verfolgung von Intentionen bzw. Zielen.

Die Aufgabe des Systems besteht darin, ein durch einen zugrunde liegenden Algorithmus generierter Plan (Sequenz von Aktionen) zu erzeugen, der es dem Agenten erlaubt, eine zielgerichtete Operation durchzuführen. Das fundamentale Problem der Erstellung von erfolversprechenden Plänen, welche letztendlich als Programme aufzufassen sind, ist, dass deren Erzeugung in nichttrivialen Umgebungen beliebig kompliziert werden kann [116].

Eine Möglichkeit, dieser Problematik zu entgehen, ist, den Prozess der Plangenerierung aus der zielgerichteten Schlussfolgerung zu entfernen, wobei stattdessen als Ersatz bereits im Vorfeld generierte Pläne eingesetzt werden. Deren Vielfalt bzw. Anwendbarkeit entspricht dann der Fähigkeit eines Agenten, auf den ihm zugetragenen Informationen zu reagieren bzw. den Intentionen nachzugehen.

Dieser Schritt impliziert, dass der Agent die notwendige Fertigkeit besitzen sollte, die für eine erfolgreiche Operation notwendigen Pläne auswählen zu können. Man könnte somit sagen, dass die Fähigkeit zur Erstellung eines zielgerichteten Plans durch die Fähigkeit der geeigneten Auswahl von vorgegebenen Plänen zur Erfüllung einer Aufgabe ersetzt worden ist.

Das Prinzip der Prägung von Plänen ist eine klare Simplifikation des oben angesprochenen Planungsprozesses. Aufgrund der Vereinfachung hat dieses Vorgehen jedoch in den meisten real existierenden Agentensystemen Einzug gehalten [86], und auch das im Rahmen dieser Arbeit implementierte Agentensystem basiert auf der vorausgehenden Erstellung von Plänen zur Intensionsverfolgung.

#### **3.1.2.2 OPERATIONSZYKLUS DES BDI-MODELLS**

Der fundamentale Operationszyklus, welcher als Richtlinie einer Realisierung eines BSI-Modells auf Grundlage des praktischen Schlussfolgerns dient, liefert der folgende Pseudo-Code [84]:

---

<sup>40</sup>Diese Informationen können auf akquirierten Sensordaten basieren, auf Nachrichten anderer Systemelemente (z. B. zusätzliche Agenten) oder auch auf Interaktionen eventueller Benutzer.

### 3 Agentenparadigma-Systemtopologie

```
1.  $B \leftarrow B_0$ ; // Mit  $B_0$  als initiale Annahmen (Beliefs).
2.  $I \leftarrow I_0$ ; // Und  $I_0$  als initiale Intentionenmenge.

// Beginn der Hauptschleife (Lebenszyklus).
3. while true do
4.    $update(\rho)$ ; // Aktualisierung von  $\rho$ 
// (Umgebungswahrnehmung durch die Sensorik).
5.    $B \leftarrow brf(B, \rho)$ ; // Aktualisierung der Annahmen
// (Einbettung neuer Informationen in die
// Menge der Annahmen durch  $brf(B, \rho)$ ).
6.    $D \leftarrow options(B, I)$ ; // Test auf Erweiterung der Optionen.
7.    $I \leftarrow filter(B, D, I)$ ; // Prozess der Erwägung (Deliberation).
8.    $\pi \leftarrow plan(B, I, A)$ ; // Prozess der zielgerichteten
// Schlussfolgerung (Means-end reasoning).

// Beginn der Planausführung.
9.   while not ( $empty(\pi) \vee success(I, B) \vee impossible(I, B)$ ) do
10.     $\alpha \leftarrow head(\pi)$ ; // Entnahme der ersten Aktion  $\alpha$  aus  $\pi$ .
11.     $exec(\alpha)$ ; // Ausführung der Aktion  $\alpha$ .
12.     $\pi \leftarrow tail(\pi)$ ; // Entfernen der ausgeführten Aktion  $\alpha$  aus
// Plan  $\pi$ .
13.     $update(\rho)$ ; // Umgebungswahrnehmung durch Sensorik.
14.     $B \leftarrow brf(B, \rho)$ ; // Aktualisierung der Annahmen.

15.    if reconsider( $I, B$ ) then // Test auf notwendige Änderung der
// Intention aufgrund veränderter
// Annahmen.
16.       $D \leftarrow options(B, I)$ ; // Test auf Erweiterung der Optionen.
17.       $I \leftarrow filter(B, D, I)$ ; // Prozess der Erwägung.
18.    endif

19.    if not sound( $\pi, I, B$ ) // Überprüfung des aktuellen Plans
// auf Konsistenz mit Zielvorgabe.
20.       $\pi \leftarrow plan(B, I, A)$ ; // Prozess der erneuten zielgerichteten
// Schlussfolgerung im Falle einer
// unzureichenden Übereinstimmung von
// Plan und Zielvorgabe.
21.    end-if // Konsistenzprüfung.
22.  end-while // Planausführung.
23. end-while // Hauptschleife.
```

Zunächst bilden die Mengen  $B, D, I$  die Annahmen, Optionen und Intentionen, die dem Agenten zur Verfügung stehen. Die Hauptschleife (3.) - (23.) beginnt ihre Operation mit den dem Agenten initial zugänglichen Informationen bezüglich seiner Umwelt und Optionen ( $B_0, I_0$ ).

Die im nächsten Schritt (4.) aktualisierten Umgebungswahrnehmungen  $\rho$  werden mittels der Funktion  $brf(B, \rho)$  bearbeitet. Letztendlich entspricht die Funktionalität von  $brf(B, \rho)$  der Transformation neuer Sensorinformationen (erzeugt durch  $update(\rho)$ ) in die Menge der Annahmen  $B$ <sup>41</sup>.

<sup>41</sup>Beispiele hierfür sind die Aktualisierung von Positionsinformationen oder das Verändern der Anzahl wahrgenommener Objekte.

### 3 Agentenparadigma-Systemtopologie

---

In Abhängigkeit der Umgebungsveränderung werden Annahmen aufgefrischt und/oder neu erzeugt bzw. entfernt (5.).

Die Schritte (6.) und (7.) dienen der eventuellen Erweiterung ( $D \leftarrow \text{options}(B, I)$ ) der Optionenmenge sowie der sich daraus ergebenden Zielverfolgung ( $I \leftarrow \text{filter}(B, D, I)$ ). Die in (7.) angegebene Funktion  $\text{filter}(B, D, I)$  repräsentiert dabei den am Anfang des Abschnitts besprochenen Prozess der Erwägung.

Die Funktion  $\text{plan}(B, I, A)$  aus Schritt (8.) stellt den Prozess der zielgerichteten Schlussfolgerung dar. Anhand der zur Verfügung stehenden Umweltinformationen  $B$ , den zu verfolgenden Zielen  $I$  und den dem Agenten zugänglichen Aktionen  $A$  wird ein Plan  $\pi$  zur Verfolgung der Intention ausgewählt. Der Plan selbst besteht dabei aus einer definierten Abfolge von Aktionen, welche durch den Prozess der zielgerichteten Schlussfolgerung erzeugt wird<sup>42</sup>.

Die Ausführung des Plans geschieht unter Beachtung der zur Verfolgung notwendigen Bedingungen in den Schritten (9.) bis (22.). Neben der reinen Ausführung der Planelemente (Aktionen) wird weiterhin die Umwelt vermöge der zugänglichen Sensorik betrachtet (13.). Fundamentale Änderungen der Umgebung, welche auch durch Nachrichten anderer Agenten mitgeteilt werden können, führen unter Umständen zu einer Veränderung von möglichen Optionen und Bedingungen, sodass die momentane Zielvorgabe auf Konsistenz überprüft werden muss (16.), was möglicherweise zu einer neuen Planauswahl führt (19. - 20.).

Eine sich ergebende Problematik ist die Häufigkeit der Umweltüberprüfung und der damit verbundene eventuelle Planabbruch bzw. die Neuauswahl. Eine zu strikte Überwachung führt möglicherweise zu einer Operation, in der eine Zielvorgabe aufgrund einer sich häufig ändernden Umwelt selten oder nie erreicht wird. Eine zu langsame Reaktion auf Umweltveränderungen könnte jedoch auch zur Folge haben, dass die gewählte Intention, bzw. das zu verfolgende Ziel, nicht mehr mit den Informationen des Agenten bzgl. seiner Umgebung konsistent ist und sein Verhalten damit einer Fehlleistung entspricht.

#### 3.1.3 PRS - PROCEDURAL REASONING SYSTEM

Unter der Vielzahl der Realisationen des BDI-Modells ist das von Michael Georgeff und Amy Lansky am Stanford Research Institute entwickelte *Procedural Reasoning System* (PRS) [117], dessen prinzipielle Architektur in Abb. 21 dargestellt ist, eine besonders häufig gewählte Variante<sup>43</sup>.

Grundelemente des PRS sind gemäß des zugrunde liegenden BDI-Modells zunächst die Strukturen, die das Wissen über die Umwelt (*Beliefs*) beinhalten und die die momentan aktiven Absichten bzw. Ziele (*Intentions*) wiedergeben sowie die, die mögliche Optionen (*Desires*) aufzeigen.

Zu erwähnen ist, dass das PRS nicht zu den Systemen gehört, welches eine grundlegende Planung, basierend auf primären Prinzipien, durchführt. Vielmehr besitzt es eine Ansammlung von Plänen, welche bereits im Vorfeld der Operation erstellt werden und deren Struktur folgende Komponenten enthalten:

1. **Ein Ziel:** Enthält die Information, für was oder für welchen Zweck der Plan verwendet werden kann. Es entspricht der Post-Kondition des Plans.
2. **Ein Kontext:** Er entspricht den Prä-Konditionen, welche gegeben sein müssen, um den Plan ausführen zu können. Anzumerken ist, dass diese Bedingungen während der gesamten Planausführung ihre Gültigkeit beibehalten müssen.
3. **Ein Ausführungskörper:** Er enthält die notwendigen Ausführungsaktionen zur Erlangung des definierten Ziels.

---

<sup>42</sup>Wie bereits am Anfang des Abschnitts besprochen, reduziert sich der Prozess der zielgerichteten Schlussfolgerung auf eine Selektion des zur Erlangung des Ziels hilfreichsten Programms.

<sup>43</sup>Eine Übersicht sowie ein Vergleich diverser auf dem BDI-Model basierenden Agentenimplementierungen findet sich in [114].

---

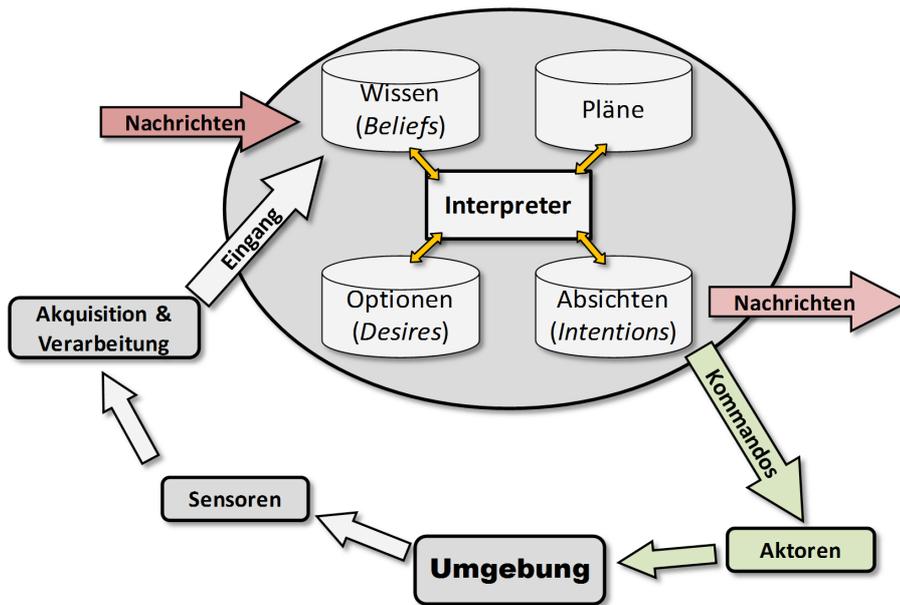


Abbildung 21: Procedural Reasoning System (PRS) Architektur.

Hauptkomponente des PRS ist der Interpreter. Seine Aufgabe ist es, vorhandene Pläne sinngemäß abzuwickeln, wobei auf die in den Wissens- und Optionsstrukturen gespeicherten Informationen zurückgegriffen wird. Es besteht dabei durchaus die Möglichkeit, dass während der Bearbeitung der Ausführungsaktionen im Ausführungskörper der Pläne zusätzliche Intentionen bzw. Ziele erzeugt werden. Die Verwaltung aller aktiver, d. h. vorhandener oder neu erzeugter Intentionen, geschieht mittels einer dedizierten Intentionsstruktur.

Die Erzeugung von zusätzlichen Intentionen im Rahmen einer Zielverfolgung durch den Agenten ist lediglich ein möglicher Effekt. Weiterhin besteht die Möglichkeit, dass im Laufe einer Abarbeitung eines Plans gewisse Ereignisse bzw. Informationen auftreten, die von besonderem Interesse sind und vermittelt werden müssen. Dieses geschieht durch Weiterleitung von Nachrichten unter Verwendung einer definierten Kommunikationsinfrastruktur (Abschnitt 3.1.4). Endpunkte können dabei zunächst andere im Gesamtsystem teilnehmende Agenten sowie Schnittstellen zu Benutzern sein. Im Falle proaktiver Agenten, welche die Möglichkeit besitzen, aktiv in die Umwelt einzugreifen, sind mögliche Endpunkte die vorhandenen Aktoren, welche eine Umsetzung der erhaltenen Informationen entsprechend durchführen<sup>44</sup>.

#### 3.1.3.1 PRS-OPERATIONSZYKLUS

Zu Beginn eines vollständigen Operationszyklus verfügt der Agent über eine gewisse Anzahl von Plänen sowie Annahmen (*Beliefs*) über seine Umwelt in Form von Relationen. Aufgrund seiner zielgerichteten Konzeption besitzt er zusätzlich ein a priori definiertes Ziel (*top-level-goal*), welches beim Start als fundamentale Intention dient und dessen Verfolgung er sich direkt widmet. Dazu trägt er dieses zunächst in die Intentionsstruktur ein, welche in Form eines Stacks implementiert ist und womit sichergestellt wird, dass das jeweils aktuelle Ziel an oberster Stelle platziert wird.

Im nächsten Schritt wird die Planbibliothek nach Plänen durchsucht, welche im Hinblick auf die Erlangung des Grundziels verwendet werden können. Für jeden infrage kommenden Plan wird anschließend eine Kontextüberprüfung durchgeführt. Getestet wird, ob die Bedingungen zur Ausführung des Plans bzw. seine Prä-Konditionen mit den vorhandenen Annahmen übereinstimmen. Im Falle des Vorhandenseins mehrerer geeigneter Pläne setzt der Prozess der

<sup>44</sup>Im Falle eines Roboters bestünde eine solche Umsetzung z. B. in der Veränderung der Richtung oder Geschwindigkeit der Bewegung.

### 3 Agentenparadigma-Systemtopologie

---

Erwägung ein. Seine Aufgabe ist es, den zur Erreichung des Ziels sich am besten eignenden Plan auszuwählen.

Ergibt sich aus Verfolgung des Grundziels die Erzeugung eines zusätzlichen Ziels (Unterziel), wird analog zum obigen Fall zunächst eine Menge anwendbarer Pläne zusammengestellt und eine Kontextprüfung durchgeführt. Liefert der angeschlossene Erwägungsprozess einen geeigneten Plan, wird das erzeugte Unterziel als das aktuell zu verfolgende betrachtet und in die Intentionsstruktur eingetragen bzw. auf den Stack gelegt.

Die Beendigung einer Zielverfolgung tritt ein, wenn alle notwendigen Aktionen des entsprechenden Plans bearbeitet worden sind. Beendet wird die Verfolgung ebenfalls für den Fall, dass Bedingungen eintreten, welche den Kontext des Plans verändern und dessen Ausführung nicht mehr erlauben. Da das Ziel keiner Verfolgung mehr unterliegt, wird es aus der Intentionsstruktur entfernt und die Verfolgung des vorherigen in die Intentionsstruktur eingetragenen Ziels wird erneut aufgenommen. Entspricht die zu entfernende Intention dem am Anfang der Operation definierten Grundziel, führt das zusätzlich zur Beendigung des gesamten PRS-Operationszyklus.

#### 3.1.4 AGENTENKOMMUNIKATION

Die Agentenkommunikation bildet den zentralen Austauschmechanismus zwischen einzelnen Agenten. Er unterscheidet sich vom herkömmlichen Datenaustausch dadurch, dass er eine Abstraktion der spezifischen Datenformate verwendet und somit eine Möglichkeit der Kommunikation zwischen diversen Systemen bietet. Die hier betrachteten Formen der Agentenkommunikation haben ihren theoretischen Ausgangspunkt in der Sprachakttheorie [118], [119], in der jede Kommunikationseinheit (Nachricht) als eine Aktion (Sprachakt) definiert wird. Zwei bekannte Umsetzungen der Übertragung von Sprachakten sind KQML und FIPA-ACL, auf die im Folgenden kurz eingegangen wird. Für eine Einführung in die Theorie der Sprachakte sei an dieser Stelle auf die Publikation von Cohen und Perrault [120] hingewiesen.

##### 3.1.4.1 KQML

Die *Knowledge Query and Manipulation Language* (KQML) wurde im Rahmen des *Knowledge Sharing Effort* (KSE) Projekts [120] entwickelt und stellt einen ersten Versuch dar, eine praktisch benutzbare Agentenkommunikationssprache zu definieren, die Elemente der Sprachakttheorie enthält bzw. vermitteln kann [86]. Im Wesentlichen ist es eine auf der Wissensebene angesiedelten Sprache (*knowledge-level messaging*), welche durch eine Anzahl von sogenannten Performativen definiert wird und in der Lage ist, Aussagen bzw. Informationen zu kommunizieren. KQML besitzt eine vordefinierte Anzahl an Performativen, deren Bedeutung über Operatoren auf einer Wissensbasis (*Virtual-Knowledge-Base*) definiert wird. Zusätzlich dazu existieren Performative, welche das Management der Agenten sowie die Kommunikationsinfrastruktur betreffen. Des Weiteren bietet KQML die Möglichkeit der Einbettung einer Inhaltssprache, die in der Regel von der jeweiligen Anwendung abhängig ist und keiner allgemein definierten Syntax unterliegt.

##### 3.1.4.2 FIPA-ACL

Eine der KQML verwandte, standardisierte Sprache zur Agentenkommunikation, welche im Rahmen dieser Arbeit Verwendung findet, stellt die *Agent Communication Language* (ACL) der *Foundation for Intelligent Physical Agents* (FIPA) dar, deren Definition 1999 verabschiedet wurde [111]. Der Unterschied zu KQML besteht in der Menge der definierten Performativen sowie der Semantik. Das zugrunde liegende Ziel dieser Standardisierung war eine Vereinfachung der Menge der Performative sowie deren Bedeutung [86]. Analog zu KQML bietet die FIPA-ACL ebenso die Möglichkeit der Einbettung einer von der jeweiligen Applikation abhängigen Inhaltssprache via definierter Schlüsselwörtern. Für eine Beschreibung ihrer Performative sowie einiger beispielhaften Nachrichten sei der Leser an dieser Stelle auf [84] verwiesen.

### 3.1.5 IMPLEMENTIERTES AGENTENSYSTEM

Das im Rahmen dieser Arbeit implementierte Agentensystem basiert auf dem von der University of Michigan entwickelte und auf dem BDI-Modell fußenden *University of Michigan-Precedural Reasoning System* (UM-PRS) [122]. Es handelt sich um eine C++-Implementierung des in Abschnitt 3.1.3 beschriebenen prozeduralen Schlussfolgerungssystems, dessen Dokumentation, Analyse sowie explizite Anwendung in [123]-[125] gefunden werden kann.

Der Grund für die Wahl dieses Systems liegt zunächst in der Tatsache begründet, dass die gewählte Implementierungssprache des UM-PRS mit der für die Entwicklung der MultiCam-Schnittstelle übereinstimmt, was wiederum eine homogene Entwicklungsstruktur ermöglicht. Durch die Wahl gleicher Implementierungssprachen ist ein zusätzliches Interface zur Kommunikation zwischen dem UM-PRS und der Sensorik nicht notwendig, womit eine enge Verzahnung ermöglicht wird, die wiederum Einfluss auf die Ausführungsgeschwindigkeit des Gesamtsystems hat.

Der größte Vorteil des UM-PRS ist jedoch die in seinem Interpreter implementierte Fähigkeit der Integration von sogenannten primitiven Funktionen in die Pläne des Agenten. Sie bietet die Möglichkeit des definierten Zugriffs auf externe Module bzw. Bibliotheken<sup>45</sup>. Durch diesen Mechanismus ist es möglich, die dem Agenten zur Verfügung stehenden Aktionen beliebig zu erweitern (Abschnitt 3.1.6). Grundsätzlich setzt sich das UM-PRS aus folgenden Konstituenten zusammen:

#### WELTMODELL

In der gewählten Implementierung wird das Wissen bzgl. der Umgebung als Weltmodell des Agenten aufgefasst. Im Wesentlichen ist es eine Ansammlung von Fakten, welche durch Relationen dargestellt bzw. definiert werden. Wie bereits im Operationszyklus des BDI-Modells angegeben, enthält jedes Weltmodell initiale Werte, welche zum Beginn der Operation dem Agenten durch den Benutzer zur Verfügung gestellt werden. Zusätzliche Informationen können im Verlauf der Zielverfolgung durch Planbearbeitung oder durch Informationen von anderen Agenten zur Laufzeit aufgenommen respektive hinzugefügt werden.

#### ZIELE

Ziele sind Zustände des Weltmodells, welche durch den Agenten zu erreichen sind und werden ebenfalls durch Relationen definiert. Im Wesentlichen wird ein Ziel durch eine konkrete Konfiguration einer Menge von Relationen, welche im Weltmodell definiert sind, repräsentiert. Grundsätzlich existieren zwei Zielarten: Top-Level-Ziele, welche den fundamentalen Operationsmodus des Agenten definieren, und Unterziele, welche im Laufe der Abarbeitung eines Plans erreicht werden müssen.

#### PLÄNE

Pläne entsprechen den Methoden, die dem Agenten zur Verfügung stehen, um Ziele zu verfolgen und zu erreichen. Im Rahmen der Implementierung werden diese als *Knowledge Areas* (KA) bezeichnet und besitzen eine wohldefinierte Struktur bestehend aus:

1. Einer Zieldefinition. Sie repräsentiert den Verwendungszweck des Plans.
2. Einem Kontext. Er enthält die notwendigen Bedingungen, die erfüllt sein müssen, um den Plan ausführen zu können.
3. Einem Plankörper. Dieser wird durch eine Menge an Aktionen definiert und kann als der prozedurale Kern des Plans angesehen werden. Im Verlauf einer Ausführung können zusätzliche Unterziele verfolgt, Fakten hinzugenommen oder gelöscht werden sowie Nachrichten an andere Objekte gesendet werden. Der Plankörper definiert die Dynamik

---

<sup>45</sup>Windows: *dynamic link libraries* (\*.dll). Linux: *shared objects* (\*.so).

### 3 Agentenparadigma-Systemtopologie

des Agenten in Bezug auf die Verfolgung des Ziels. Ein beispielhafter Agentenplan findet sich in Abb. 22.

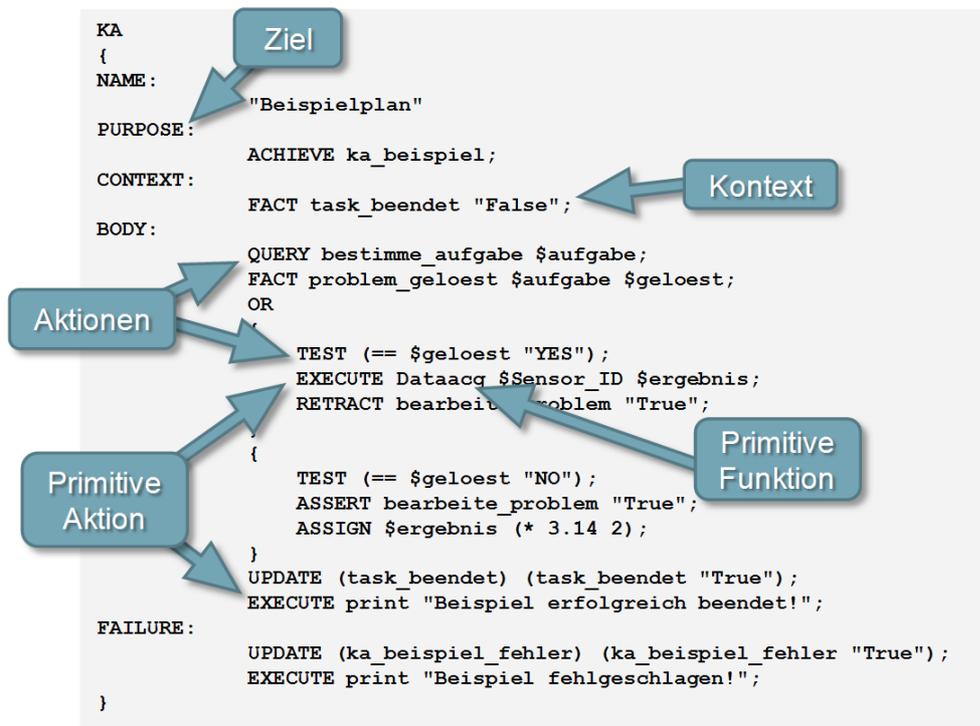


Abbildung 22: Beispielhafter Agentenplan.

Der angegebene Plan dient lediglich der Darstellung seines prinzipiellen Aufbaus sowie der Syntax unterschiedlicher Aktionen. Dementsprechend enthalten die Aktionen selbst keinen tieferen Sinn (z. B. `ASSIGN $ergebnis (* 3.14 2);`).

#### AKTIONEN

Aktionen sind die grundlegenden Bausteine der Agentenoperation und können in zwei Klassen unterteilt werden:

1. Primitive Aktionen können direkt ausgeführt werden. Primitive Funktionen sind Bestandteile primitiver Aktionen. Im Rahmen der gewählten Implementierung entspricht eine primitive Funktion einem Aufruf in eine interne oder externe Bibliothek.
2. Nichtprimitive Aktionen sind z. B. zu erreichende Unterziele und/oder Anfragen, auf die gewartet werden muss, Relationsprüfungen bzw. Tests sowie das Hinzufügen oder Entnehmen von Weltmodellrelationen.

#### INTENTIONSSTRUKTUR

Die Intensionsstruktur agiert als ein Laufzeit-Stack des PRS-Systems. Auf ihr werden die zu erreichenden Haupt- sowie eventuell zu verfolgende Unterziele abgelegt. Sie enthält die Information, welches Ziel gerade verfolgt wird und welche Intentionen noch zu verfolgen sind.

#### INTERPRETER

Der Interpreter ist die zentrale Instanz des PRS und ist zuständig für die kontrollierte Operation des gesamten Systems durch Abarbeitung der in den Plänen angegebenen Aktionen (Abb. 22). In

### 3 Agentenparadigma-Systemtopologie

ihm ist die komplette Logik untergebracht, die für eine Interaktion mit anderen Systemen verantwortlich ist. Das beinhaltet sowohl die Interaktion mit externen Modulen<sup>46</sup> via primitiven Funktionen als auch die Kommunikation mit anderen Agenten oder Benutzern mittels einer zugrunde liegenden Kommunikationsplattform, welche ebenfalls mittels primitiver Funktionen angesprochen wird.

Eine weitere fundamentale Aufgabe des Interpreters ist die der Erwägung. Wann immer sich die Informationen des Weltmodells ändern oder die Liste der Ziele sich ändert, ist es die Zuständigkeit des Interpreters einen passenden Plan zu selektieren, um auf die veränderten Gegebenheiten zu reagieren. Eine detaillierte Beschreibung der Funktionalität des Interpreters findet sich in [123].

#### 3.1.6 AGENTENARCHITEKTUR

Die im Rahmen dieser Arbeit eingesetzte und auf dem UM-PRS aufbauende Agentenarchitektur besteht aus drei Grundmodulen (Abb. 23):

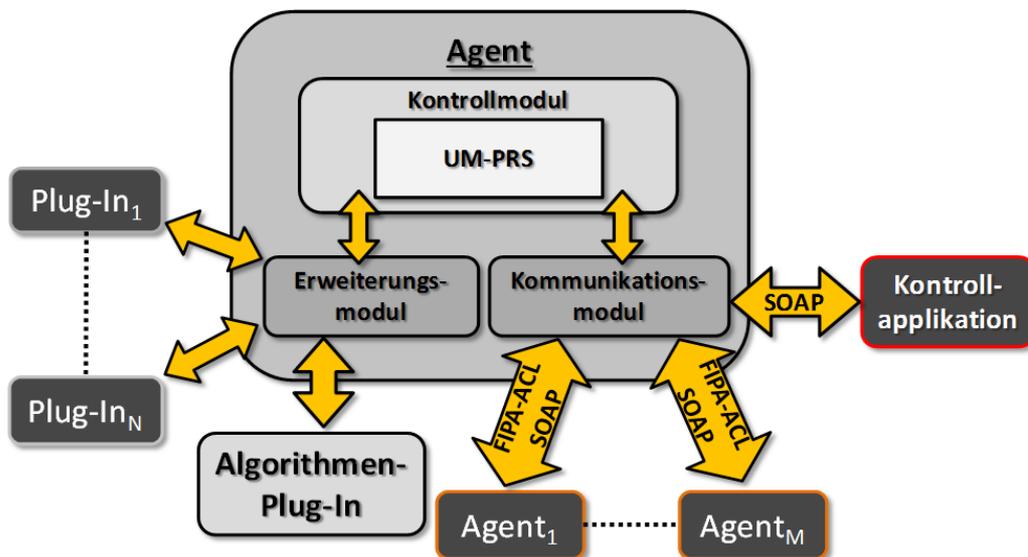


Abbildung 23: Skizze der implementierten Agentenarchitektur.

#### KONTROLLMODUL

Das Kontrollmodul ist für die fundamentale Operation des Agenten zuständig. In der Anfangsphase übernimmt es die Aufgabe, den Agenten mittels einer Konfigurationsdatei in einen definierten Zustand zu bringen. Dieses beinhaltet die Initialisierung des Erweiterungs- als auch des Kommunikationsmoduls mit den in der Konfigurationsdatei angegebenen Werten. Im laufenden Betrieb ist es für die Abarbeitung zugeführter Steuerkommandos zuständig, welche ihm z. B. durch eine Kontrollapplikation via *Simple Object Access Protocol* (SOAP) zugesendet werden können<sup>47</sup>.

#### KOMMUNIKATIONSMODUL

Das Kommunikationsmodul besitzt eine zweifache Aufgabe. Die Erste besteht darin, sowohl Steuerbefehle, z. B. von benutzergesteuerten Kontrollapplikationen, entgegenzunehmen und diese dem Kontrollmodul in definierter Weise zur Verfügung zu stellen als auch Ergebnisse eventuell durchgeführter Berechnungen an fordernde Instanzen, wie einer Visualisations- oder Kontrollapplikation, zu liefern.

<sup>46</sup>Die Verwendung externer Module ermöglicht die Ansteuerung bzw. Verwendung von beliebiger Hardware.

<sup>47</sup>Eine solche Kontrollapplikation kann sowohl auf dem lokalen als auch auf einem entfernten Rechner angesiedelt sein.

### 3 Agentenparadigma-Systemtopologie

---

Die zweite Aufgabe ist es, den Agenten eine gegenseitige Kommunikation zu ermöglichen und ihnen eine standardisierte Kommunikationsplattform zur Verfügung zu stellen. Als fundamentale Übertragungsschicht dient das TCP/IP-Protokoll. Dementsprechend ist jeder in einem System befindliche Agent durch eine eindeutige IP-Adresse gekennzeichnet. Darüber aufgesetzt ist das SOAP, auf der sämtliche Datenübermittlung basiert. Diesbezüglich besitzt jeder Agent einen dedizierten Serverthread, der für die Beantwortung von Anfragen via SOAP zuständig ist. Anfragen seinerseits sowie Übermittlung von Informationen an eine Kontrollapplikation, z. B. zur Visualisation von Informationen, werden vermöge der inhärenten Client-Fähigkeiten des Agenten durchgeführt.

Im Falle einer Kommunikation via FIPA-ACL-Nachrichten werden diese in SOAP-Mitteilungen eingebettet und an den entsprechenden Agenten gesendet. Anzumerken ist, dass die Nachrichtenübermittlung unter Verwendung der FIPA-ACL optional ist. Bedingt durch die hierarchische Kapselung der Informationen (Abb. 24) besteht durchaus die Möglichkeit der Versendung von Daten und Events unter lediglicher Verwendung von SOAP-Nachrichten und führt aufgrund der verringerten Overheads zu einer Reduktion der zur Dekodierung der Informationen notwendigen Rechenlast.

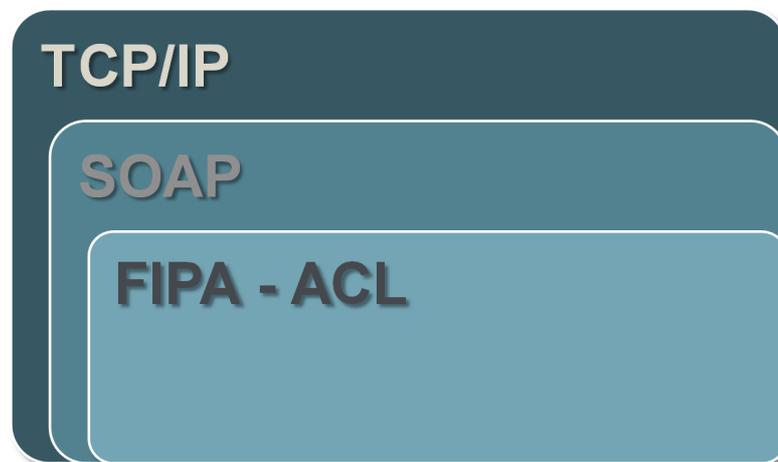


Abbildung 24: Einbettung der verschiedenen Nachrichtenebenen.

#### ERWEITERUNGSMODUL

Das Erweiterungsmodul dient der Administration der vorhandenen funktionalen Erweiterungen (Plug-Ins), welche den funktionalen Rahmen des Agenten definieren. Es besitzt eine standardisierte Schnittstelle, an der die zusätzlichen Module andocken können, um ihre Funktionalität dem Agenten zur Verfügung zu stellen. Mittels dieser Schnittstelle ist es weiterhin möglich, aus den Plänen des UM-PRS via den bereits erwähnten primitiven Funktionen direkt auf die Erweiterungsmodule bzw. deren Funktionen zuzugreifen. Plug-Ins bieten somit eine einfache Möglichkeit der Einbindung beliebiger Hard- und Software in die Infrastruktur des Agenten und sind als dynamisch zur Laufzeit zuladbare Bibliotheken implementiert.

Eine besondere Stellung nimmt das Algorithmenmodul ein. Es enthält die notwendige Intelligenz bzw. die notwendigen Algorithmen, um auf den Daten der Plug-Ins zu operieren. Analog zu den Plug-Ins zur Erweiterung der Funktionalität besitzt es ebenfalls eine standardisierte Schnittstelle, sodass sowohl das UM-PRS als auch das Kontrollmodul auf die beinhaltete Algorithmik zugreifen kann. Eventuelle Parameter, welche der Algorithmik zur Verfügung zu stellen sind, befinden sich XML-codiert in einer dedizierten Konfigurationsdatei. Aufgrund der Vielfalt der möglichen Implementierungen ist deren interne Struktur lediglich bis zu einem gewissen Grad standardisierbar, was aber keine Einschränkung darstellt, da der Inhalt der Konfigurationsdatei in der Regel von der jeweilig durchgeführten Realisierung der Algorithmik bestimmt wird und diese

### 3 Agentenparadigma-Systemtopologie

---

naturgemäß die Informationen zur Prozessierung der in der Konfigurationsdatei enthaltenen Informationen besitzt.

Eine Verknüpfung des Datenflusses zwischen den funktionalen Erweiterungen und den Algorithmen wird durch die Pläne des Agenten, unter Verwendung entsprechender primitiver Aktionen, vorgenommen. Ebenfalls in den Plänen codiert ist die Versendung der durch die Algorithmik erzeugten lokalen Informationen an den Fusionsknoten. Der Vorteil, welcher sich durch eine Auslagerung der Algorithmik in eine Bibliothek ergibt, ist, dass sich im Falle eines Updates oder Erweiterung die entsprechende Datei sehr einfach austauschen lässt.

Die Separation von Datenakquisition und Algorithmik erlaubt eine relativ simple Erweiterung bzw. Veränderung der Funktionalität. Während die Hinzunahme zusätzlicher Informationen durch die Verwendung weiterer Plug-Ins ermöglicht wird, kann eine darauf aufbauende Algorithmik unabhängig davon in Form eines angepassten Algorithmen-Plug-Ins zur Verfügung gestellt werden.

#### 3.1.7 KONSTITUTION DER VERWENDETEN AGENTEN

##### 3.1.7.1 MULTISENSORKNOTEN (MSK)

Informationsliefernde Agenten bzw. Multisensorknoten (MSK) übernehmen gemäß der in Abschnitt 1.2 diskutierten Modellierung die Aufgabe der Erzeugung sämtlicher Informationen zur Lokalisation, Verfolgung sowie Klassifikation von Objekten. Hardwaremäßig besteht ein im Rahmen dieser Arbeit verwendeter MSK aus zwei Komponenten:

1. MultiCam, zuständig für die Datenakquisition.
2. Ein dedizierter Rechner, zuständig für Datenprozessierung und Kommunikation.

Die zu einer Operation notwendige Software eines MSK, implementiert auf einem Windows-System, entspricht der in Abb. 23 dargestellten Agentenarchitektur und setzt sich zusammen aus:

1. **Agentenimplementierung.** Die Implementierung des BDI-Agenten setzt sich zusammen (Abschnitt 3.1.6) aus einem Kontroll-, einem Kommunikations- sowie einem Erweiterungsmodul, welche mittels einer fest definierten Konfigurationsdatei<sup>48</sup> initialisiert werden. In ihr enthalten sind zunächst die Informationen bezüglich der Topologie des Gesamtsystems, bestehend aus den eindeutigen Identitäten aller an der Operation beteiligten Einheiten in Form einer ID-Nummer und IP-Adresse. Zusätzlich dazu beinhaltet die Konfigurationsdatei die Informationen bzgl. der zu verwendenden funktionalen Erweiterungen und der einzusetzenden Algorithmik. Ebenfalls in der Konfigurationsdatei enthaltend sind die Informationen über die zu verwendende primäre Zielvorgabe (*top-level-goal*) des Agenten sowie über sein initiales Weltmodell<sup>49</sup>.
2. **Funktionale Erweiterungen (Plug-Ins):** Die für eine sinnvolle Operation notwendigen funktionalen Erweiterungen bzw. Plug-Ins in Form von zur Laufzeit dynamisch zuladbaren Bibliotheken.
3. **Algorithmen Plug-In:** Die zur Prozessierung der Daten notwendige Algorithmik, welche ebenfalls als eine zur Laufzeit dynamisch zuladbare Bibliothek implementiert ist.

Die im Rahmen dieser Arbeit durchgeführte Lokalisation, Verfolgung und Klassifikation von Objekten ist unter Verwendung zweier unterschiedlicher Plug-Ins durchgeführt worden. Während das MultiCam-Plug-In dazu verwendet wurde, synchronisierte Videos zu erzeugen, wurde das MultiCamVideo-Plug-In zur Akquisition dieser Daten eingesetzt.

---

<sup>48</sup>Fest definiert bedeutet hierbei, dass die Konfigurationsdatei einen festgelegten Namen sowie eine feste Position relativ zu der Agentenimplementierung (ausführbare Datei) besitzt.

<sup>49</sup>Im Wesentlichen besteht diese Information aus dem Namen der Datei. Diese enthält die notwendigen Pläne und Relationen zur Verfolgung des Grundziels und wird zu Beginn der Operation geladen und vom Parser übersetzt.

---

#### 3.1.7.2 FUSIONSKNOTEN (FSK)

Der sowohl für die Fusion der Daten als auch für die Verfolgung von Objekten auf Grundlage der fusionierten Informationen zuständige Agententyp besitzt keine Sensorik zur Akquisition von Informationen. Dementsprechend besteht die Hardware eines FSK lediglich aus einem dedizierten Rechner. Wie auch im Falle der MSK setzt sich der Softwareteil aus einer Agentenimplementierung sowie einem Algorithmen-Plugin zusammen, welcher die zur Prozessierung der MSK-Daten notwendige Algorithmik beherbergt. Aufgrund der Tatsache, dass der FSK keine Sensorik beherbergt, sind zusätzlichen Plug-Ins nicht notwendig.

Zusätzlich zu den Informationen bzgl. der Netzwerktopologie, der zu verwendenden Algorithmik sowie der Zielvorgabe des Fusionsknotens, enthält die Konfigurationsdatei des FSK die Information bezüglich der Positionen aller MSK-Koordinatensysteme relativ zu dem globalen, welches der FSK als Grundlage der Darstellung verwendet. Anhand dieser Daten ist es dem FSK möglich, die auf den jeweiligen Koordinatensystemen der MSK basierten lokalen Informationen in das globale Koordinatensystem zu transformieren (Abschnitt 3.2.2). Die Positionsinformationen der MSK sind während der Aufstellung des Gesamtsystems zu erzeugen (Abschnitt 1.2), wobei die Position des globalen Koordinatensystems beliebig gewählt werden kann. Wichtig zu erwähnen ist, dass grundsätzlich der FSK die Schnittstelle zum Gesamtsystem repräsentiert, und somit er es ist, welcher die erzeugten Informationen an eventuell nachgeschaltete Fremdsysteme weiterreicht.

Zwecks Minimierung der erforderlichen Hardware ist es weiterhin möglich, die Funktionalitäten des FSK auf einen ausgewählten MSK zu übertragen, welcher dann die Aufgabe der synchronisierten Akquisition und Fusion der Informationen übernimmt. Durch Einbettung der FSK/MSK-Kombination in eine Kontroll- bzw. Visualisationsapplikation, welche direkten Zugriff auf die Informationen des FSK besitzt, lässt sich weiterhin der Kommunikationsaufwand minimieren und entspricht der im Rahmen dieser Arbeit gewählten Vorgehensweise.

#### 3.1.8 ZEITLICHE SYNCHRONISATION

Die zeitliche Synchronisation aller MSK sowie des FSK wird durch die Verwendung des *Network-Time-Protokolls* (NTP) [174] erreicht. Der im Rahmen dieser Arbeit verfolgte und bereits im POmSe-Projekt erfolgreich getestete Ansatz beinhaltet die Verwendung eines MSK oder des FSK als NTP-Server, wobei die restlichen Knoten den NTP-Clients entsprechen. Verwendet wird eine frei verfügbare Server- [172] und Client-Software [173].

Im Allgemeinen ist die Konfiguration der NTP-Client-Server-Topologie ein Schritt, der einmalig zu Beginn der Systemoperation durchgeführt werden muss. Eine zusätzliche Administration ist im Rahmen einer zeitlichen Synchronisierung nicht mehr notwendig. Durch die Verwendung eines isolierten Netzes kann in der Regel ein zeitlicher Versatz der teilnehmenden Systeme zum NTP-Server von ca. 1 ms erreicht und gehalten werden.

### 3.2 INFORMATIONSFUSION

Ziel der Multisensor-Datenfusion ist es, durch Kombination bzw. Zusammenführung der einzelnen Daten einen messbaren Informationsgewinn zu erhalten [87], [92]. Des Weiteren bietet die Fusion der Daten mehrerer Quellen die Möglichkeiten der Minimierung von Messfehlern, etwa durch redundante Anordnungen, der Erhöhung der Gewissheit bezüglich des Vorhandenseins eines Objekts unter Zuhilfenahme verschiedener Modalitäten sowie der Reduzierung der Unvollständigkeit der Informationen in Bezug auf die Umgebung durch die Benutzung verteilter Sensoren mit verschiedenen Sichtweisen auf die Szene.

Eine Übersicht bzgl. des prinzipiell durch Datenfusion zu erwartenden Nutzens ist in Tab. 4, Anhang F, angegeben. Applikationen sowie Projekte zum Thema Multisensor-Datenfusion finden sich in Tab. 5.

### 3.2.1 FUSIONSTYPEN

Gemäß [91] kann die Fusion von Sensordaten in drei Grundtypen eingeteilt werden:

1. **Komplementäre Fusion:** Fusion von Daten mehrerer verteilter Sensoren (z. B. CCD-Kameras, IR-Kameras, etc.) mit unterschiedlichen Ausrichtungen bzw. Blickwinkeln. Da jeder Sensor seine Teilinformation bzgl. der Szene liefert, verringert dieser Fusionstyp die Unvollständigkeit bzgl. der Umwelt.
2. **Konkurrierende Fusion:** Fusion von Daten von verschiedenen Quellen mit gleicher Ausrichtung bei gleichzeitiger Beobachtung. Durch das Vorhandensein multipler Repräsentationen eines Objekts oder einer Szene kann eine erhöhte Aussagekraft bzgl. der gesuchten Eigenschaften bzw. Signatur (z. B. Position, Geschwindigkeit) getroffen werden. Diese Art der Fusion dient primär zur Verringerung von fehlerhaften und unsicheren Daten. Schätzungen gesuchter Parameter lassen sich durch Fusion mehrerer Datenquellen in Hinblick auf ihre Genauigkeit erhöhen.
3. **Kooperative Fusion:** Fusion von Daten verschiedener Quellen mit bestehenden gegenseitigen Abhängigkeiten. Um Beobachtungen bzw. Rückschlüsse zu ziehen, werden alle Informationen benötigt. Beispielhaft sei hier die Fusion von Daten eines Stereo-Kamerasystems genannt. Der Vorteil eines kooperativ fusionierenden Systems ist die Möglichkeit, eventuell multimodale Daten in einer geeigneten Weise zu kombinieren, um Informationen zu erhalten, welche durch die Verwendung lediglich einzelner Sensoren nicht zugänglich sind.

Im Allgemeinen wird in einem implementierten System eine Mischung der oben angegebenen Fusionstypen anzutreffen sein. Das im Rahmen dieser Arbeit verwendete Verfahren beinhaltet den Einsatz kooperierender Teilsysteme, welche eine Sensorik mit teilweise überlappendem Sichtbereich verwendet.

Aufgrund der unterschiedlichen Ausrichtungen ergibt sich bei der Konsolidierung der Teilinformationen eine Fusionierung sowohl im komplementären als auch im konkurrierenden Sinn (Abschnitt 5.3). Des Weiteren wird gezeigt (Abschnitt 4.3), dass durch kooperative Fusion von 2D- und 3D-Daten eine Steigerung der Performanz der verwendeten Algorithmen zur Personenklassifikation erreicht werden kann.

### 3.2.2 DATENANGLEICHUNG

Eine grundsätzliche Notwendigkeit, welche sich im Rahmen einer Fusion von Informationen verschiedener Quellen ergibt, ist ihre Darstellung in einer übergreifenden, datenangleichenden Form [91].

Der im Rahmen dieser Arbeit verfolgte Ansatz basiert auf der Darstellung der durch die MSK erzeugten Informationen in einem globalen Referenzsystem, wobei gilt:

1. Informationen wie Position und Geschwindigkeit eines Objekts werden aus der Sicht eines MSK zunächst bezogen auf sein lokales Koordinatensystem erzeugt.
2. Informationen des Gesamtsystems werden durch den FSK, nach Transformation aller MSK-Informationen, in einem globalen Referenzsystem dargestellt.

Einen schematischen Überblick des verwendeten Fusionsprozesses liefert Abb. 25.

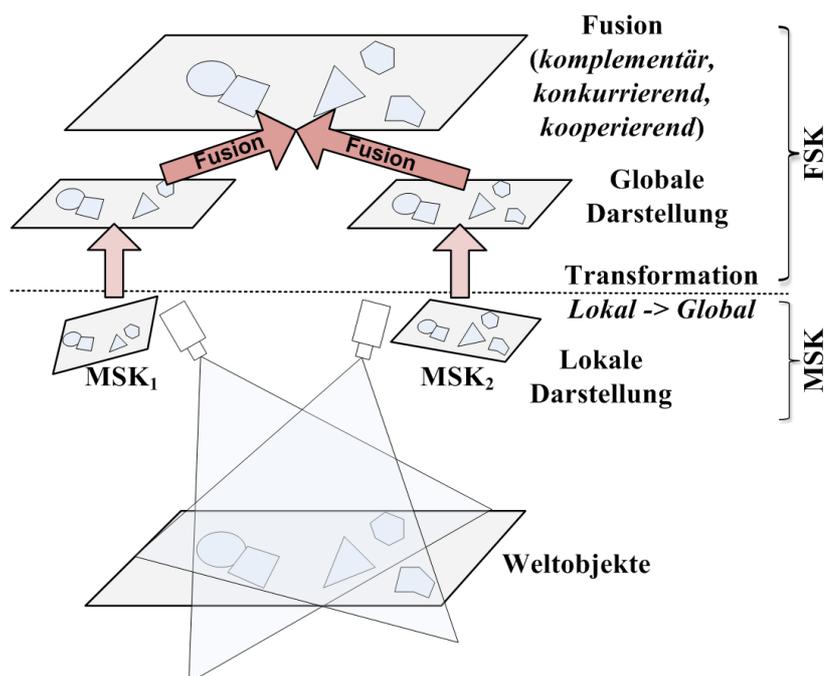


Abbildung 25: Fusion auf Grundlage angeglicher Daten.

#### 3.2.3 JDL DATENFUSIONSPROZESS-MODELL

Aufgrund der Komplexität eines Gesamtprozesses der Lokalisation, Verfolgung und Klassifikation von Objekten ist es sinnvoll, verschiedene Stufen der Informationsverarbeitung einzuführen. Naturgemäß befinden sich die Ebenen mit geringem Schlussfolgerungspotenzial, wie die Akquisition der Grundinformationen, am Anfang eines jeden Prozesses. Nachgeschaltete Bearbeitungsstufen ergeben zusätzliche Ebenen mit einer gesteigerten Informationsdichte und erlauben ein höheres Maß an Deduktion (Abb. 26).

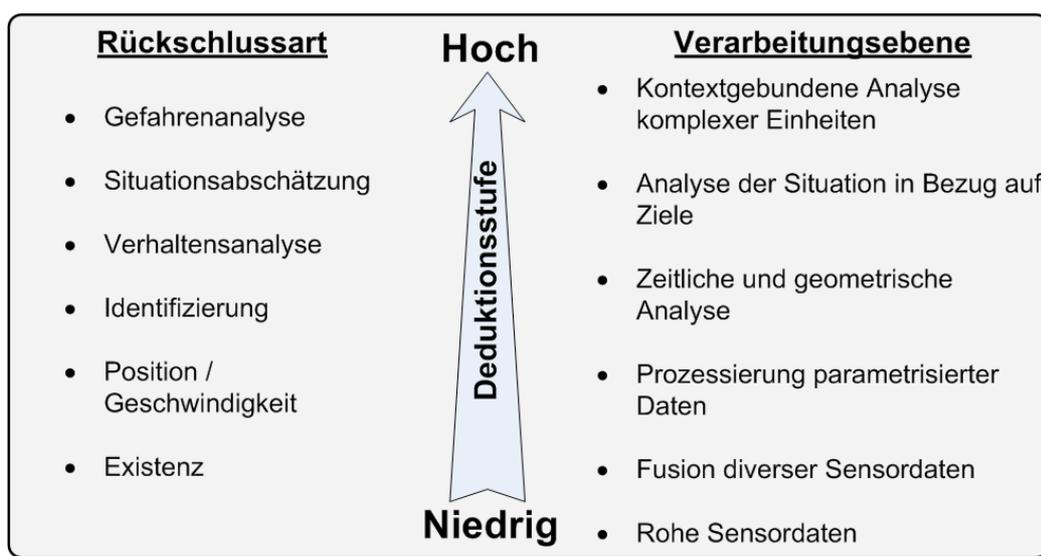


Abbildung 26: Hierarchie der Fusionsrückschlüsse [87].

### 3 Agentenparadigma-Systemtopologie

Während niedrige Deduktionsstufen lediglich triviale Rückschlüsse erlauben, bieten höhere Verarbeitungsebenen die Möglichkeit einer umfassenden Betrachtung. Eine Modellierung der in Abb. 28 dargestellten Rückschlusshierarchie erlaubt hierbei das *Joint Directors of Laboratories* (JDL) Datenfusionsprozess-Modell (Abb. 27), unter Verwendung verschiedener Stufen der Datenprozessierung<sup>50</sup>.

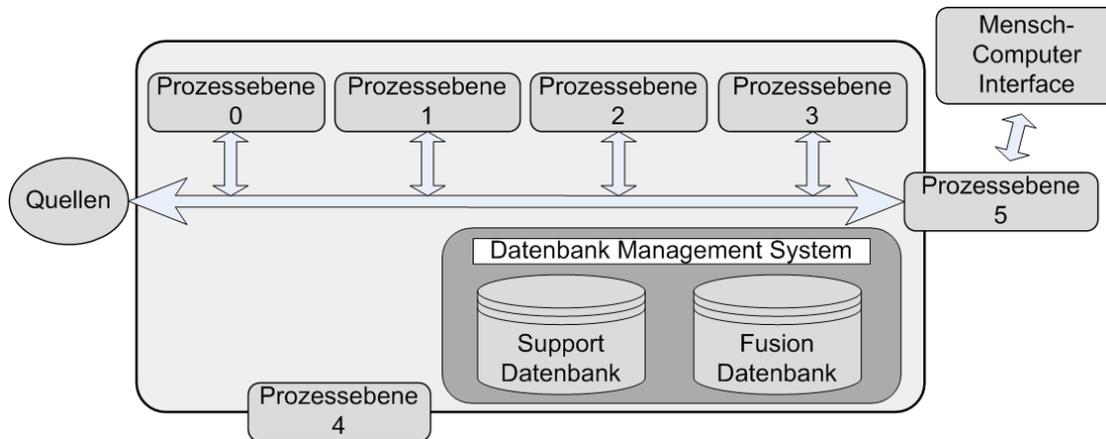


Abbildung 27: Das Joint Directors of Laboratories (JDL) Datenfusionsprozess-Modell.

Eine detaillierte Erklärung des kompletten Modells findet sich in [87]. Zusätzliche Informationen bezüglich der Verwendung unterschiedlicher Prozessstufen sind zu finden in [94]-[96].

#### PROZESSEBENE 0:

**Präprozessierung der Datenquellen.** Im Allgemeinen bedürfen die Informationen, welche direkt von den Datenquellen akquiriert werden, einer Vorbehandlung, bevor sie einer anschließenden Fusion mit Daten von anderen Quellen zugeführt werden. Beispiele der Vorbehandlung sind u. a. Bildbearbeitungsalgorithmen zur Korrektur von Linsenverzerrungen, Erzeugung von Farbinformationen aus Mustern, z. B. der Bayer-Matrix, Angleichung der Daten bzgl. Raum und Zeit, Konversion der Einheiten oder plattformspezifische Korrekturen.

#### PROZESSEBENE 1:

**Objektverfeinerung.** Basierend auf den Ergebnissen der vorigen Prozessebene dient diese der Kombination von Sensordaten, um eine möglichst exakte und verlässliche Schätzung bzgl. der gesuchten Eigenschaften, wie Bewegung, Position sowie Attribute wie Größe oder Klasse des Zielobjekts, zu erhalten.

#### PROZESSEBENE 2:

**Situationsverfeinerung.** Prozessebene 2 verwendet die Ergebnisse vorausgegangener Ebenen, um eine Beschreibung der momentanen Beziehungen zwischen den detektierten Zielobjekten sowie deren Umgebung zu entwickeln. Während Ebene 1 hauptsächlich den Fokus auf die Existenz der zu detektierenden Objekte legt, liegt der Schwerpunkt dieser Ebene auf der Interpretation des Vorhandenseins eines oder mehrerer Zielobjekte. Beispielhafte Einsatzgebiete sind z. B. kontextbezogene Interpretation (Vorhandensein mehrerer Zielobjekte in einem bestimmten Gebiet), Schlussfolgerung unter Benutzung mehrerer Perspektiven (Auflösung von Mehrdeutigkeiten), Detektion spezifischer Aktivitäten (auffälliges Verhalten) bzw. Ereignisse.

<sup>50</sup>Die Beschreibung der Prozessstufen erfolgt in Anlehnung an [87].

#### PROZESSEBENE 3:

**Beurteilung der Auswirkung.** Prozessebene 3 benutzt die von den vorigen Ebenen bereitgestellten Informationen, um eine Projektion der momentanen Situation in die Zukunft zu erzeugen und um somit mögliche Auswirkungen oder Bedrohungen vorherzusehen. Paradebeispiel sind hierfür militärische Anwendungen, in denen aufgrund einer gegebenen Situation versucht wird, eine Einschätzung der Lage zu generieren oder zukünftige potenzielle Bedrohungen vorherzusagen, etwa durch die Detektion etwaiger feindlicher Objekte in einem bestimmten Zielgebiet.

#### PROZESSEBENE 4:

**Prozessverfeinerung und Ressourcenmanagement.** Prozessebene 4 kann als eine Validierungsstufe des gesamten Fusionsprozesses angesehen werden. Angedeutet ist dieses als ein Block in Abb. 26 außerhalb der Prozessierungskette (Ebene 0-3). Ziel ist es, durch Überwachung des gesamten Fusionsprozesses etwaige Schwachstellen aufzudecken und entsprechende Maßnahmen einzuleiten. Als Beispiel kann eine Optimierung eines Algorithmus zur Erhöhung der Echtzeit-Performanz herangezogen werden. Diese Ebene ist im Allgemeinen ein optionaler Schritt, da es nicht in jedem Fusionsprozess möglich sein wird, eine entsprechende Optimierung im Nachhinein zu implementieren.

#### PROZESSEBENE 5:

**Kognitive Verfeinerung.** Wie auch die vorige Prozessebene nimmt diese ebenfalls eine Sonderstellung ein. Motiviert wird die Einführung dieser Ebene mit der häufig auftretenden Notwendigkeit, eine menschliche Entscheidung in den kompletten Fusionsprozess einfließen lassen zu müssen, respektive zu wollen. Im Rahmen dieser Prozessierungsstufe werden die Daten in einer sinnvollen Art und Weise dem Benutzer zugänglich gemacht, z. B. durch eine entsprechende Visualisierung. Zusätzlich können Methoden zur Verfügung gestellt werden, die eine direkte Interaktion des Benutzers mit dem System ermöglichen (Human-Computer-Interfaces). Denkbar ist, dass Systeme sowohl auf die physische Verfassung als auch auf den mentalen Zustand eines Benutzers reagieren [94].

Anzumerken ist, dass die im Rahmen dieser Dissertation durchgeführten Bearbeitungsschritte zu den Prozessstufen 0-2 des JDL-Datenfusionsmodells gezählt werden können. Die Verarbeitungsschritte welche den höheren Prozessebenen entsprechen, werden in dieser Arbeit nicht behandelt.

#### 3.2.4 PROZESSABLAUF

Begonnen wird die für einen Zeitpunkt  $k$  durchzuführende Prozessierung der Informationen durch eine zeitsynchrone Datenakquisition. Jeder MSK sammelt nach Erhalt einer vom FSK gesendeten Triggernachricht eigenständig die entsprechenden Sensorinformationen. Anschließend dazu findet auf jedem MSK eine lokale Präprozessierung der Sensordaten statt (Abschnitt 4.1), deren Notwendigkeit in Abschnitt 2.2.3 diskutiert worden ist. Gemäß des JDL-Modells sind diese beiden initialen Schritte der Prozessebene 0 zuzuordnen (Abschnitt 3.2.3).

Der nächste Verarbeitungsschritt betrifft die Lokalisation, Verfolgung und Klassifikation der detektierten Objekte. Gleich der vorigen Stufe finden diese Vorgänge ebenfalls lokal und parallel auf jedem MSK statt. Die auf Basis des lokalen Koordinatensystems berechneten Eigenschaften der lokalisierten Objekte wie Position, Geschwindigkeit sowie deren Eigenart (Person / keine Person) werden in einem Statusvektor zusammengefasst und anschließend an den FSK gesendet. Die dadurch erreichte Reduktion der pro Zeitpunkt an den FSK zu übertragenden Daten ermöglicht eine gleichzeitige Operation einer Vielzahl von MSK, ohne dass eine Beschränkung durch die zugrunde liegende TCP/IP-basierte Kommunikationsinfrastruktur befürchtet werden muss. Bedingt durch die Parallelität des Ansatzes wird durch eine Hinzunahme zusätzlicher MSK der zur Bearbeitung notwendige Zeitraum  $t_k^{Proc}$  idealerweise nur unwesentlich verlängert, da lediglich der

### 3 Agentenparadigma-Systemtopologie

Fusionsprozess eine Erhöhung der abzuarbeitenden Datenmenge erfährt (Abschnitt 5.5). Im Kontext des JDL-Modells entsprechen diese Verarbeitungsschritte einer Prozessierung auf der ersten Ebene.

Nach Erhalt aller MSK-Zustandsvektoren und anschließender Datenangleichung (Abschnitt 3.2.2), werden die Informationen durch den FSK zu einem globalen Zustandsvektor fusioniert, welcher die momentane Umgebung aus Sicht des kompletten Systems wiedergibt. Dieses impliziert, dass der Fusionsprozess erst nach Erhalt aller lokalen Zustandsvektoren beginnt. Da die notwendige Bearbeitungszeit  $t_k^{Proc}$  im Wesentlichen von der lokalen Berechnungsdauer auf den MSK bestimmt wird, ist die Verwendung einer homogenen Rechnerarchitektur ratsam. Nach erfolgter Erzeugung der globalen Beschreibung schließt sich die Verfolgung von Objekten auf Grundlage des globalen Zustandsvektors an, welche ebenfalls auf dem FSK durchgeführt wird. Es ist die Information bezüglich der globalen Objekte, welche das Endprodukt der Prozessierung eines Zeitpunktes  $k$  darstellt und welche zwecks einer sich anschließenden Verarbeitung vom FSK zur Verfügung gestellt wird. Gemäß der Definition des JDL-Modells sind diese Operationsschritte der Prozessebene 2 zuzuordnen.

Der komplette Informationsfluss des Systems auf Basis der in Abschnitt 3.1.7 und 3.1.7.2 definierten MSK und FSK und auf Grundlage des JDL-Prozessmodells ergibt sich aus Abb. 28.

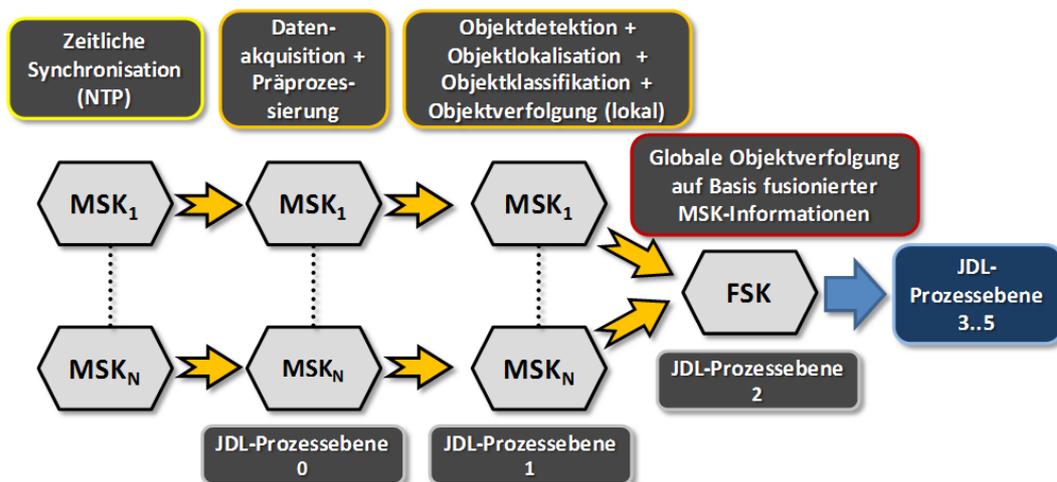


Abbildung 28: Informationsfluss des Gesamtsystems.

Bei Betrachtung des Informationsflusses (Abb. 28) fällt auf, dass eine gewisse konzeptionelle Ähnlichkeit zu Blackboard-Modellen besteht [98], welche bereits seit einiger Zeit zur Lösung verteilter Probleme herangezogen werden, auf die aber im Weiteren nicht näher eingegangen wird. Für einen Vergleich von Blackboard- mit Multiagentensystemen sei diesbezüglich auf [137] verwiesen.

### 3.3 ZUSAMMENFASSUNG

Das auf dem Agentenparadigma beruhende System bietet aufgrund seiner Konzeption die Möglichkeit einer einfachen Skalierung. Bezogen auf Überwachungssysteme impliziert Skalierbarkeit sowohl eine Erweiterbarkeit des überwachten Gebietes als auch eine damit verbundene Steigerung des Informationsgehaltes. Der Vorteil, sich nicht von vornherein auf eine gewisse Konfiguration bzw. Topologie festlegen zu müssen, ist systeminhärent.

Ein zusätzlicher Vorteil ergibt sich durch die modulare Konzeption des Agentensystems, welches eine funktionale Erweiterung in fast beliebiger Form zulässt. Die Hinzunahme zusätzlicher Informationsquellen (z. B. Sensoren) in Form von Plug-Ins sowie die Verlagerung der Algorithmik

### 3 Agentenparadigma-Systemtopologie

in ein extern zur Laufzeit zuladbares und einfach austauschbares Modul potenzieren die Anwendungsmöglichkeiten des gewählten Ansatzes.

Aufgrund der gewählten Kommunikationsstruktur kann der FSK seine globalen Informationen in einfacher Weise an eine eventuell vorhandene Kontrollapplikation versenden, welche die erfassten Daten z. B. visuell aufbereiten und präsentieren kann. Denkbar ist weiterhin ein dem Prozess der Personen- bzw. Objektverfolgung auf Fremdsystemen nachgeschalteter Bearbeitungsschritt aus den höheren Prozessierungsebenen des JDL-Modells (Abschnitt 3.2.3), z. B. Verhaltensanalyse oder Situationsabschätzung.

#### ANWENDUNG DES AGENTENPARADIGMAS IM PROJEKT POMSE

Aufgrund der modularen Konzeption des implementierten, agentenbasierten Grundsystems ist es nicht auf den Einsatz von spezieller Sensorik, z. B. der MultiCam, beschränkt. Im Rahmen des POMSe-Projektes (Detektion und Lokalisation von Personen und Objekten unter Verwendung mobiler Sensorik) [217] wurde es mehrfach, unter Verwendung eines verschiedenen Satzes von Sensoren und unter Verwendung eines von Abb. 28 leicht abweichenden Flusses, erfolgreich eingesetzt.

Die grundsätzliche Zielsetzung von POMSe war die Untersuchung der Möglichkeit des Einsatzes eines mobilen Netzwerks von Multisensorknoten zur Lokalisation und Detektion sowohl von Personen als auch Objekten. Der leitende Gedanke dabei war, mithilfe eines mobilen Sensornetzwerks ein von der Umgebung abhängiges und zeitlich beschränktes Sperrgebiet zu definieren und das Sensornetzwerk zur Detektion und Klassifikation der in das Zielgebiet eindringenden Objekte einzusetzen. Eine zu beachtende Einschränkung war dabei das Verbot der Verwendung von Sensorsystemen, z. B. reguläre Kameras, welche die potenzielle Möglichkeit böten, einzelne Personen zu identifizieren.

Die erste Projektphase von POMSe (AP1) diente der Evaluation existierender Sensorverfahren im Hinblick auf ihre Eignung zur Personendetektion unter Berücksichtigung der genannten Einschränkung. Unter Beachtung verschiedener menschlicher Signaturen (Abb. 1) wurde zunächst eine Auswahl infrage kommender Sensorverfahren zusammengestellt. Nach der Bildung der infrage kommenden Sensorik war es Aufgabe jedes einzelnen teilnehmenden technischen Partners, sich auf Grundlage der vorhandenen Expertise auf eine Kombination von maximal zwei Sensorsystemen zu fokussieren. Die Wahl der Sensorkombinationen der jeweiligen technischen Partner (TP) enthält Tab. 1.

Tab. 1: Sensorauswahl der technischen Partner (Projekt POMSe).

Technischer Partner	Verwendete Sensorik
Zentrum für Sensorsysteme (ZESS) Universität Siegen	<ul style="list-style-type: none"><li>• Passive Infrarotsensorsysteme (PIR)<sup>51</sup><ul style="list-style-type: none"><li>◦ Kommerzielle Systeme (PIR1, PIR2)</li><li>◦ Eigenentwicklungen</li></ul></li><li>• Thermalkameras (TK) (Auflösung 16x16 Pixel)</li></ul>
Institut für elektronische Messtechnik (EMT) Universität Ilmenau	<ul style="list-style-type: none"><li>• Radar (Doppler)</li><li>• Radar (UWB)<sup>52</sup></li></ul>
Institut für Sensorik und Messsysteme (SMS) Universität der Bundeswehr	<ul style="list-style-type: none"><li>• Radar (FMCW)<sup>53</sup></li><li>• Laser-Scanner (LIDAR)</li></ul>
Sinus GmbH	<ul style="list-style-type: none"><li>• Trittschallsensoren</li></ul>

<sup>51</sup>PIR-Sensor: *Passive Infra-Red*-Sensor.

<sup>52</sup>UWB-Radar: *Ultra Wideband*-Radar.

<sup>53</sup>FMCW-Radar: *Frequency Modulated Continuous Wave*-Radar.

### 3 Agentenparadigma-Systemtopologie

Auf Basis der ausgewählten Sensorik sollte im Rahmen der ersten Projektphase jeder Partner zusätzlich ein eigenes mobiles Sensornetzwerk errichten und seine Eignung im Hinblick auf eine Personendetektion untersuchen.

Das im Rahmen von P0mSe entwickelte Demonstrationsnetzwerk<sup>54</sup> wurde auf Grundlage des in dieser Arbeit vorgestellten Agentensystems implementiert. Jeder der drei vorhandenen Knoten bestand in seiner Grundform aus einem kompakten Rechner und verwaltete eine variierende Kombination aus kommerziellen und selbst entwickelten PIR-Systemen sowie einer Thermalkamera mit geringer Auflösung. Die für die Operation des Netzwerks implementierte Algorithmik auf den Sensorknoten übernahm die Akquisition, Vorprozessierung und Klassifikation. Die Ergebnisse der Prozessierung wurden an einen im Netzwerk vorhandenen Fusionsknoten (FSK) gesendet<sup>55</sup>.

Der erste Einsatz des agentenbasierten Demonstrationsnetzwerks fand anlässlich der initialen, mehrtägigen Messkampagne in Siegen statt. Zielsetzung dieser Kampagne war die Verifikation der Netzwerkfunktionalität sowie der Akquisition von Testdaten zwecks Entwicklung von Algorithmen zur Objektdetektion und Objektklassifikation auf Grundlage der verwendeten Sensorik (Tab. 1). Die betrachteten Testszenarien sind in den Abb. 29 und 30 dargestellt.

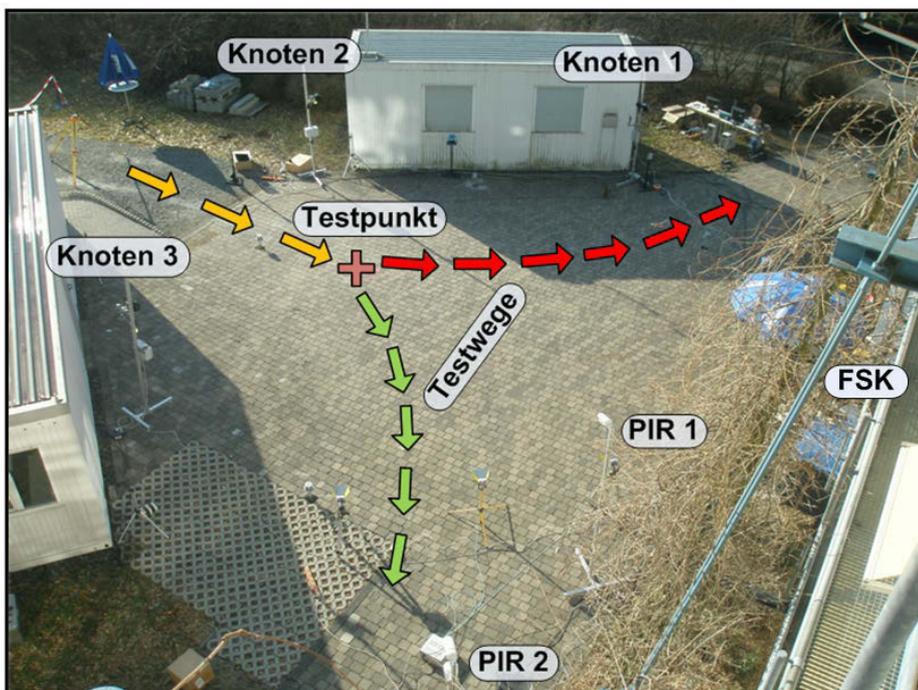


Abbildung 29: Einsatz agentenbasierter Multisensorknoten I.

Erste Anordnung der Multisensorknoten des ZESS-Sensornetzes anlässlich der ersten Messkampagne im Rahmen des P0mSe-Projektes (API). Siegen, Januar 2009. Die Bewegung der Personen bzw. Objekte erfolgte entlang der zwei eingezeichneten Testwege.

In jedem dieser Szenarien bewegten sich die Testobjekte in definierten Versuchsabläufen auf verschiedene Weise (gehend, kriechend, etc.) entlang der eingezeichneten Testwege. Um Vergleichswerte zu besitzen, wurden zusätzlich zu den Personen diverse Objekte (Puppe, Koffer, etc.) entlang der Testwege bewegt. Zusätzlich dazu stoppte jedes Objekt für eine kurze Zeit an einem definierten Testpunkt. Analog zur Vorgehensweise in dieser Arbeit wurden die

<sup>54</sup>Jeder teilnehmende technische Partner war selbst für die Entwicklung seines Sensornetzwerks verantwortlich. Die in diesem Kapitel getroffenen Aussagen bzgl. der Konfiguration sowie des Prozessablaufs betreffen somit nur das eigene, agentenbasierte Sensornetz.

<sup>55</sup>Die Kommunikation der am Netz beteiligten Knoten erfolgte auf Grundlage von TCP/IP.

### 3 Agentenparadigma-Systemtopologie

Akquisitionssignale durch den Fusionsknoten in regelmäßigen Intervallen an die teilnehmenden Sensorenknoten übermittelt.

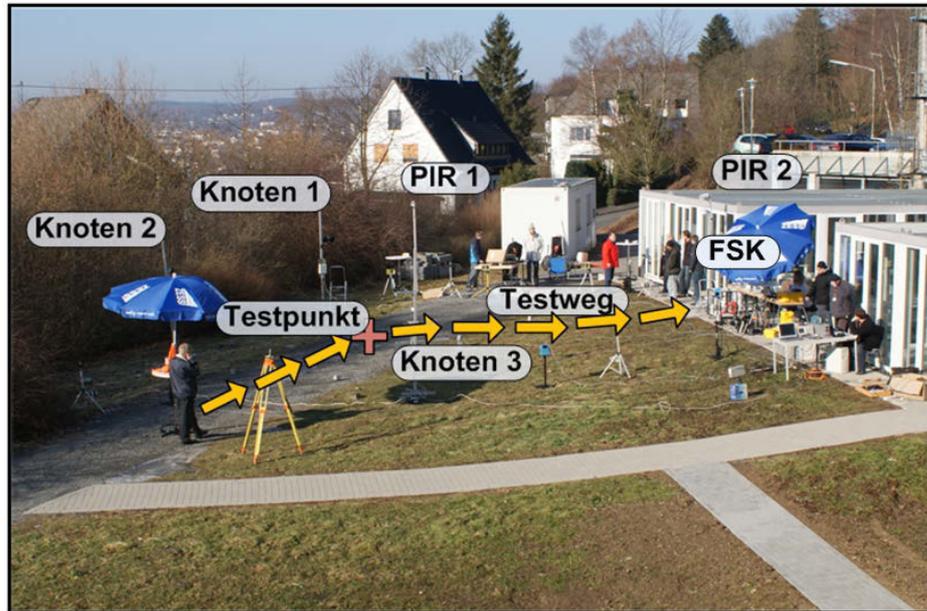


Abbildung 30: Einsatz agentenbasierter Multisensorknoten II.

Zweite Anordnung der Multisensorknoten (ZESS-Sensornetz) anlässlich der ersten Messkampagne im Rahmen des P0mSe-Projektes (AP1). Siegen, Januar 2009.

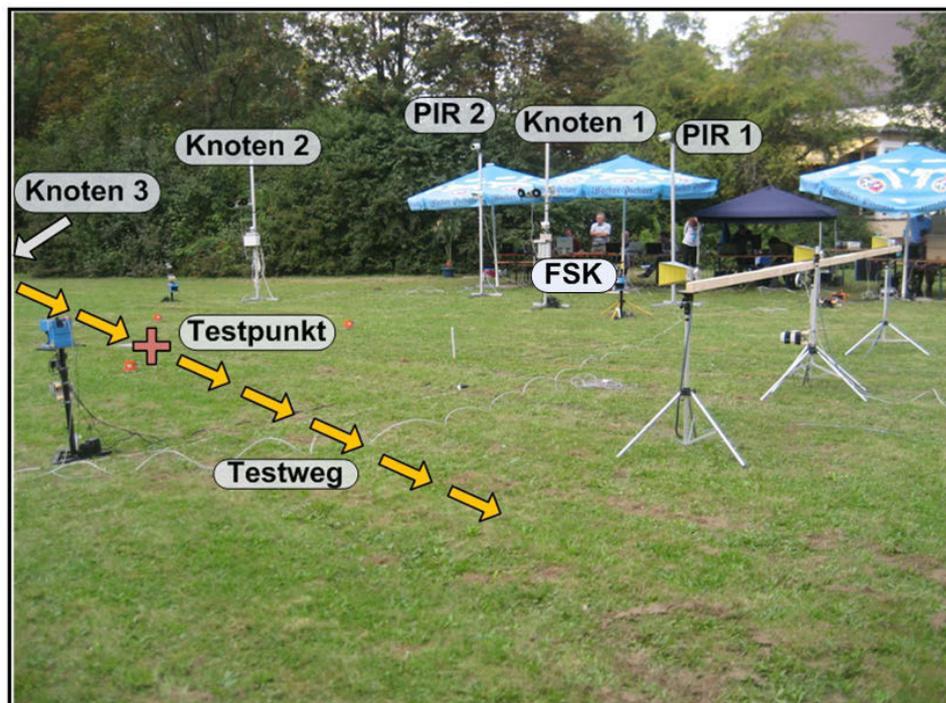


Abbildung 31: Einsatz agentenbasierter Multisensorknoten III.

Anordnung der Multisensorknoten des ZESS-Sensornetzes während der abschließenden Messkampagne im Rahmen des P0mSe-Projektes (AP2). München, September 2009.

### 3 Agentenparadigma-Systemtopologie

---

Die Sensorknoten wiederum hatten die Aufgabe, die akquirierten Sensordaten in einem definierten Format lokal abzuspeichern, was letztendlich einer Vorprozessierung der Daten entsprach. Bezogen auf das JDL-Prozessmodell (Abb. 27) operierten die Netze aller technischen Partner während dieser Messkampagne im Bereich der Prozessebene 0. Detaillierte Angaben zur ersten Projektphase finden sich in [218].

Auf Grundlage der gesammelten Informationen wurden im Verlauf der zweiten Projektphase (AP2) Algorithmen zur Detektion von Objekten und zur Klassifikation zwischen Person und Objekt entwickelt. Anlässlich einer finalen, mehrtägigen Messkampagne in München (Abb. 31), unter Teilnahme des BSI, wurden die erreichten Fähigkeiten der durch die technischen Partner entwickelten Sensornetzwerke im Hinblick auf Lokalisation und Klassifikation von Objekten demonstriert.

Analog zur ersten Messkampagne in Siegen wurde diesbezüglich erneut ein Testweg definiert, auf dem sich Personen und Objekte durch das zu überwachende Gebiet bewegen mussten. Im Gegensatz zur ersten Kampagne war die Aufgabe der Netzwerke diesmal jedoch die Detektion und Klassifikation der Objekte auf Grundlage der durch das Objekt pro Messung erzeugten Informationen in Echtzeit, d. h. direkt nach Beendigung der Messung. Jede Messung beinhaltete diesbezüglich einen zeitlich definierten Start und Stoppunkt, welcher in Form eines elektronischen Signals an die Netzwerke aller technischen Partner simultan gesendet wurde.

Die während der Messung anfallenden Sensordaten wurden vom jeweiligen Sensorknoten lokal gesammelt. Nach Beendigung der Messung wurden die akquirierten Daten zunächst auf jedem Knoten lokal vorbehandelt und dem Algorithmus zur Klassifikation zugeführt. Anschließend wurden die lokal erzeugten Ergebnisse an einen Fusionsknoten gesendet, welcher diese zu einem Endergebnis kombinierte und visualisierte. Zusätzlich dazu wurden alle Ergebnisse der Detektion und Klassifikation pro Sensorknoten gespeichert.

Die Ergebnisse der Detektion und Klassifikation von Objekten unter Verwendung des diesbezüglich entwickelten agentenbasierten Multisensornetzwerks auf Grundlage passiver Sensorik wurde im Rahmen einer Abschlusspräsentation dem BSI vorgestellt. Details zur Implementation der Algorithmik zur Detektion und Klassifikation sowie der verwendeten Konfigurationen der Sensorknoten finden sich [219].

Basierend auf den Ergebnissen ergab sich die Möglichkeit einer zusätzlichen Kooperation zwischen BSI und ZESS im Rahmen des Folgeprojekts MOVEDETECT [229], auf das hier aber nicht weiter eingegangen wird.

## 4 AGENTENBASIERTE DETEKTION, LOKALISATION UND KLASSEIFIKATION VON OBJEKTEN

Gemäß der verfolgten Methodik zur Detektion, Lokalisation und Klassifikation von Objekten bedarf es aufeinander aufbauender Prozessstufen, wobei der zugrunde liegende Informationsfluss des Gesamtsystems in Abschnitt 3.2.4 vorgestellt wurde. Ziel dieses Kapitels ist die Diskussion der implementierten Schritte der JDL-Prozessebenen 0 und 1, deren Zusammenhang in Abb. 32 wiedergegeben ist.

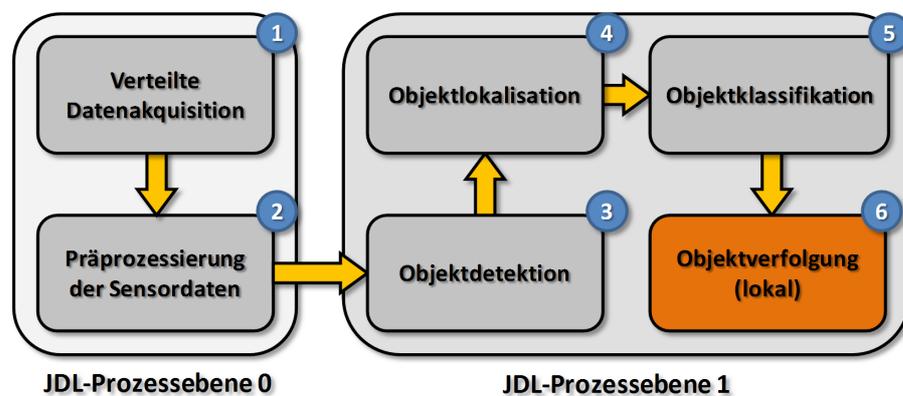


Abbildung 32: Schritte der JDL-Prozessebenen 0 und 1.

Zunächst behandelt Abschnitt 4.1 die Bearbeitungsschritte der JDL-Prozessebene 0, bestehend aus der verteilten Datenakquisition sowie der Präprozessierung akquirierter Sensordaten.

Abschnitt 4.2 befasst sich mit den zur JDL-Prozessebene 1 zugehörigen Vorgängen und bildet den Hauptteil dieses Kapitels. Diskutiert werden die implementierten Verfahren zur Detektion und Lokalisation von Objekten auf Grundlage der Informationen des PMD-Chips. Die in Abb. 32 ebenfalls angegebene lokale Objektverfolgung wird in Kapitel 5 Gegenstand der Diskussion sein.

Abschnitt 4.3 diskutiert den im Rahmen dieser Arbeit verfolgten Ansatz zur Klassifikation von Objekten, wobei im Anschluss dazu die Vorteile der kooperativen Fusion der unterschiedlichen Sensordaten beleuchtet werden.

### 4.1 OPERATIONEN AUF JDL-PROZESSEBENE 0

#### 4.1.1 VERTEILTE DATENAKQUISITION

Zu Beginn einer Operationssequenz übernimmt der FSK die Aufgabe der Benachrichtigung der kooperierenden Multisensorknoten (Abb. 29), welche diese zur Bildakquisition veranlasst. Eine durchgeführte Untersuchung in Bezug auf die zu erwartenden Latenzen ergab, dass die notwendige Zeitspanne zur Benachrichtigung eines Agenten bzw. Multisensorknotens ca. 0.3 ms beträgt, was gewährleistet, dass die getriggerten Szenenaufnahmen eine hinreichende zeitliche Synchronizität besitzen, um sich bewegende Objekte bzw. Personen eindeutig zu lokalisieren. Bei einer angenommenen Geschwindigkeit von 6 km/h, was schnellem Gehen entspricht, liegt die während 1 ms zurückgelegte Weglänge bei etwa 1.6 mm. In Bezug auf eine Lokalisierung durch mehrere,

## 4 Agentenbasierte Detektion, Lokalisation und Klassifikation von Objekten

sequenziell getriggerte Agenten<sup>56</sup>, ist der so entstehende räumliche Versatz der Objekte zu vernachlässigen. Zwecks eindeutiger Identifizierung der Aufnahmen besitzt jedes Bild einen 32 Byte großen Header mit eingebettetem Zeitstempel (Tab. 6, Anhang F).

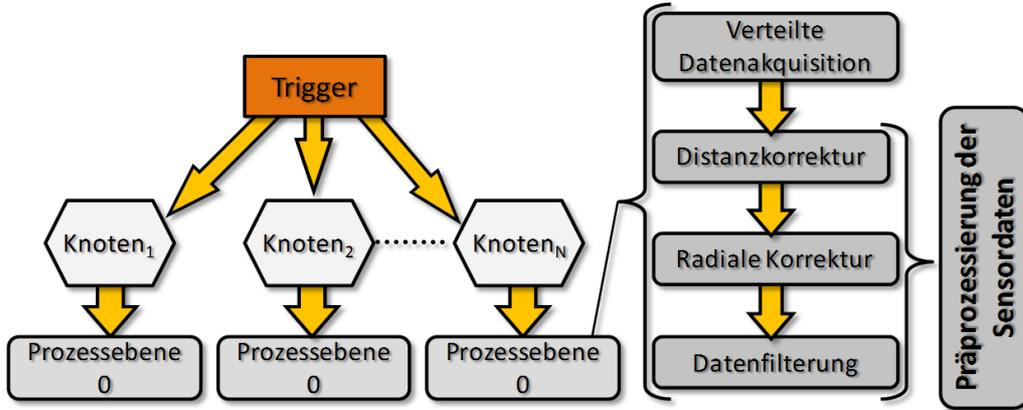


Abbildung 33: Operationen der JDL-Prozessebene 0.

### 4.1.2 PRÄPROZESSIERUNG DER SENSORDATEN

#### 4.1.2.1 DISTANZKORREKTUR

Der erste Schritt zur Korrektur der akquirierten Informationen ist eine Berichtigung der Entfernungsdaten, welche aufgrund inhomogener Modulationssignale eine charakteristische Abweichung von der idealen Linie besitzen (Abschnitt 2.2.3). Für jeden gemessenen PMD-Pixelwert  $z_{u,v}^{Mess}$  wird ein korrigierter Wert  $z_{u,v}^{Korr}$  mittels

$$\tilde{z}_{u,v} = z_{u,v}^{Mess} - \sum_{i=0}^{10} p_i (z_{u,v}^{Mess})^i \quad (17)$$

berechnet. Die Indizes  $u, v$  repräsentieren dabei die Pixelpositionen auf dem PMD-Chip. Die Koeffizienten  $p_i$  sind durch die charakteristische Abweichfunktion (14) der MultiCam bestimmt.

#### 4.1.2.2 RADIALE KORREKTUR

Die Berechnung der radialen Korrektur ist lediglich von der Brennweite der verwendeten Optik  $f$  sowie den Ausmessungen der PMD-Chip-Pixel  $s_u, s_v$  abhängig, sodass in der Regel eine Berechnung der Korrekturwerte lediglich ein Mal durchgeführt werden muss. Die radiale Korrektur wird auf den bereits distanzkorrigierten PMD-Pixeln  $\tilde{z}_{u,v}$  (17) mittels

$$z_{u,v}^{Korr} = \tilde{z}_{u,v} \cos \left( \arctan \left( \frac{\sqrt{(s_u(|u|-0.5))^2 + (s_v(|v|-0.5))^2}}{f} \right) \right) \quad (18)$$

<sup>56</sup>Es sei an dieser Stelle darauf hingewiesen, dass im Falle der Verwendung des TCP/IP-Protokolls keine Broadcast-Nachricht zur Verfügung steht. Eine sequenzielle Benachrichtigung aller MSK ist durch den FSK durchzuführen. Eine Möglichkeit der Erweiterung wäre die zusätzliche Verwendung von UDP als Signaltransportschicht, welche aber im Rahmen dieser Arbeit nicht implementiert wurde.

## 4 Agentenbasierte Detektion, Lokalisation und Klassifikation von Objekten

---

durchgeführt. Für die Korrekturberechnung wird jeweils der Mittelpunkt eines PMD-Pixels als Referenz verwendet. Die in  $Z^{Korr} = \{z_{u,v}^{Korr}\}$  enthaltenen, zweifach korrigierten Pixelwerte bilden die Basis für eine weitergehende Verarbeitung der Distanzinformationen.

### 4.1.2.3 ENTFERNUNGSDATENFILTERUNG, PLAUSIBILITÄTSTEST

Die Notwendigkeit einer Filterung der Entfernungsdaten ergibt sich aufgrund des definierten Überwachungsvolumens  $z_{min} \leq z \leq z_{max}$  (Abschnitt 1.2) und ist durch die von der MultiCam gelieferten Tiefeninformationen in einer einfachen Weise möglich.

Sei  $D$  eine Matrix mit in Vergleich zu  $Z^{Korr}$  gleicher Dimensionalität. Für jedes Element  $z_{u,v}^{Korr}$  aus  $Z^{Korr}$  existiert ein korrespondierendes Element  $d_{u,v}$  aus  $D$ , mit

$$d_{u,v} = \begin{cases} 1 & z_{min} \leq z_{u,v}^{Korr} \leq z_{max}, \\ 0 & \text{sonst,} \end{cases} \quad (19)$$

welches eine Aussage bzgl. der Validität des entsprechenden Pixels im Sinne der Distanzfilterung trifft.

### 4.1.2.4 MODULATIONSDATENFILTERUNG

Der hauptsächliche Grund für eine Filterung der Modulationsinformation liegt in der Entfernung von Distanzinformationen mit geringer Zuverlässigkeit. Gemäß (13) ergibt sich, dass die Präzision der Entfernungsmessung sowohl durch Minimierung der Hintergrundintensität als auch durch Maximierung der Modulationsintensität zu steigern ist. Dieses impliziert, dass Entfernungsmessungen mit geringer Modulationsamplitude bei angenommen gleicher Hintergrundintensität eine hohe Varianz aufweisen und dementsprechend als unzuverlässig einzustufen sind. Die Wahl einer unteren Schwelle  $\mu_{min}$  ist in der Regel abhängig vom gewählten Einsatz.

Basierend auf den Erfahrungen, welche durch die Verwendung der MultiCam in diversen Anwendungsbereichen [62], [74], [82], [83], [188] gewonnen wurden, ist ein Wert von 20 ADC-Counts als untere Grenze im Rahmen dieser Arbeit als sinnvoll erachtet worden. Eine obere Filterschwelle für Modulationsamplituden ist im Falle der Benutzung einer minimalen Distanz als Filterkriterium für Distanzdaten nicht notwendig, da dadurch automatisch eine entsprechende Begrenzung der Modulationsdaten erfolgt.

Wie auch im Falle der Distanzfilterung sei zunächst die Matrix  $M$  mit zu  $Z^{Korr}$  gleicher Dimensionalität eingeführt. Des Weiteren sei  $Z^{Mod}$  die Matrix der Modulationsdaten des Bildes. Durch

$$m_{u,v} = \begin{cases} 1 & z_{u,v}^{Mod} > \mu_{min}, \\ 0 & \text{sonst,} \end{cases} \quad (20)$$

beinhalten die Elemente  $m_{u,v}$  von  $M$  die Information, ob die durchgeführte Messung des PMD-Pixels an Position  $u, v$  eine hinreichende Verlässlichkeit besitzt (Abb. 34).

#### 4 Agentenbasierte Detektion, Lokalisation und Klassifikation von Objekten

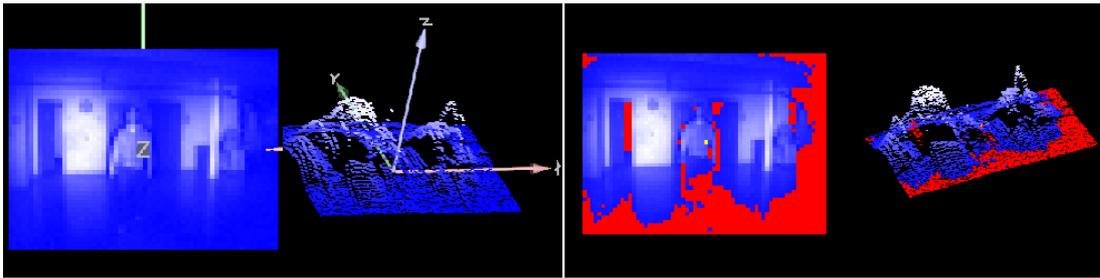


Abbildung 34: Filterung der Modulationsamplitude.

Links: Modulationsdaten vor Filterung. Rechts: Nach Filterung. Rot gefärbte Pixel entsprechen den verworfenen Distanzwerten mit einer Modulationsamplitude kleiner als  $\mu_{min} = 20$ .

Der letzte Schritt der Präprozessierung besteht in der Erzeugung einer Matrix  $Z^{Valid}$ , welche die gültigen Distanzdaten des Bildes eines Zeitpunktes enthält. Benutzt werden dazu die Informationen der Maskenmatrizen  $D$  und  $M$ , welche auf den Distanzdaten der Matrix  $Z^{Korr}$  angewendet werden:

$$z_{u,v}^{Valid} = \begin{cases} z_{u,v}^{Korr} & d_{u,v} > 0 \wedge m_{u,v} > 0, \\ 0 & \text{sonst.} \end{cases} \quad (21)$$

Abb. 35 zeigt die Ergebnisse der präprozessierten PMD-Distanzdaten. Zur besseren Erkennbarkeit wurden die gefilterten Distanzdaten  $Z^{Valid}$ , wo sinnvoll, orthografisch projiziert.

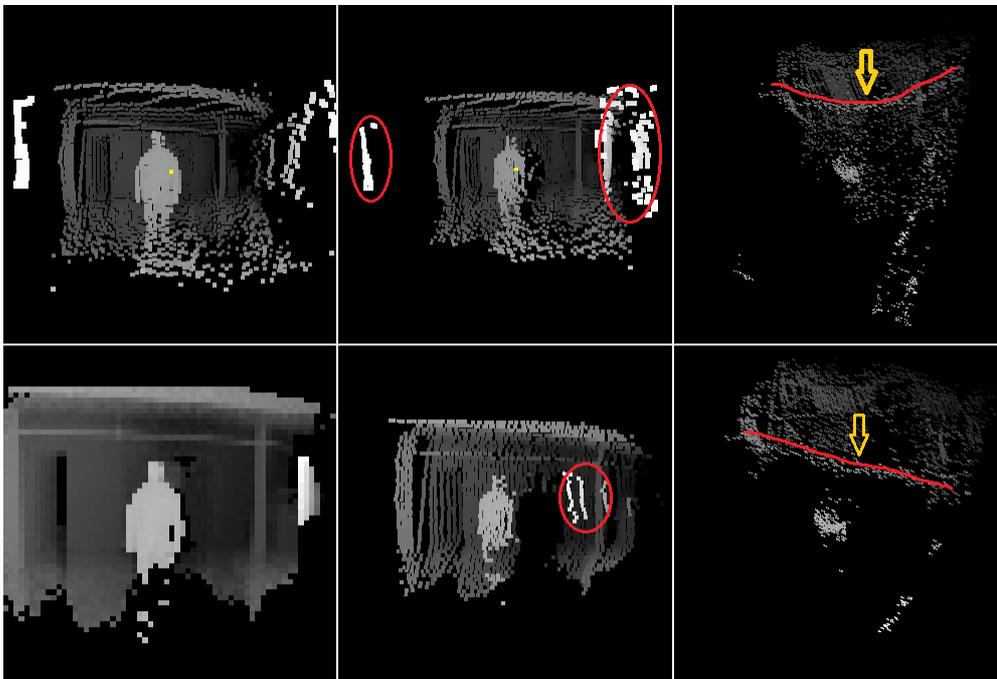


Abbildung 35: Effekte der Präprozessierung der PMD-Sensordaten.

Obere Zeile: Distanzwerte der Matrix  $Z^{Korr}$ . Untere Zeile: Distanzwerte der Matrix  $Z^{Valid}$ . Linke Spalte: Entfernungswerte vor (oben) und nach einer Präprozessierung der Sensordaten (unten, orthografische Projektion). Mittlere Spalte: Durch Präprozessierung reduzierte Menge der uneindeutigen Distanzdaten. Rechte Spalte: Effekte der radialen Korrektur.

## 4 Agentenbasierte Detektion, Lokalisation und Klassifikation von Objekten

Die linke Spalte von Abb. 35 enthält im oberen Bild die Distanzdaten von  $Z^{Valid}$ . Darunter sind die Informationen der Matrix  $Z^{Korr}$  in orthografischer Projektion abgebildet. Erkennbar ist die Reduktion der Informationen durch die Präprozessierung der Sensordaten. Die mittlere Spalte von Abb. 35 beleuchtet die Ergebnisse der Filterung der Distanzwerte mit  $Z_{min} > z_{u,v}^{Korr}$  sowie Tiefeninformationen mit zu geringen Modulationsamplituden. Ebenfalls gefiltert werden Distanzen, welche aufgrund des limitierten Eindeutigkeitsbereichs der MultiCam wieder in den definierten Überwachungsbereich der Kamera eingeblenndet wurden, da sie eine im Vergleich zu Distanzwerten aus diesem Bereich weitaus geringere Modulationsamplitude besitzen. In der rechten Spalte von Abb. 35 zeigen sich die Effekte der durchgeführten radialen Korrektur durch Anwendung von (18).

### 4.2 OPERATIONEN AUF JDL-PROZESSEBENE 1

Gemäß der Definition der JDL-Prozessebene 1 liegt die Hauptaufgabe dieser Stufe darin, basierend auf den Operationen der Prozessebene 0, eine möglichst verlässliche Schätzung der objektspezifischen Charakteristika der einzelnen Zielobjekte zu erzeugen (Abschnitt 3.2.3). Die Grundstruktur des verfolgten Ansatzes liefert Abb. 36. Die Prozessierungsstufen, welche für eine Detektion und Lokalisation von Objekten notwendig sind, basieren auf einem klassischen Bottom-Up-Modell, bestehend aus einer Hintergrundsubtraktion mit anschließender Merkmalsextraktion und werden in den folgenden Abschnitten behandelt werden. Die Diskussion des Prozesses der Klassifikation erfolgt anschließend.

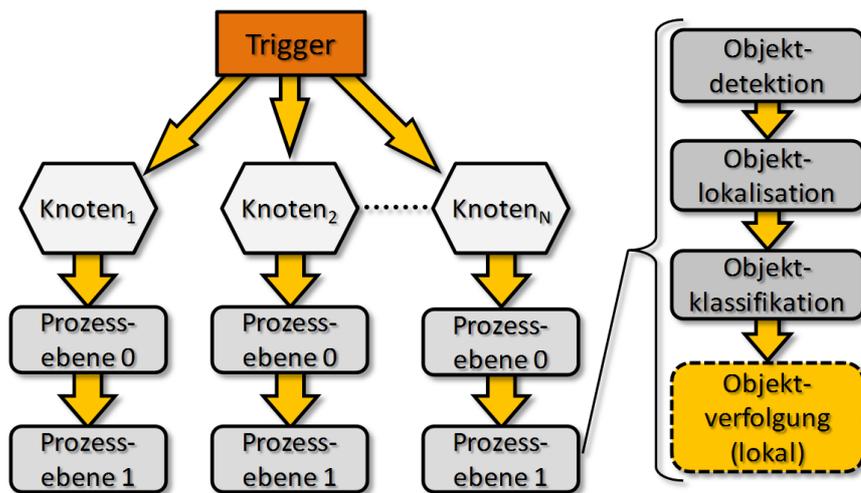


Abbildung 36: Operationen der JDL-Prozessebene 1.

#### 4.2.1 OBJEKTDETEKTION

Ein für stationäre Überwachungssysteme häufig verwendetes Verfahren zur Bestimmung von informationstragenden Daten bildet die Methode der Hintergrundsubtraktion [138], [139]. Das zugrunde liegende Prinzip ist, dass nach erfolgter Modellierung eines Hintergrundbilds dieses mit dem aktuellen Bild unter Verwendung einer Distanzoperation

$$\Delta(s, k) = \begin{cases} 1 & d(\underline{I}_{s,k}, \underline{H}_{s,k}) > \delta, \\ 0 & \text{sonst,} \end{cases} \quad (22)$$

## 4 Agentenbasierte Detektion, Lokalisation und Klassifikation von Objekten

---

verglichen wird, wobei die sich ergebenden Unterschiede bzw. Abweichungen in einer binären Vordergrundmaske abgelegt werden. In ihr sind die Informationen enthalten, welche Elemente des aktuellen Bildes durch (22) als Vordergrundinformation bestimmt worden sind und als Grundlage der Detektion von Objekten verwendet werden können.

Der Term  $\Delta(s, k)$  in (22) entspricht dem Wert des Maskenelementes mit Index  $s$  bezogen auf das Bild  $I$  des Zeitpunktes  $k$ , welcher durch den Distanzoperator  $d$  unter Verwendung eines Schwellenwertes  $\delta$  bestimmt worden ist. Ein von Null verschiedener Wert gibt an, dass das entsprechende Element als zum Vordergrund zugehörig betrachtet wird. Im Allgemeinen ist die Wahl des Operators  $d$  sowie des Schwellenwertes  $\delta$  von der Repräsentation der Bildinformationen abhängig. Beispiele verschiedener Distanzoperatoren finden sich u. a. in [138]-[143], [211].

Wie aus (22) ersichtlich, hängt die Erzeugung von Vordergrundinformationen wesentlich von der Wahl des verwendeten Hintergrundmodells ab. Während in einfachen Sonderfällen eine einmalige, statische Berechnung des Hintergrunds ausreichen mag, ist im Allgemeinen eine Anpassung an sich verändernde Umgebungen notwendig. Dieses impliziert, dass die Modellierung des Hintergrunds sich an die Veränderungen anpasst und diesen in das zu bildende Modell adaptiert. Diskutiert werden im Folgenden zwei im Rahmen der Implementierung eingesetzte Verfahren zur Hintergrundmodellierung, welche die Fähigkeit besitzen, Änderungen der Umgebung in die Modellbildung mit einzubeziehen.

### 4.2.1.1 DETERMINISTISCHE HINTERGRUNDMODELLIERUNG

Eine rechentechnisch effektive Methode, Änderung des Hintergrunds zu berücksichtigen, bietet das Verfahren des laufenden Mittelwerts (*Running Average*). Gegeben sei zunächst das Modell des Hintergrunds  $H_{k-1}^{RA}$  des vorherigen Zeitpunktes<sup>57</sup>  $k-1$  sowie das Bild des momentanen Zeitpunktes  $k$ . Unter Verwendung von

$$H_{s,k}^{RA} = (1-\alpha)H_{s,k-1}^{RA} + \alpha I_{s,k} \quad \text{mit } \alpha \in [0,1] \quad (23)$$

wird ein Update des Hintergrunds erreicht, wobei  $\alpha$  als Adaption- oder Lerngeschwindigkeit bezeichnet wird. Durch Variation von  $\alpha$  zur Laufzeit ist es möglich, den Einfluss neuer Informationen in die Modellbildung direkt zu verändern. Analog zu (22) entspricht der Index  $s$  in (23) dem zugehörigen Bildelement in  $I_k$  sowie  $H_k^{RA}$ . Als Kriterium, ob ein Bildelement teil des Hintergrunds ist, kann die Distanz<sup>58</sup> des betrachteten Bildelements  $I_{s,k}$  zu dem entsprechenden Element  $H_{s,k}^{RA}$  des Hintergrunds herangezogen werden. Werte in einer gewissen Nähe des Hintergrunds werden entsprechend klassifiziert.

### 4.2.1.2 PROBABILISTISCHE MODELLIERUNG DES HINTERGRUNDS

Der im Rahmen dieser Arbeit betrachtete Ansatz zur probabilistischen Modellierung des Hintergrunds basiert auf dem Verfahren vom Stauffer und Grimson [144], [145]. In ihrem Ansatz wird der Wert eines jeden Bildelementes durch eine gewichtete Summe aus  $n \in \mathbb{N}$  multivariaten Normalverteilungen (*Gaussian Mixture Model*, GMM) modelliert. Für ein Bildelement  $I_{s,k}$  ergibt sich damit, dass die Wahrscheinlichkeit seines Wertes unter Verwendungen seines ihn beschreibenden Modells durch

---

<sup>57</sup>Das Hintergrundmodell  $H_0^{RA}$  wird zu Beginn mit dem Bild  $L_0$  initiiert.

<sup>58</sup>Die zu wählende Distanz wird im Allgemeinen sowohl durch die Wahl des Operators aus (22) als auch von den herrschenden Umgebungsbedingungen sowie der Applikation bestimmt. Denkbar sind z. B. Distanzen auf Grundlage beliebiger l-Normen [138] oder statistischen Momenten wie Mittelwert und Varianz [141]-[143].

---

#### 4 Agentenbasierte Detektion, Lokalisation und Klassifikation von Objekten

---

$$\begin{aligned}
 f_{L_{s,k}}(\xi_{s,k}) &= \sum_{i=1}^n w_{i,s,k} f_{L_{s,k}|\underline{\mu}_{i,s,k}, \Sigma_{i,s,k}}(\xi_{s,k} | \tilde{\underline{\mu}}_{i,s,k}, \tilde{\Sigma}_{i,s,k}) \\
 &= \sum_{i=1}^n \frac{w_{i,s,k}}{(2\pi)^{h/2} |\tilde{\Sigma}_{i,s,k}|^{1/2}} \exp\left(-\frac{1}{2}(\xi_{s,k} - \tilde{\underline{\mu}}_{i,s,k})^T \tilde{\Sigma}_{i,s,k}^{-1} (\xi_{s,k} - \tilde{\underline{\mu}}_{i,s,k})\right)
 \end{aligned} \tag{24}$$

bestimmt wird. Vorausgesetzt wird dabei eine statistische Unabhängigkeit der Komponenten von  $L_{s,k}$ , was zur  $n$  diagonalen Kovarianzmatrizen  $\Sigma_{i,s,k} = \sigma_{i,s,k}^2 \mathbf{I}$  führt und ihre Invertierung rechentechnisch beschleunigt. Für die Bildelemente  $L_{s,k}$  in (24) wird angenommen, dass  $L_{s,k} \in \mathbb{R}^h$  und damit  $\underline{\mu}_{i,s,k} \in \mathbb{R}^h$ ,  $\Sigma_{i,s,k} \in \mathbb{R}^{h \times h}$  gilt. Weiterhin gilt für die Gewichtungsfaktoren bzw. Mischkoeffizienten [144], dass

$$\sum_{i=1}^n w_{i,s,k} = 1. \tag{25}$$

Streng genommen erfolgt in (24) eine theoretisch nicht exakte Gleichsetzung der unbedingten Wahrscheinlichkeitsdichte mit einer Summe aus gewichteten bedingten Dichten. Dieses bedarf einer Erklärung. Unter Verwendung der fundamentalen Summen- und Produktregel der Wahrscheinlichkeitstheorie [228] kann eine unbedingte Dichte geschrieben werden als

$$\begin{aligned}
 f_{L_{s,k}}(\xi_{s,k}) &= \sum_{i=1}^n f_{L_{s,k}|\underline{\mu}_{i,s,k}, \Sigma_{i,s,k}}(\xi_{s,k}, \tilde{\underline{\mu}}_{i,s,k}, \tilde{\Sigma}_{i,s,k}) \\
 &= \sum_{i=1}^n f_{\underline{\mu}_{i,s,k}, \Sigma_{i,s,k}}(\tilde{\underline{\mu}}_{i,s,k}, \tilde{\Sigma}_{i,s,k}) f_{L_{s,k}|\underline{\mu}_{i,s,k}, \Sigma_{i,s,k}}(\xi_{s,k} | \tilde{\underline{\mu}}_{i,s,k}, \tilde{\Sigma}_{i,s,k}).
 \end{aligned} \tag{26}$$

Die Form der unbedingten Dichte  $f_{L_{s,k}}(\xi_{s,k})$  aus (24) ergibt sich letztendlich durch Interpretation der  $n$  in (26) angegebenen multivariaten Wahrscheinlichkeiten  $f_{\underline{\mu}_{i,s,k}, \Sigma_{i,s,k}}(\tilde{\underline{\mu}}_{i,s,k}, \tilde{\Sigma}_{i,s,k})$  als den  $n$  Wichtungsfaktoren bzw. Mischkoeffizienten  $w_{i,s,k}$  von (24):

$$\begin{aligned}
 f_{L_{s,k}}(\xi_{s,k}) &= \sum_{i=1}^n f_{\underline{\mu}_{i,s,k}, \Sigma_{i,s,k}}(\tilde{\underline{\mu}}_{i,s,k}, \tilde{\Sigma}_{i,s,k}) f_{L_{s,k}|\underline{\mu}_{i,s,k}, \Sigma_{i,s,k}}(\xi_{s,k} | \tilde{\underline{\mu}}_{i,s,k}, \tilde{\Sigma}_{i,s,k}) \\
 &= \sum_{i=1}^n p\{\omega : \underline{\mu}_{i,s,k} = \tilde{\underline{\mu}}_{i,s,k}, \Sigma_{i,s,k} = \tilde{\Sigma}_{i,s,k}\} f_{L_{s,k}|\underline{\mu}_{i,s,k}, \Sigma_{i,s,k}}(\xi_{s,k} | \tilde{\underline{\mu}}_{i,s,k}, \tilde{\Sigma}_{i,s,k}) \\
 &= \sum_{i=1}^n w_{i,s,k} f_{L_{s,k}|\underline{\mu}_{i,s,k}, \Sigma_{i,s,k}}(\xi_{s,k} | \tilde{\underline{\mu}}_{i,s,k}, \tilde{\Sigma}_{i,s,k}).
 \end{aligned} \tag{27}$$

Die Mischkoeffizienten  $w_{i,s,k}$  entsprechen somit formal der multivariaten Wahrscheinlichkeit, dass die Parameter  $\underline{\mu}_{i,s,k} = \tilde{\underline{\mu}}_{i,s,k}$  und  $\Sigma_{i,s,k} = \tilde{\Sigma}_{i,s,k}$  sind.

Angenommen sei jetzt, dass zum Zeitpunkt  $k-1$  jedes Bildelement  $L_{s,k-1}$  durch (24) modelliert wurde. Im Falle des Vorhandenseins eines neuen Bildes  $L_k$  zum Zeitpunkt  $k$  dient der erste Schritt in diesem Verfahren der Auffrischung sämtlicher Modelle. Für jedes Bildelement  $L_{s,k}$  wird zunächst der Abstand seines Wertes zu den  $n$  Mittelwerten seiner Normalverteilungen berechnet.

#### 4 Agentenbasierte Detektion, Lokalisation und Klassifikation von Objekten

Jede Verteilung, deren Mittelwert weniger als das 2.5-fache<sup>59</sup> der Standardabweichung entfernt vom Wert von  $\underline{I}_{s,k}$  liegt, erfährt eine Auffrischung ihrer Komponenten gemäß

$$\begin{aligned} w_{i,s,k} &= (1-\alpha)w_{i,s,k-1} + \alpha \\ \underline{\mu}_{i,s,k} &= (1-\beta_{i,k})\underline{\mu}_{i,s,k-1} + \beta_{i,k}\underline{I}_{s,k} \\ \sigma_{i,s,k}^2 &= (1-\beta_{i,k})\sigma_{i,s,k-1}^2 + \beta_{i,k}(\underline{I}_{s,k} - \underline{\mu}_{i,s,k})^T(\underline{I}_{s,k} - \underline{\mu}_{i,s,k}), \end{aligned} \quad (28)$$

mit  $\alpha$  als der primären Lernrate und  $\beta_{i,k}$  dem sekundären Lernfaktor definiert durch

$$\beta_{i,k} = \alpha f_{L_{i,k}|\underline{\mu}_{s,k-1}, \Sigma_{s,k-1}}(\underline{\xi}_{s,k} | \underline{\tilde{\mu}}_{s,k-1}, \underline{\tilde{\Sigma}}_{s,k-1}). \quad (29)$$

Wie auch im Falle der Methode des laufenden Mittelwertes ist durch die Variation des Lernparameters  $\alpha$  eine Beeinflussung der Zeitspanne möglich, die der Algorithmus zur Adaption neuer Informationen benötigt. Für die Normalverteilungen eines Modells, deren Mittelwerte weiter als das 2.5-fache ihrer Standardabweichung vom Wert von  $\underline{I}_{s,k}$  entfernt sind, gilt, dass lediglich ihre Wichtungen gemäß

$$w_{i,s,k} = (1-\alpha)w_{i,s,k-1} \quad (30)$$

verringert werden, wobei sie ihre Mittelwerte und Varianzen beibehalten. Der Sonderfall, dass die Mittelwerte aller Verteilungen mehr als das 2.5-fache ihrer Standardabweichung von  $\underline{I}_{s,k}$  entfernt liegen, führt dazu, dass die Normalverteilung mit der geringsten Wichtung entfernt und durch eine neue, mit Mittelwert gleich  $\underline{I}_{s,k}$ , sowie niedrigem Initialgewicht und hoher anfänglicher Varianz ersetzt wird. Im Anschluss an die Auffrischung werden die  $n$  Verteilungen des Modells gemäß ihrer Fitnesswerte  $w_{i,s,k}/\sigma_{i,s,k}$  geordnet.

Zwecks Berechnung der Entscheidung, ob es sich bei dem Bildelement  $\underline{I}_{s,k}$  um ein zum Vordergrund zugehöriges Element handelt, wird mittels eines weiteren Parameters  $T \in \mathbb{R}^+$  die Anzahl  $m \leq n$  der Distributionen bestimmt, welche zur Modellierung des Hintergrundes tatsächlich beitragen. Für die Bestimmung von  $m$  gilt:

$$m = \arg \min_l \left( \sum_{j=1}^l w_j > T \right), l \leq n. \quad (31)$$

Ein Bildelement  $\underline{I}_{s,k}$  wird als Vordergrund betrachtet, falls sein Wert für jede der  $m$  Distributionen seines Modells mindestens ein definiertes Vielfaches der Standardabweichung des entsprechenden Mittelwertes entfernt liegt [144].

In [144], [145] geben Stauffer und Grimson an, dass im Falle moderater Veränderungen der Umwelt eine hinreichende Separation der Zielobjekte vom Hintergrund möglich ist und deren verlässliche Detektion erlaubt. Schnelle Änderungen des Umgebungslichtes sowie Szenen mit hochüberlappenden Objekten führen jedoch zu einer reduzierten Performanz. Des Weiteren ist der Algorithmus anfällig im Bezug auf den von den bewegten Objekten geworfenen Schatten. Eine Erweiterung des Verfahrens im Hinblick auf schnellere Ausführung und robusteres Verhalten

<sup>59</sup>Die Wahl bzw. Größe der Standardabweichung wird in [144] diskutiert.

## 4 Agentenbasierte Detektion, Lokalisation und Klassifikation von Objekten

gegenüber Schattenwurf wird von KaewTraKulPong et al. in [146] beschrieben. Eine Implementierung dieses Verfahren findet sich in der frei verfügbaren OpenCV-Bibliothek [178] und wird im Rahmen des verfolgten Ansatzes zur Objektdetektion verwendet.

### 4.2.2 IMPLEMENTIERTER ALGORITHMUS ZUR OBJEKTDETEKTION

Der im Rahmen dieser Arbeit implementierte Algorithmus zur Detektion von Objekten basiert auf zwei aufeinander aufbauenden Komponenten, der Bewegungsdetektion sowie der Modellierung des Hintergrunds, welche im Folgenden diskutiert werden.

#### 4.2.2.1 BEWEGUNGSDETEKTION

Die implementierte Bewegungsdetektion verwendet das in Abschnitt 4.2.1.2 diskutierte, adaptive Verfahren der probabilistischen Modellierung des Hintergrundes mittels gewichteter Summen von Normalverteilungen (GMM).

Die Anzahl der nicht als Hintergrund klassifizierten Bildelemente liefert dabei einen Hinweis über das Ausmaß der Bewegung und wird im Rahmen der Bewegungserkennung dazu benutzt, sporadisches Erkennen von Aktivität zu unterdrücken. Als Datengrundlage der Bewegungserkennung dienen die Intensitätsbilder beider Sensoren, welche jeweils getrennt voneinander behandelt werden (Abb. 37).

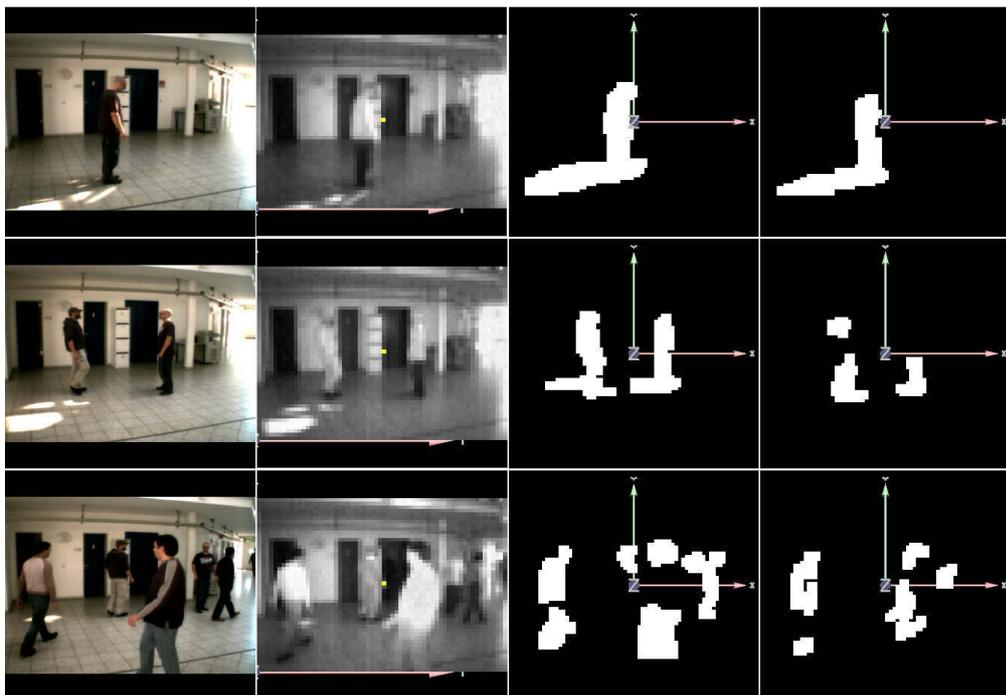


Abbildung 37: Bewegungsdetektionen.

1. Spalte: Hoch aufgelöstes Intensitätsbild des CMOS-Chips. 2. Spalte: Intensitätsbild des PMD-Chips. 3. Spalte: Detektionsresultat basierend auf den Intensitätsdaten des PMD-Chips. 4. Spalte: Detektionsresultat basierend auf den skalierten Intensitätsdaten des CMOS-Chips. In beiden Intensitätsbildern ist der Schattenwurf der Personen deutlich zu erkennen.

#### 4 Agentenbasierte Detektion, Lokalisation und Klassifikation von Objekten

Es gilt:

$$A_k = \begin{cases} 1 & N_t^{PMD} > \eta^{PMD} \vee N_t^{CMOS} > \eta^{CMOS} ; N_t^{PMD}, N_t^{CMOS}, \eta^{PMD}, \eta^{CMOS} \in \mathbb{N}, \\ 0 & \text{sonst,} \end{cases} \quad (32)$$

mit  $A_k$  als dem Aktivitätsindikator des Bildes des Zeitpunkts  $k$ ,  $N_k^{PMD}$  und  $N_k^{CMOS}$  als die Anzahl der als Vordergrund klassifizierten Bildelemente im entsprechenden Sensorbild zum Zeitpunkt  $k$  und  $\eta^{CMOS}$  sowie  $\eta^{PMD}$  als parametrisierbare Schwellwerte, welche die Anzahl an notwendigen Vordergrundpixel festlegen, welche vorhanden sein müssen, um eine Bewegung anzuzeigen.

Der Vorteil dieser parallelen Operation auf den Intensitätsbildern der Sensoren ist, dass eine Bewegungsdetektion selbst im Falle extrem schwacher Tageslichtintensität problemlos möglich ist (Abb. 38). Das Fehlen des Umgebungslichts führt aufgrund von (13) sogar dazu, dass die Güte der erhaltenen Distanzdaten erhöht wird.

Da die Bewegungsdetektion lediglich zur Definition von Perioden der Aktivität im Detektionsfeld der MultiCam benutzt wird, findet der Vorgang der Modellauffrischung permanent statt, d. h. mit jedem neuen Bild wird ein Update der Hintergrundmodelle sowohl der PMD- als auch der CMOS-Bildpixel vorgenommen. Die Wahl der modellspezifischen Lernraten ist in Anbetracht der benutzten Informationen so zu wählen, dass für die Adaption an die Szene einige Sekunden<sup>60</sup> notwendig sind.



Abbildung 38: Bewegungsdetektion in dunklen Umgebungen.

Obere Zeile: Aufnahme eines dunklen Raums ohne Bewegung. Untere Zeile: Sich bewegende Person im Detektionsbereich der MultiCam. 1. Spalte: Informationen des CMOS-Chips. 2. Spalte: Informationen des PMD-Chips. 3. Spalte: Bewegungsdetektion auf Basis der PMD-Daten. 4. Spalte: Fehlende Bewegungsdetektion aufgrund unzureichender Informationen des CMOS-Chips.

Aufgrund der Tatsache, dass lediglich das Wissen, ob eine Bewegung stattgefunden hat, von Interesse ist, wird zum Zwecke der Reduktion der Rechenlast jedes Intensitätsbild des 2D-Chips auf die zum PMD-Chip korrespondierende Größe herunter skaliert ( $640 \times 480 \rightarrow 64 \times 48$ ), was dazu führt, dass die Zeitspanne, welche für die Detektion von Bewegungen pro Bild angesetzt werden muss, unter 1ms liegt. Ein zusätzlicher Vorteil, welcher sich durch die Reduktion der Auflösung ergibt, ist, dass die Empfindlichkeit der Bewegungsdetektion reduziert wird. Die Erzeugung von

<sup>60</sup>Im Verlauf der Akquisition der Daten wurde die Lernrate auf einen Wert gesetzt, sodass das Modell ca. 5s benötigte, um Änderungen in dem Detektionsgebiet vollständig zu integrieren und somit keine neuen Vordergrundinformationen zu erzeugen.

## 4 Agentenbasierte Detektion, Lokalisation und Klassifikation von Objekten

Vordergrundpixel aufgrund minimaler Veränderungen im Hintergrund wird unterdrückt bzw. reduziert.

Zu erwähnen ist, dass eine weitere Verwendung der auf den Intensitätsdaten der Sensoren basierten Bewegungsinformationen nicht stattfindet. Gründe hierfür sind die geringe Auflösung der Intensitätsbilder, die unzureichende Separierung von Vordergrundobjekten sowie der durch die Objekte induzierte Schattenwurf (siehe dazu insbesondere die untere Zeile von Abb. 37).

### 4.2.2.2 HINTERGRUNDMODELLIERUNG

Der Implementierung eines Algorithmus zur Generierung von Hintergrunddaten sind folgende Anforderungen zugrunde gelegt worden:

1. Unabhängigkeit von den Tageslichtverhältnissen.
2. Schnelle Ausführung.
3. Fähigkeit zur Adaption mit variabler Lernrate.

Um die erste Anforderung zu erfüllen, wird im Falle der Hintergrundmodellierung auf die Informationen des PMD-Sensors zurückgegriffen. Verwendung finden dabei die Distanzwerte sowie die Modulationsamplituden der Sensorpixel.

Das gewählte Verfahren, welches die zweite und dritte Bedingung erfüllt, basiert auf der Methode des laufenden Mittelwertes (Abschnitt 4.2.2.1). Hinzugenommen werden zusätzlich die Varianzen der Distanz sowie der Modulationsamplitude der Bildelemente, welche auf Grundlage der mitgeführten Bildhistorie<sup>61</sup> berechnet werden. Die Festlegung, dass die Modellierung des Hintergrunds lediglich in Perioden nichtvorhandener Aktivität<sup>62</sup> erfolgt, wird dadurch begründet, dass die Separation von Vordergrundobjekten und Hintergrund im Wesentlichen auf den naturgemäß unterschiedlichen Tiefendistanzen beruht. Eine Einbindung der Vordergrundobjekte in das Hintergrundmodell ist unerwünscht, da sie zu einer Verringerung der Separationsleistung führt. Weiterhin gilt, dass durch Objektbewegungen die Distanzwerte des Hintergrunds beeinflusst werden. Eine besonders drastische Veränderung der Hintergrunddistanzen ergibt sich dabei durch den Effekt der Multipropagation (Abschnitt 2.2.3), welcher im Falle von Objektbewegungen in geringer Distanz zur MultiCam in Erscheinung tritt.

Zwecks Verminderung der Einflüsse auf die Hintergrunddistanzdaten durch aus dem Sichtfeld der Kamera austretende Objekte erfolgt die Auffrischung des Hintergrunds erst nach einer Periode vorhandener Inaktivität  $\Delta t^{aus}$  (Abb. 39).

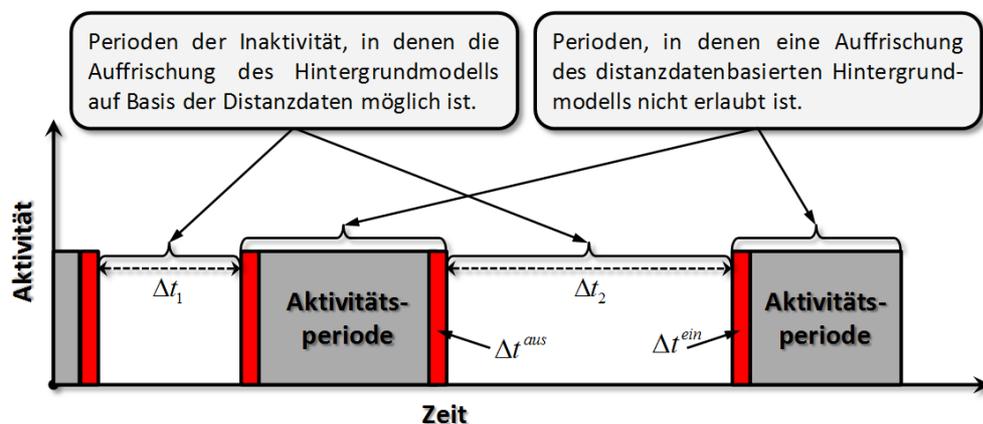


Abbildung 39: Auffrischungsperioden des distanzdatenbasierten Hintergrundmodells.

<sup>61</sup>Im Rahmen der Implementierung wurde die Länge der Historie auf 20 Bilder gesetzt.

<sup>62</sup>Das Vorhandensein von bereits detektierten Objekten entspricht ebenfalls einer Periode der Aktivität.

#### 4 Agentenbasierte Detektion, Lokalisation und Klassifikation von Objekten

Für den Eintritt von Objekten gilt, dass bereits eine bestimmbare Zeitspanne  $\Delta t^{ein}$  vor dem Ereignis die Auffrischung des Hintergrunds abgebrochen wird<sup>63</sup>.

Die Informationen diesbezüglich lassen sich aus der mitgeführten Bildhistorie entnehmen. Eine zusätzlich getroffene Annahme ist, dass zu Beginn der Systemoperation eine gewisse Zeitspanne<sup>64</sup> existiert, in denen lediglich der Hintergrund zu erkennen ist, und welche zur Erzeugung eines initialen Hintergrundmodells verwendet werden kann. Der implementierte Ablauf der distanzdatenbasierten Hintergrundmodellierung zeigt Abb. 40.

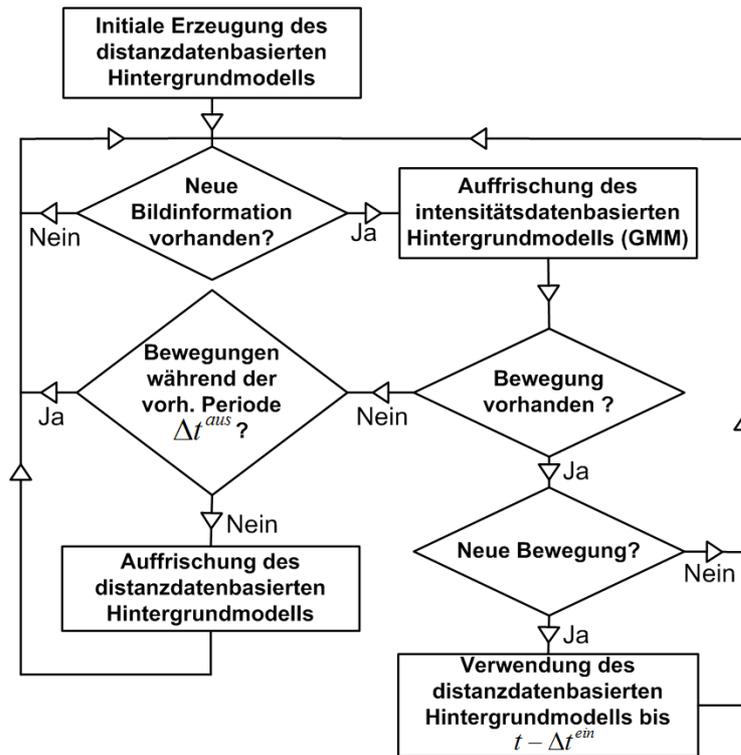


Abbildung 40: Ablauf der Hintergrundmodellerzeugung.

Die gewählte Implementierung zur Modellierung des distanzdatenbasierten Hintergrunds verwendet eine Variation des laufenden Mittelwerts (23). Ersetzt wird die Lernrate  $\alpha$  durch eine gegen sie konvergierende Funktion:

$$f(\alpha, \epsilon, k) = e^{-k\epsilon} \Theta(e^{-k\epsilon} - \alpha) + \alpha \Theta(\alpha - e^{-k\epsilon}) \quad (33)$$

mit  $\alpha \in \mathbb{R}^+$ ,  $\alpha \leq 1$  als Adaptionsgeschwindigkeit,  $\epsilon \in \mathbb{R}^+$ ,  $\epsilon \leq 1$  als Konvergenzfaktor,  $k \in \mathbb{N}$  der Bildnummer und  $\Theta$  der Heaviside-Stufenfunktion. Durch (76) wird gewährleistet, dass im Falle einer notwendigen Neuberechnung des Hintergrundmodells bereits nach wenigen Bildern eine gewisse Information vorhanden ist. Im Wesentlichen benötigt der Algorithmus etwa 10 Bilder, um eine verlässliche Information bezüglich des Hintergrunds generieren zu können (Abb. 41).

Zu erwähnen ist, dass eine komplette Neuberechnung des Hintergrundmodells nicht stattfindet. Setzt die Auffrischung des Hintergrundmodells nach einer Periode der Aktivität wieder ein, wird das vor der Aktivitätsperiode erzeugte Modell als Operationsgrundlage weiterverwendet.

<sup>63</sup>Die Werte  $\Delta t^{aus}$  und  $\Delta t^{ein}$  entsprechen einer zur Akquisition von einer definierbaren Anzahl von Bildern notwendigen Zeitspanne und können dem System während der Initialisierungsphase als zusätzliche Parameter übergeben werden.

<sup>64</sup>Die notwendige Zeitspanne wird in der Implementierung auf die zur Akquisition der vollständigen Bildhistorie (20 Bilder) benötigten Zeit gesetzt und entspricht einer Dauer von ca. 1s.

## 4 Agentenbasierte Detektion, Lokalisation und Klassifikation von Objekten

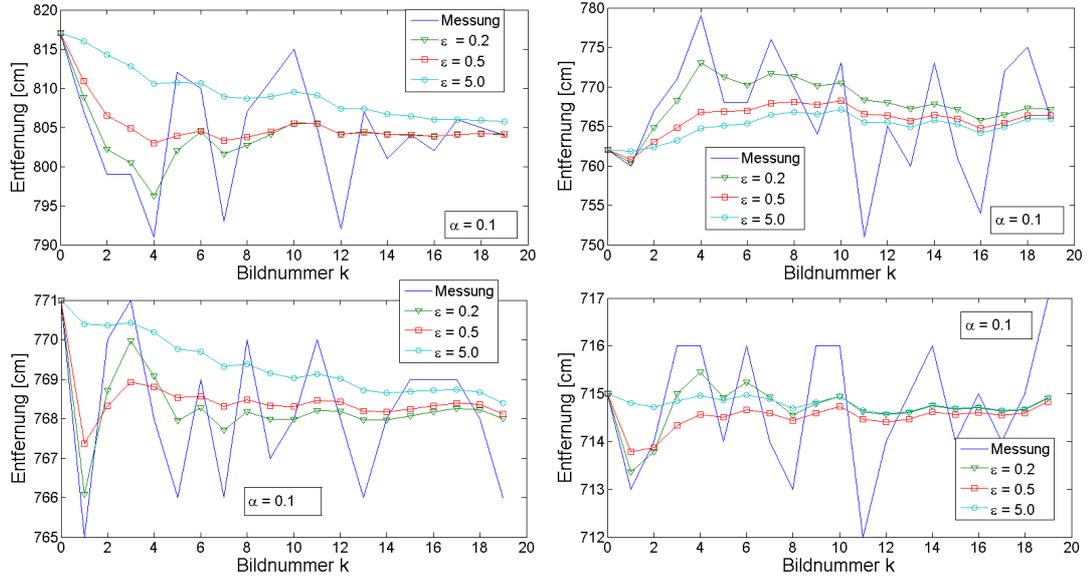


Abbildung 41: Hintergrundmodellierung.

Modellierung des Hintergrundes von vier Pixel mit unterschiedlicher Distanzvarianz unter Verwendung einer Bildsequenz bestehend aus 20 aufeinanderfolgenden Bildern ( $k=0, \dots, 19$ ). Oben links: Hohe Varianz (geringe Modulationsamplitude). Oben rechts: Hohe Varianz. Unten links: Mittlere Varianz. Unten rechts: Geringe Varianz. Die als Messung gekennzeichnete Linie entspricht dem Wert des Pixels per Bild, die restlichen Graphen dem berechneten Hintergrund. Gut zu erkennen ist die durch die Methode des laufenden Mittelwertes durchgeführte Glättung der Entfernungsdaten.

### 4.2.2.3 BESTIMMUNG DER VORDERGRUNDPIXEL

Sei  $H_k^{RA}$  das zum Zeitpunkt  $k$  zur Verfügung stehende Hintergrundbild. Weiterhin sei  $Z_k^{Front}$  eine vorläufige Vordergrundmaske mit im Vergleich zu  $Z_k^{Korr}$  gleicher Dimensionalität. Zur Bestimmung der Vordergrundpixel wird zunächst die Differenz zwischen dem zu betrachtenden Bildelement  $z_{u,v,k}^{Korr} \in Z_k^{Korr}$  und dem Hintergrundmodellpixel  $H_{u,v,k}^{RA}$  bestimmt. Befindet sich die Differenz außerhalb einer durch die Standardabweichung  $\sigma_{u,v,k}^{RA}$  des Hintergrundelementes  $H_{u,v,k}^{RA}$  definierten Grenze, wird das zugehörige Element in der Vordergrundmaske markiert:

$$z_{u,v,k}^{Front} = \begin{cases} z_{u,v,k}^{Korr} & |H_{u,v,k}^{RA} - z_{u,v,k}^{Korr}| > \sigma_{u,v,k}^{RA} c_D \wedge H_{u,v,k}^{RA} - z_{u,v,k}^{Korr} \geq 0, \\ -z_{u,v,k}^{Korr} & |H_{u,v,k}^{RA} - z_{u,v,k}^{Korr}| > \sigma_{u,v,k}^{RA} c_D \wedge H_{u,v,k}^{RA} - z_{u,v,k}^{Korr} < 0, \\ 0 & \text{sonst.} \end{cases} \quad (34)$$

Die Variable  $c_D \in \mathbb{R}^+$  in (34) entspricht dabei einem wählbaren Skalierungsfaktor, dessen Größe durch Analyse der Vordergrundinformationen zum Zeitpunkt der Implementierung zu bestimmen ist. Im Allgemeinen führt ein zu geringer Wert dazu, dass im Falle vorhandener Objektbewegungen eine hohe Zahl von Hintergrundpixel falsch eingestuft werden (Abb. 42).

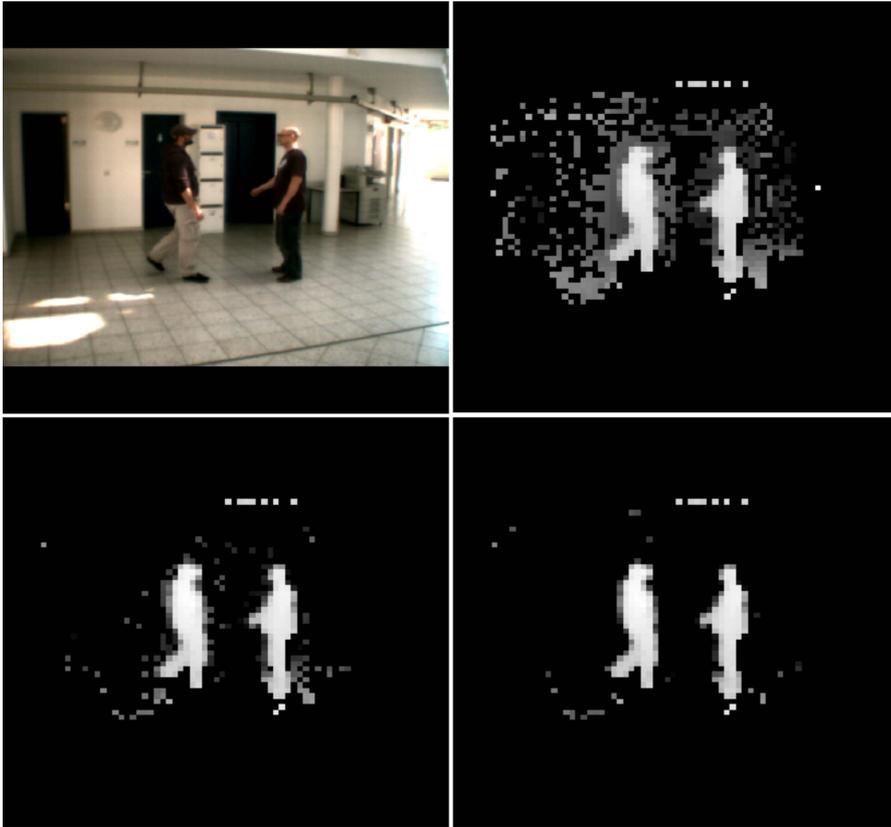


Abbildung 42: Vordergrundpixel unter Verwendung unterschiedlicher Schwellen.

Oben links: Intensitätsbild CMOS-Sensor. Oben rechts: Vordergrundpixel für einen Skalierungsfaktor von  $c_D=1$ . Unten links: Skalierungsfaktor  $c_D=3$ . Unten rechts: Skalierungsfaktor  $c_D=6$ . Für die restlichen Parameter gelten:  $c_M=2.5$ ,  $\alpha=0.1$ ,  $\epsilon=0.5$ .

Eine falsche Einstufung begründet sich durch die Tatsache, dass Objektbewegungen eine veränderte Beleuchtung des Hintergrunds implizieren. Diese veränderte Bestrahlungsintensität hat zur Folge, dass die entsprechenden Pixel auf dem PMD-Chip sowohl veränderte Distanzwerte als auch Modulationsamplituden aufweisen und irrtümlich als Vordergrund klassifiziert werden. Eine Untersuchung bzgl. der Veränderung einer Distanz unter Verwendung unterschiedlicher Modulationsamplituden findet sich in [62]. Zu erwähnen ist noch, dass die Berechnung der Varianzen der Tiefenwerte  $\sigma_{u,v,k}^{RA}$  parallel zu der Erzeugung des Hintergrundmodells unter Verwendung der Informationen aus der mitgeführten Bildhistorie durchgeführt wird.

Im Folgeschritt werden alle Elemente aus  $Z_k^{Front}$  mit  $z_{u,v,k}^{Front} \neq 0$  betrachtet. Für Werte größer 0 wird angenommen, dass die entsprechenden Pixel dem Vordergrund zuzuordnen sind. Pixelwerte mit  $z_{u,v,k}^{Front} < 0$  bedürfen einer zusätzlichen Untersuchung.

Sei zunächst  $m_{u,v,k} \in M_k$  die Modulationsamplitude des betrachteten Pixel und  $H_{u,v,k}^{RA-Mod}$  die Modulationsamplitude des Modulationsdaten-Hintergrundmodells, welche ebenfalls mithilfe des laufenden Mittelwertes gebildet wird. Durch erneute Schwellwertbildung  $\forall z_{u,v,k}^{Front} < 0$  mittels

$$z_{u,v,k}^{Front} = \begin{cases} -z_{u,v,k}^{Front} & H_{u,v,k}^{RA-Mod} - m_{u,v,k} > H_{u,v,k}^{RA-Mod} c_M ; c_M \in \mathbb{R}^+, \\ 0 & \text{sonst,} \end{cases} \quad (35)$$

## 4 Agentenbasierte Detektion, Lokalisation und Klassifikation von Objekten

---

können die Pixel herausgefiltert werden, welche durch das Verlassen des Eindeutigkeitsintervalls erneut in den zu untersuchenden Distanzbereich hineintransformiert werden. Ihre Detektion ist, bedingt durch die relativ geringe Modulationsamplitude aufgrund ihrer großen Entfernung, einfach.

Die letzte Stufe im Rahmen einer Vordergrundseparation dient der Entfernung isolierter Pixel. Zunächst werden mithilfe eines simplen Kerns jegliche Pixel von  $Z_k^{Front}$  entfernt, welche weniger als eine definierte Anzahl  $\eta \in \mathbb{N}$  an direkten Nachbarn besitzen.

Der zweite Schritt dient der Elimination von Distanzspitzen. Für jedes übrig gebliebene Element aus  $Z_k^{Front}$  wird untersucht, wie weit dieses von seinen direkten Nachbarn entfernt ist. Überschreitet die maximale Distanz eine definierte Entfernung  $\zeta \in \mathbb{R}^+$ , so gilt dieses Pixel als nicht valide und wird entfernt. Übrig bleibt eine Distanzmatrix  $Z_k^{Front}$  mit von 0 verschiedenen Elementen, die den Vordergrund des momentanen Bildes enthält (Abb. 43).



Abbildung 43: Vordergrundseparation in Szenen mit mehreren Personen.

Separation wurde durchgeführt unter Verwendung des Parametersatzes:  $c_D=6$ ,  $c_M=2$ ,  $\alpha=0.1$ ,  $\zeta=200$ .

### 4.2.3 OBJEKTLOKALISATION

Der im Rahmen dieser Arbeit verfolgte Ansatz zur Lokalisation von Objekten basiert auf der mittels Clustering durchgeführten Untersuchung der Daten von Aktivitätsclustern, welche auf Grundlage einer Analyse des Tiefenverlaufs der separierten Vordergrunddaten erzeugt werden.

Clustering kann abstrakt als ein Prozess der Einteilung von Objekten in Ähnlichkeitsklassen beschrieben werden [148]. Anwendungsgebiete sind u. a. das Data-Mining sowie der Bereich des maschinellen Lernens, in dem das Clustern als eine Methode des unbeaufsichtigten Lernens interpretiert wird. Eine allgemeine Übersicht existierender Verfahren wird in [149] gegeben, eine Übersicht räumlicher Clusteringalgorithmen findet sich in [148]. Da die von der MultiCam gelieferten Daten räumliche Informationen beinhalten, werden im Folgenden zwei häufig im Bereich des spatialen Clusterings verwendete Algorithmen diskutiert, wobei im Anschluss dazu der implementierte Lokalisationsalgorithmus präsentiert wird.

#### 4.2.3.1 PROBABILISTISCHES CLUSTERING

Der Ansatz des probabilistischen Clusterings liegt darin, die zu analysierenden Daten als durch ein unbekanntes Modell generierte Proben zu betrachten [150]. Angenommen wird dabei, dass das Modell sich aus einer Mischung von Wahrscheinlichkeitsverteilungsfunktionen zusammensetzt. In der Regel wird auf Normalverteilungen zurückgegriffen, was letztendlich keine Einschränkung darstellt, da jede Verteilungsfunktion durch eine Mixtur aus Normalverteilungen angenähert werden kann [152].

Unter diesen Annahmen ergibt sich die Möglichkeit einer einfachen Zuordnung eines Clusters zu der Wahrscheinlichkeitsverteilung, welche die Elemente des Clusters mit der höchsten Wahr-

## 4 Agentenbasierte Detektion, Lokalisation und Klassifikation von Objekten

---

scheinlichkeit erzeugt. Dieses führt automatisch auch dazu, dass die generierten Daten, neben ihren offensichtlichen Attributen, zusätzlich latente Information besitzen, welche die Zugehörigkeit zu einem Cluster beinhalten. Ziel des Clusterings ist es, diese nicht offensichtliche Eigenschaft in den zur Verfügung stehenden Daten herauszufinden.

### EXPECTATION-MAXIMIZATION-ALGORITHMUS (EM-ALGORITHMUS)

Ein häufig verwendetes probabilistisches Verfahren zur Erzeugung eines Modells zwecks Bestimmung latenter Variablen in Daten ist die Verwendung einer Maximum-Likelihood-Methode in Form des *Expectation-Maximization-Algorithmus*. In jeweils zwei aufeinanderfolgenden Schritten wird eine Verfeinerung der Modellparameter iterativ durchgeführt, bis ein notwendigerweise a priori vorhandenes Konvergenzkriterium erfüllt wird. Im Folgenden wird der Ablauf des EM-Algorithmus in knapper Form wiedergegeben. Eine detaillierte Beschreibung sowie Analyse des Konvergenzverhaltens kann in [151] gefunden werden.

Sei  $X = \{\underline{x}_i\}$ ,  $\underline{x}_i \in \mathbb{R}^n$ ,  $i = 1, \dots, N$  die Menge der zu analysierenden Daten, von denen angenommen wird, dass sie durch eine Mischung von  $M$  Normalverteilungen erzeugt werden:

$$f_{X|\Theta}(\Xi | T) = \sum_{m=1}^M w_m f_{X|\Theta_m}(\Xi | \tau_m = \{\underline{\mu}_m, \Sigma_m\}) \quad \text{mit} \quad \sum_{m=1}^M w_m = 1. \quad (36)$$

$\Theta = \{\Theta_m, w_m\}$ ,  $m = 1, \dots, M$  entspricht dabei der gesuchten Parametermenge des Modells, mit  $w_m$  als Wichtung und  $\Theta_m$ , der Parametrisierung der  $m$ -ten Wahrscheinlichkeitsfunktion. Die Menge  $\Xi = \{\xi_i\}$  enthält die zu den Elementen der Menge  $X$  zugehörigen Dummyvariablen  $\xi_i \in \mathbb{R}^n$  [180]. Analog dazu verhält sich  $T = \{\tau_1, \dots, \tau_m\}$  zu  $\Theta$ .

### INITIALISIERUNGSPHASE

Der erste Schritt besteht in der Initialisierung der gesuchten Parameter der einzelnen Verteilungen, bestehend aus Erwartungswert  $\underline{\mu}_j$ , Kovarianzmatrix  $\Sigma_j$  und Wichtung  $w_j$ , mit zufälligen oder durch etwaiges Vorwissen erzeugten Werten.

#### 1. SCHRITT: BERECHNUNG DER ERWARTUNGSWERTE (*EXPECTATION*)

Basierend auf den momentanen Parametern werden  $\forall \underline{x}_i \in X$  zunächst die Einflüsse der  $M$  Cluster im Hinblick auf ihre Erzeugung berechnet:

$$p\{\omega: j = \kappa \mid \underline{x}_i = \xi_i, \Theta = T\} = \frac{w_j f_{\underline{x}_i|\Theta_j}(\xi_i | \tau_\kappa)}{\sum_{m=1}^M w_m f_{\underline{x}_i|\Theta_m}(\xi_i | \tau_m)}, \quad (37)$$

mit  $\kappa$  als einer zur  $j$ -ten Verteilung zugehörigen Dummyvariablen.

#### 2. SCHRITT: MAXIMIERUNGSPHASE (*MAXIMIZATION*)

Die dritte Phase entspricht einer Parameterverfeinerung zwecks Maximierung einer Zielfunktion, wobei auf die Erwartungswerte der Estimationsphase (37) zurückgegriffen wird:

#### 4 Agentenbasierte Detektion, Lokalisation und Klassifikation von Objekten

$$w_j^{Neu} = \frac{1}{n} \sum_{i=1}^n p\{\omega: j=\kappa | x_i = \xi_i, \Theta = T\}, \quad (38)$$

$$\mu_j^{Neu} = \frac{\sum_{i=1}^n \xi_i p\{\omega: j=\kappa | x_i = \xi_i, \Theta = T\}}{\sum_{i=1}^n p\{\omega: j=\kappa | x_i = \xi_i, \Theta = T\}}, \quad (39)$$

$$\Sigma_j^{Neu} = \frac{\sum_{i=1}^n p\{\omega: x_i \in C_j | x_i(\omega) = \xi_i, \Theta_i = T_i\} (\xi_i - \mu_{j,t+1}) (\xi_i - \mu_{j,t+1})^T}{\sum_{i=1}^n p\{\omega: j=\kappa | x_i = \xi_i, \Theta = T\}}. \quad (40)$$

Als Zielfunktion dient der Logarithmus von (36):

$$\Lambda = \ln(f_{X|\Theta}(\Xi|T)) = \ln \prod_{i=1}^n f_{x_i|\Theta}(\xi_i|T) = \sum_{i=1}^n \ln \left( \sum_{m=1}^M w_m f_{x_i|\Theta_m}(\xi_i|\tau_m) \right). \quad (41)$$

Nach jedem Maximierungsschritt wird die Differenz von (41) mit dem Ergebnis der vorherigen Iteration verglichen. Unterschreitet die Differenz  $\Lambda^{Neu} - \Lambda$  einen bestimmten Schwellenwert, dann entspricht das einer hinreichenden Konvergenz und der Algorithmus stoppt. Falls nicht, erfolgt eine erneute Iteration aus Estimation und Maximierung. Die generierten Parameter  $\Theta^{Neu}$  definieren eine Modellkonfiguration und bieten die Möglichkeit, latente Variablen der Elemente von  $X$ , z. B. Clusterzugehörigkeit, zu bestimmen.

Aufgrund der Eigenschaft des EM-Algorithmus gegen lokale Minima zu konvergieren [151], ist es durchaus möglich, für einen Datensatz  $X$ , je nach Initialisierung der Parameter  $\Theta$ , unterschiedliche Modelle zu erhalten (Abb. 44). Eine schlechte Wahl der Initialwerte, z. B. mittels des Zufallsprinzips, führt aufgrund des Konvergenzverhaltens gegen lokale Minima unter Umständen zu ungeeigneten Modellparametern (Abb. 44, rechts). Es ist daher sinnvoll, etwaiges Vorwissen in die Initialisierung mit einfließen zu lassen, zumal sich dadurch auch die Berechnungsdauer verkürzen lässt.

Aufgrund seines probabilistischen Ansatzes wird der EM-Algorithmus zu der Klasse der weichen Klassifikatoren gezählt. Betrachtet man Gleichung (37), dann erkennt man, dass für ein beliebiges Element  $x_i \in X$  keine explizite Zugehörigkeit zu einem Cluster existiert. Vielmehr ist jedes Element von  $X$  aufgrund des probabilistischen Ansatzes gleichzeitig Mitglied eines jeden Clusters, jedoch mit unterschiedlicher Wahrscheinlichkeit. Eine Eigenschaft des probabilistischen Ansatzes ist somit, dass leichte Variationen von Datenelementen nicht zu einer sprunghaften Änderung der Clusterzugehörigkeit führen.

Vorteile, welche sich aus der Verwendung des probabilistischen Clusterings ergeben, sind [149]:

1. Die Möglichkeit der Adaption an komplexe Datenstrukturen.
2. Der Algorithmus kann beliebig gestoppt und wiedergestartet werden, was Operationen auf Datenblöcken ermöglicht (*batch processing*).
3. Zu jedem Zeitpunkt der Iteration kann auf das gewonnene Modell zurückgegriffen werden, um etwaige Unterscheidungen bereits im Vorfeld durchzuführen.

## 4 Agentenbasierte Detektion, Lokalisation und Klassifikation von Objekten

4. Das Resultat ist ein einfach zu interpretierendes Modell. Insbesondere liefert eine Untersuchung der Kovarianzmatrizen  $\Sigma_j$  Hinweise bezüglich der Form der untersuchten Daten bzw. der jeweiligen Cluster.

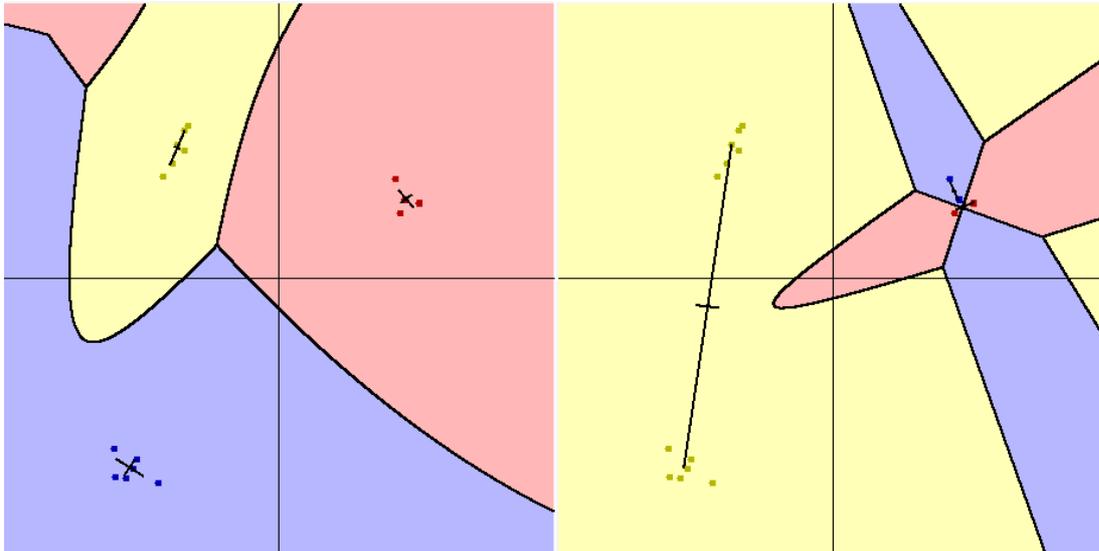


Abbildung 44: Clusteringergebnisse unterschiedlicher Initialisierungen.

Variationen in den Initialisierungswerten des EM-Algorithmus können aufgrund der Konvergenz gegen lokale Minima zu unterschiedlichen Modellen führen (Erstellung der Graphen erfolgte mithilfe der in [153] zu findenden Demoapplikation).

### 4.2.3.2 DETERMINISTISCHES CLUSTERING K-MEANS

Ein im Vergleich zum probabilistischen Ansatz unterschiedliches, ebenfalls iteratives Verfahren ist das deterministische k-Means Clustering [154], [155]. Gemäß [149] gehört k-Means, nicht zuletzt aufgrund der geringen Komplexität sowie der damit bedingten schnellen Ausführbarkeit, zu einer der am häufigsten eingesetzten Methoden zum Clustern von Daten in wissenschaftlichen und industriellen Anwendungen.

Sei zunächst wieder  $X = \{x_i\}$ ,  $x_i \in \mathbb{R}^k$ ,  $i = 1, \dots, n$ , die in  $M \in \mathbb{N}$  Klassen zu separierende Datenmenge. Im k-Means-Algorithmus wird jeder dieser Klassen bzw. Cluster  $C_j$ ,  $j = 1, \dots, M$ , durch den Mittelwert  $c_j \in \mathbb{R}^k$  der darin enthaltenen Punkte (Zentroide) repräsentiert, welche es gemäß eines Optimierungskriteriums zu berechnen gilt. Der prinzipielle Ablauf ist gegeben durch:

1. **Initialisierung:** Initialisierung der  $c_1, c_2, \dots, c_M \in \mathbb{R}^k$  (plausibel oder zufällig).
2. **Iteration** (bis zum Erfüllen des Konvergenzkriteriums):

$$a) \quad \forall i \text{ definiere } \mu_i := \operatorname{argmin}_j \|x_i - c_j\|_p,$$

$$b) \quad \forall j \text{ setze } c_j := \frac{\sum_{i=1}^n x_i \delta(\mu_i - j)}{\sum_{i=1}^n \delta(\mu_i - j)}.$$

#### 4 Agentenbasierte Detektion, Lokalisation und Klassifikation von Objekten

Der prinzipielle Ablauf besitzt eine gewisse Ähnlichkeit zum EM-Algorithmus. Anfänglich werden die Mittelwerte der Cluster unter Zuhilfenahme eventuell vorhandenem Vorwissen initialisiert<sup>65</sup>. Im ersten Iterationsschritt werden die Elemente der Datenmenge  $X$  gemäß ihres Abstands zu den Clusterzentren geordnet.

Im zweiten Schritt werden die Zentroiden neu berechnet und mit dem Wert der vorigen Iteration verglichen. Findet eine Veränderung unterhalb eines definierten Konvergenzkriteriums statt, wird die Berechnung abgebrochen, falls nicht, startet eine weitere Iteration. Die obig angegebene Ausführungsvorschrift dient der Optimierung der totalen "Inter-Cluster-Varianz" [148]:

$$E = \sum_{j=1}^m \sum_{x \in C_j} \|x - c_j\|_p. \quad (42)$$

Diese kann unter Einführung einer auf den Mengen  $C = \{c_1, c_2, \dots, c_M\}$  und  $M = \{\mu_1, \mu_2, \dots, \mu_n\}$  als operierende Funktion formuliert werden:

$$F(M, C) = \sum_{i=1}^n \|x_i - c_{\mu_i}\|_p. \quad (43)$$

Der Vorteil dieser Formulierung lässt sich durch die erneute Betrachtung des Iterationsschrittes des k-Means wiedergeben. Zunächst entspricht Schritt a) einer Minimierung von  $F$  in Bezug auf  $M$  unter Beibehaltung von  $C$ . Schritt b) kehrt diesen Sachverhalt um und es findet eine Minimierung von  $F$  in Bezug auf  $C$  statt, wobei nun  $M$  festgesetzt ist. Dieses lässt sich als ein Gradientenabstieg auf den Parametern von  $F$  interpretieren und führt aufgrund der fallenden Monotonie zu einer Konvergenz gegen ein Minimum. Die Bewegung entlang des Gradienten stellt jedoch keineswegs sicher, dass k-Means gegen ein absolutes Extremum konvergiert, was dazu führen kann, dass dieser Algorithmus ebenfalls anfällig für lokale Minima ist (Abb. 45).

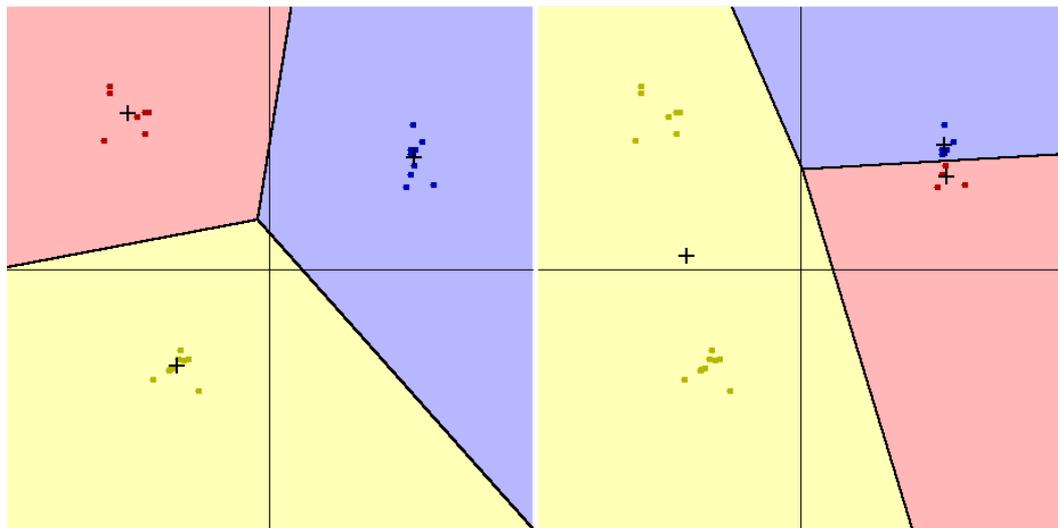


Abbildung 45: Ergebnisse eines auf drei Clustern operierenden k-Means-Algorithmus.

Zu erkennen ist die Sensitivität des Algorithmus bzgl. der verwendeten Initialisierungswerte (Erstellung der Graphen erfolgte mithilfe der in [153] zu findenden Demoapplikation).

<sup>65</sup>Im Falle von nicht vorhandenem Vorwissen findet eine zufällige Initialisierung statt.

## 4 Agentenbasierte Detektion, Lokalisation und Klassifikation von Objekten

---

Die Vorteile, die für eine Verwendung von k-Means sprechen, sind [149]:

1. Operation auf beliebiger  $p$ -Norm.
2. Parallelisierbarkeit und die damit einhergehende Beschleunigung der Berechnung.
3. Nichtsensitiv in Bezug auf die Anordnung bzw. Reihenfolge der Daten.

Im Gegensatz zum EM-Algorithmus gehört k-Means zu der Klasse der harten Klassifikatoren. Das bedeutet, dass jedes Element einer Menge  $X$  eine exklusive Zugehörigkeit zu einem Cluster besitzt. Die kann dazu führen, dass minimale Verschiebungen der Vektoren, im Gegensatz zu EM, in einer veränderten Clusterzuweisung münden.

Obwohl k-Means, bedingt durch seine einfache Struktur, in Bezug auf die benötigte Rechenleistung relativ schnell konvergiert, ist der gelieferte Informationsgehalt bzgl. der segmentierten Daten, aufgrund der harten Klassifikation, gering. Eine Möglichkeit, die Informationsmenge zu steigern ist, den EM-Algorithmus zusätzlich als einen nachgeschalteten Prozessschritt einzusetzen. Dazu wird zunächst k-Means verwendet, um rasch Clusterzentren zu bilden, welche dann im Folgeschritt als Erwartungswerte dem EM-Algorithmus zur Verfügung gestellt werden. Der Vorteil dieser Vorgehensweise liegt in der beschleunigten Berechnung der latenten Informationen mittels EM. Eine Implementierung dieser Vorgehensweise findet sich z. B. in der freien OpenCV-Bibliothek [178].

### 4.2.4 IMPLEMENTIERTER ALGORITHMUS ZUR OBJEKTLOKALISATION

Eine grundsätzliche Problematik bei der Verwendung von Clustering-Algorithmen zur Bildung von Ähnlichkeitsklassen ist das benötigte Vorwissen bezüglich der zu verwendenden Anzahl an Cluster sowie der Initialisierung der Startparameter. Die bereits diskutierte Sensibilität der Algorithmen bezüglich der gewählten Anfangsparameter führt dazu, dass eine zufällige Initialisierung eventuell in einer unzureichenden Modellbildung mündet und eine sinnvolle Separation der Daten verhindert (Abb. 44, 45). Wünschenswert ist es daher, bereits vorab berechnete Informationen bezüglich der Anzahl der Cluster sowie plausible Startwerte zu erzeugen, um auf Grundlage dieses Vorwissens eine verlässliche Modellierung durchführen zu können. Der im Rahmen dieser Arbeit verwendete Ansatz zur Generation von Vorwissen basiert auf der Erzeugung von Aktivitätsclustern. Die folgenden Abschnitte dienen der Erläuterung dieser Erzeugung.

#### 4.2.4.1 ERZEUGUNG DER AKTIVITÄTSKONTUR

Grundlage der Erzeugung von Aktivitätsclustern sind die Informationen einer Gradientenanalyse einer auf Basis der Daten aus  $Z_k^{Front}$  (35) gewonnenen Aktivitätskontur. Die verwendeten Distanzdaten beruhen auf dem nativen Koordinatensystem  $K^{3D}$  des PMD-Chips, dessen Auflösung durch  $u_{max}$  in horizontaler und  $v_{max}$  in vertikaler Richtung festgelegt ist (Abb. 46).

Zur Bestimmung der Aktivitätskontur werden zunächst die dreidimensionalen Werte von  $Z_k^{Front}$  auf die  $u-z$ -Ebene von  $K^{3D}$  projiziert. Der erste Schritt diesbezüglich besteht in der Bildung des Medians für jede der  $u_{max}$  Spalten von  $Z_k^{Front}$ . Die Wahl des Medians wird durch sein stabiles Verhalten gegenüber eventuell vorhandenen Ausreißern motiviert. Eine Annahme, welche getroffen wird, ist, dass die Spalte zur Bildung des Medians eine Mindestanzahl von  $N_{median}$  Distanzwerte größer null aufweisen muss. Ist das nicht der Fall, wird der entsprechende Spaltenmedian auf 0 gesetzt. Das Ergebnis ist ein Medianvektor

$$\underline{\mu}_k = [\mu_1, \mu_2, \dots, \mu_{u_{max}}]^T, \quad (44)$$

#### 4 Agentenbasierte Detektion, Lokalisation und Klassifikation von Objekten

dessen Abbildung über die  $u$ -Koordinate von  $K^{3D}$  das Aktivitätsprofil des Zeitpunkts  $k$  wiedergibt (Abb. 47).

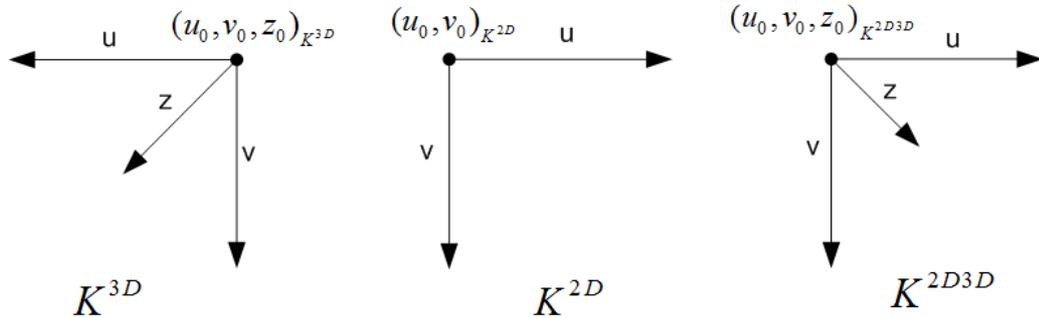


Abbildung 46: Sensorkoordinatensysteme der MultiCam.

Links: Sensorkoordinatensystem des PMD-Sensors. Mitte: Sensorkoordinatensystem des CMOS-Sensors. Rechts: Gemeinsames Koordinatensystem beider MultiCam-Sensoren. Zu erwähnen ist, dass die Koordinatensysteme so gewählt worden sind, dass die  $u-v$ -Ebenen von  $K^{2D}$  und  $K^{2D3D}$  sowie die  $z$ -Richtung von  $K^{3D}$  und  $K^{2D3D}$  übereinstimmen.

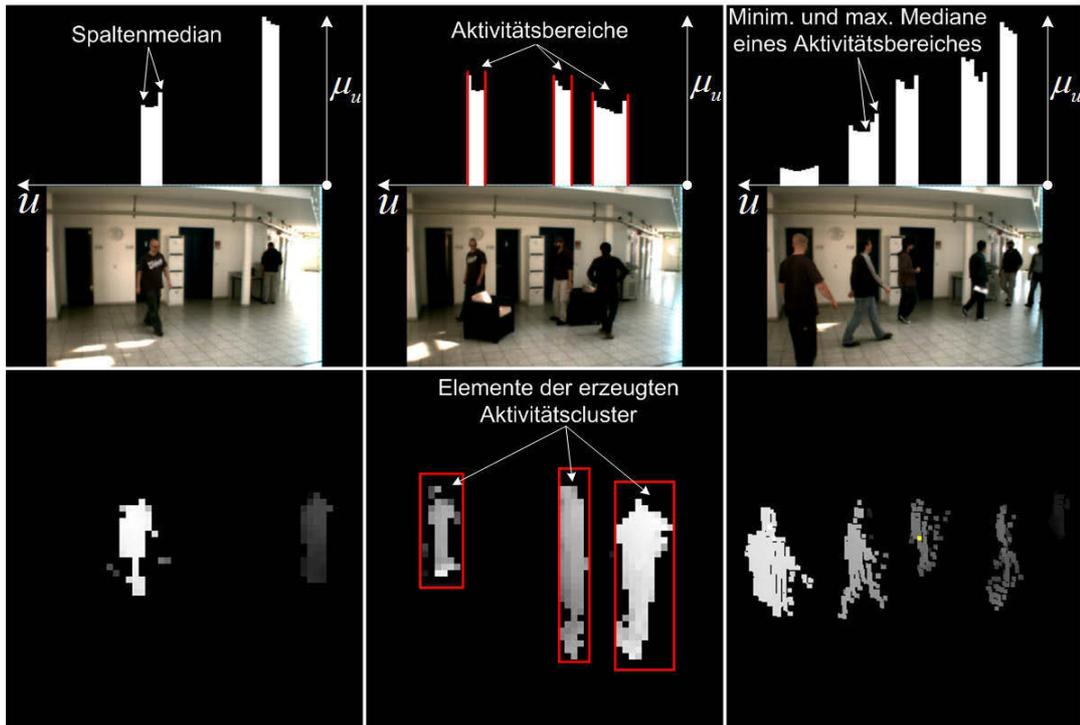


Abbildung 47: Aktivitätsprofile für drei Szenen mit unterschiedlicher Personenzahl.

Obere Zeile: 2D-Bild mit überlagertem Aktivitätsprofil. Durch Gradientenanalyse des Aktivitätsprofils ergibt sich die jeweilige Anzahl von Aktivitätsclustern und entspricht nur im Idealfall nichtvorhandener Verdeckung der Anzahl der im Detektionsfeld vorhandenen Objekte bzw. Personen. Für die Erzeugung von Medianwerten wurde der empirische Wert  $N_{median} = 4$  verwendet. Untere Zeile: Ergebnisse der Hintergrundseparation.

### 4.2.4.2 BILDUNG DER AKTIVITÄTSCLUSTER

Durch Lokalisation von Übergängen von oder nach 0 von Medianwerten größer Null im Medianvektor (44) des momentanen Bildes  $Z_k^{Front}$ , ergeben sich Informationen bezüglich der Position sowie Breite vorhandener Aktivitäten (Abb. 47). Anschließend werden für jeden der sich pro Zeitpunkt ergebenden  $n$  Aktivitätsbereiche die entsprechenden Bildelemente aus  $Z_k^{Front}$  in einem Aktivitätscluster  $C_k^j$ ,  $j=1, \dots, n$ , mit  $C_k^j \subseteq Z_k^{Front}$ , gesammelt.

Der verfolgte Ansatz zur Lokalisation eines Objekts basiert auf der Analyse der durch die Menge  $C_k^j$  definierten Distanzdaten (Abb. 47) unter Verwendung des EM-Algorithmus (Abschnitt 4.2.3.1). Der Grund für die Wahl dieses probabilistischen Verfahrens liegt in der weichen Klassifizierung der Nutzdaten. Die zusätzlichen Informationen, welche der EM-Algorithmus im Vergleich zu hart klassifizierenden Algorithmen, wie k-Means (Abschnitt 4.2.3.2) liefert, wird dazu benutzt, Untersuchungen der Plausibilität durchzuführen, welche Gegenstand der Diskussion des folgenden Abschnittes sind.

Auf Basis der vorhandenen Tiefeninformationen eines Aktivitätsclusters wird eine Einteilung in einen Vorder- und einen Hintergrundcluster durchgeführt, wobei Ersterem eine besondere Bedeutung zukommt. Die von dem EM-Algorithmus in den Vordergrundcluster eingeteilten Distanzdaten entsprechen den Daten eines lokalisierten Objekts und werden im Folgenden als die primäre Partition des Aktivitätsclusters bezeichnet. Aufgrund vorhandenen Vorwissens, definiert durch den minimalen und maximalen Wert  $\mu_{min,t}^j$ ,  $\mu_{max,t}^j$  der Spaltenmediane des zum Aktivitätscluster zugehörigen Aktivitätsbereiches (Abb. 47), bietet sich die Möglichkeit der gezielten Initialisierung der Modellparameter. Eine Estimation durch einen zusätzlichen Algorithmus, z. B. k-Means [188], ist nicht notwendig und führt aufgrund der verringerten benötigten Rechenleistung zu einer beschleunigten Ausführung.

Jedem vorhandenen Aktivitätscluster  $C_k^j$  wird nach Bestimmung seiner primären Partition durch den EM-Algorithmus eine Parametermenge  ${}^{2D3D}c_k^j$  zugewiesen, welche die Informationen bezüglich der Größe des Aktivitätsclusters sowie der primären Partition enthalten. Codiert werden diese Informationen in Form von vier Punkten im Koordinatensystem  $K^{2D3D}$ , von denen jeweils zwei ein Rechteck definieren, welches die entsprechenden Daten der primären Partition bzw. des gesamten Aktivitätsclusters umranden (Abb. 48). Da die ursprünglichen Positionen dieser vier Punkte sich auf das Koordinatensystem  $K^{3D}$  des PMD-Sensors beziehen, werden sie im Anschluss an ihre Berechnung in das Koordinatensystem  $K^{2D3D}$  transformiert (Abb. 46).

Für die zu erzeugende Objektposition gilt, dass sie sich aus zwei Informationen zusammensetzt. In der  $u-v$ -Ebene von  $K^{2D3D}$  ergibt sich die Objektposition durch den Mittelpunkt des Rechtecks der primären Partition (Abb. 48). Als Position in  $z$ -Richtung wird das durch den EM-Algorithmus berechnete Zentrum des Vordergrundclusters verwendet.

Zwecks Erhöhung des Identifikationspotenzials des Aktivitätsclusters werden der Parametermenge zwei zusätzliche Vektoren  $\eta_k^j$  und  $\iota_k^j$  mit  $\eta_k^j, \iota_k^j \in \mathbb{R}^{N_{median}}$  hinzugefügt. Die Komponenten von  $\eta_k^j$  enthalten die  $N_{median}$  höchsten Modulationsamplituden der Elemente der primären Partition, wobei diese im Vektor nach absteigender Größe geordnet werden. Der Intensitätsvektor  $\iota_k^j$  enthält die zu den ausgewählten  $N_{median}$  Pixel zugehörigen Intensitätswerte, wobei die Ordnung dieser in dem Vektor  $\iota_k^j$  durch die Reihenfolge der Elemente in  $\eta_k^j$  bestimmt wird. Für die Parametermenge  ${}^{2D3D}c_k^j$  eines Aktivitätsclusters  $C_k^j$  im Koordinatensystem  $K^{2D3D}$  gilt somit, dass seine Zusammensetzung durch

$${}^{2D3D}c_k^j = \{ {}^{2D3D}P_k^j, {}^{2D3D}P_{1,k}^{j,PP}, {}^{2D3D}P_{2,k}^{j,PP}, {}^{2D3D}P_{1,k}^{j,AC}, {}^{2D3D}P_{2,k}^{j,AC}, \eta_k^j, \iota_k^j \} \quad (45)$$

## 4 Agentenbasierte Detektion, Lokalisation und Klassifikation von Objekten

bestimmt ist. Sie bildet die Basisinformation für den nachgeschalteten Prozess der Objektklassifikation sowie der Objektverfolgung (Kapitel 5).

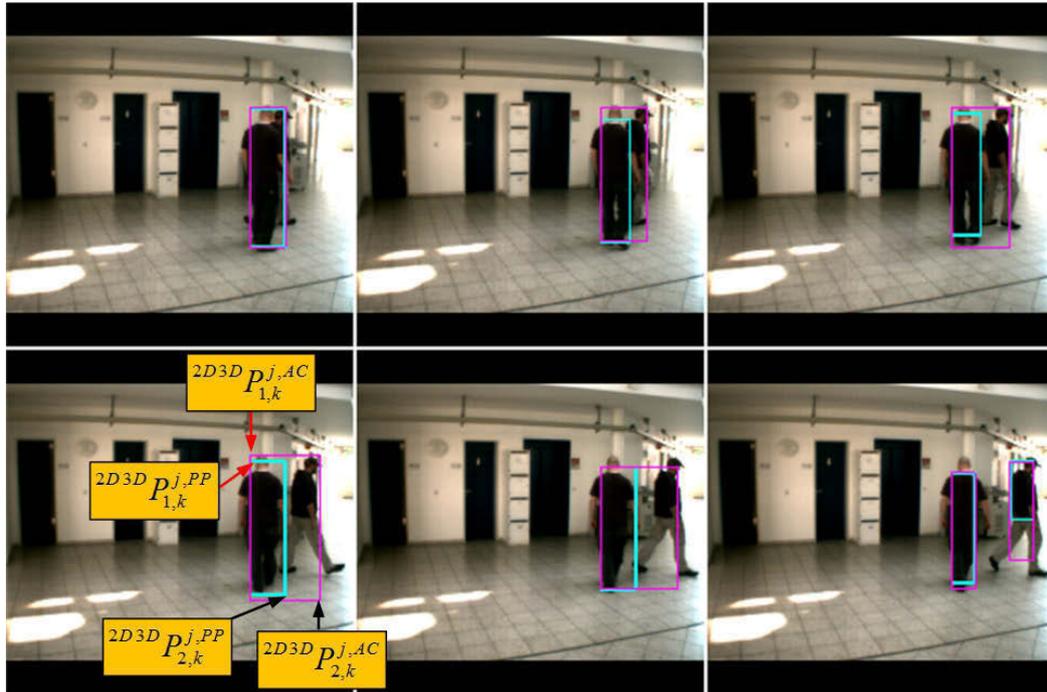


Abbildung 48: Objektlokalisierung I.

Das als violett gekennzeichnete und durch  ${}^{2D3D}P_{1,k}^{j,AC}$  sowie  ${}^{2D3D}P_{2,k}^{j,AC}$  definierte Gebiet entspricht der Fläche des gesamten Aktivitätsclusters. Die Fläche der primären Partition wird durch das türkise Rechteck mittels der Punkte  ${}^{3D}P_{3,t}^{j,AC}$  und  ${}^{2D3D}P_{2,k}^{j,PP}$  festgelegt und entspricht der Umrandung der Daten des lokalisierten Objekts. Durch die Fokussierung auf den Vordergrundcluster kann gewährleistet werden, dass das Augenmerk auf das zur Kamera nächste Objekt erhalten bleibt. Das letzte Bild der obigen Szene zeigt den Fall der Erzeugung eines zusätzlichen Aktivitätsclusters aufgrund von Objektabsplattung.

### 4.2.5 ZUSÄTZLICHE BETRACHTUNGEN ZUR OBJEKTLOKALISATION

Prinzipiell bildet die Information der Parametermenge (45) bereits eine adäquate Beschreibung eines lokalisierten Objekts. Mithilfe der intrinsischen Kameraparameter, wie Pixelabmessungen  $s_u$ ,  $s_v$  und Brennweite  $f$  des verwendeten Objektivs und unter Verwendung von (16), können Aussagen über Position und Größe des Objekts im realen Raum getroffen werden.

Dass eine zusätzliche Analyse der zur Verfügung stehenden Daten bzw. Informationen im Sinne einer Plausibilität notwendig ist, wird im Folgenden anhand einiger Beispiele gezeigt.

Abb. 49 spiegelt den Fall der Generation von Aktivitätsclustern mit unzureichender Information wider. Durch Analyse der durch die primären Partition abgedeckten Fläche können Objekte von geringer Größe und mangelndem Informationsgehalt unterdrückt werden. Im Allgemeinen sind diese Objekte bewegungsinduziert und flüchtig. Sie liefern im Sinne einer nachgeschalteten Objektverfolgung keinen nennenswerten Beitrag.

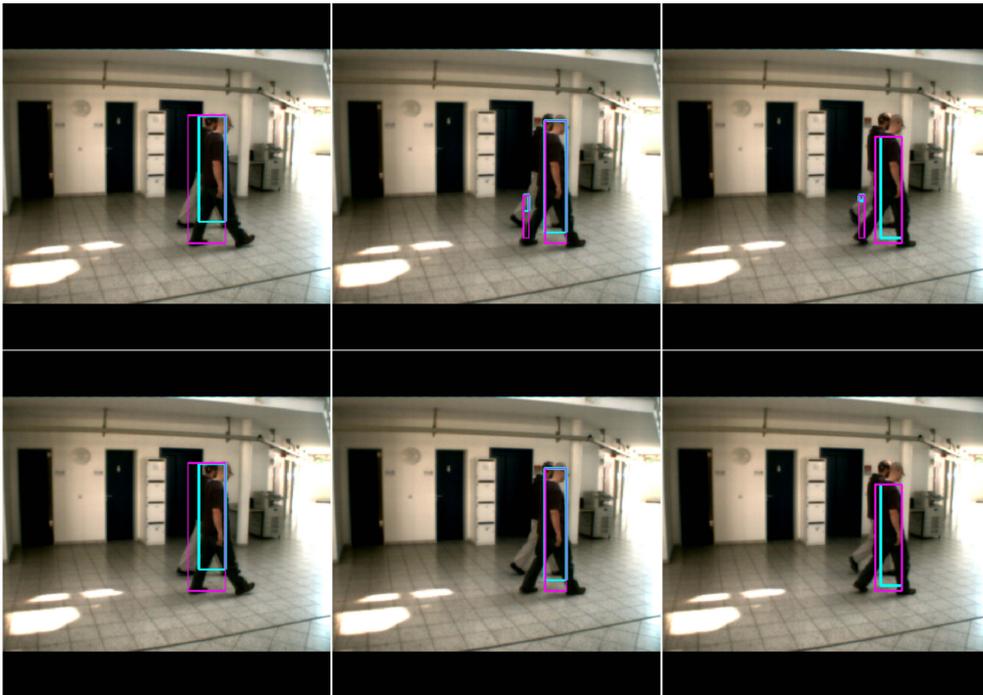


Abbildung 49: Lokalisationsproblematik I.

Oben: Generation von Aktivitätsclustern mit geringem Informationsgehalt. Unten: Unterdrückung von Aktivitätsclustern geringer Information durch Analyse der Größe der primären Partition.

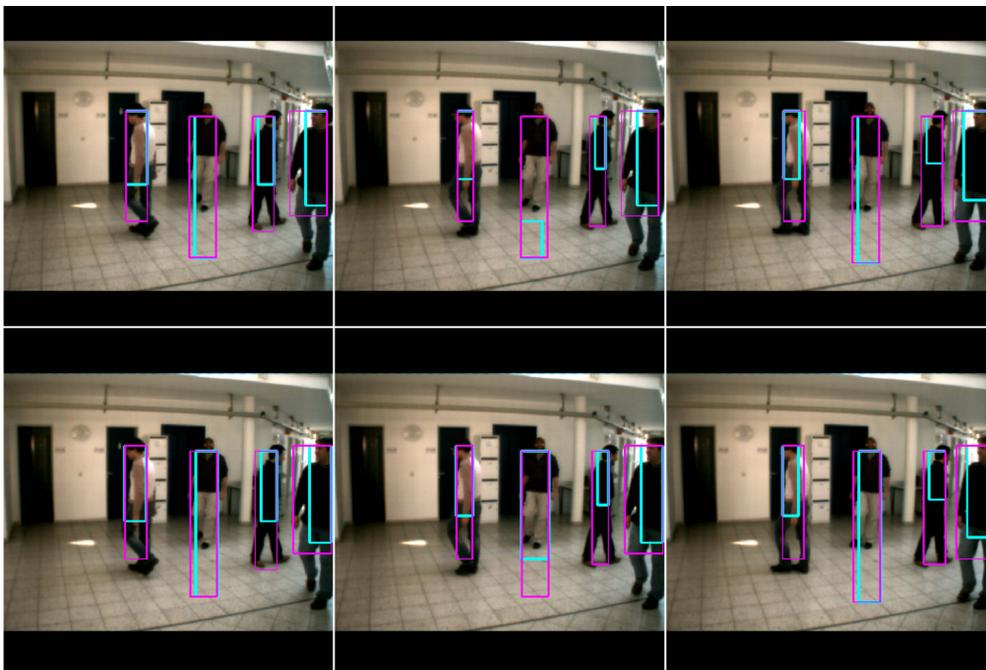


Abbildung 50: Lokalisationsproblematik II.

Oben: Die veränderte Reflexion des modulierten Lichts durch den Untergrund aufgrund von Objektbewegungen führt zu Veränderungen der Hintergrunddaten und hat einen direkten Einfluss auf die Bildung von Aktivitätsclustern. Unten: Korrektur der Fehllokalisierung nach Analyse der Verteilungsgewichte des EM-Modells sowie Positionsanalyse der primären Partition auf dem Aktivitätscluster. Parameter:  $w_{min}=0.25$ .

#### 4 Agentenbasierte Detektion, Lokalisation und Klassifikation von Objekten

Eine weitere Lokalisationsproblematik wird durch Abb. 50 erkennbar. Durch die Rückstreuung des Lichts eines sich bewegenden Objekts bzw. Person ergibt sich unter Umständen, besonders im Falle stark reflektierender Oberflächen, eine Veränderung der Hintergrunddistanzdaten in dem Gebiet vor dem zu lokalisierenden Objekt, und das führt dazu, dass das Zielobjekt nicht mehr in der primären Partition enthalten ist.

Eine Korrektur der Fehllokalisierung wird zum Einen durch die Analyse der Gewichte der Normalverteilungen des EM-Modells und zum Anderen durch eine zusätzliche Betrachtung der Position und Größe der primären Partition im Vergleich zur Fläche des Aktivitätsclusters in der  $u-v$ -Ebene von  $K^{2D3D}$  erreicht.

Unterschreitet das Gewicht der Normalverteilung des Distanzclusters mit geringster Entfernung einen empirisch ermittelten Schwellenwert  $w_{min}$  und kann das Zentrum der primären Partition in einem unplausiblen Bereich<sup>66</sup> des Gebietes des Aktivitätsclusters gefunden werden, wird die anfängliche Annahme bzgl. der Lokalisation verworfen und die Daten des Hintergrundclusters als lokalisiertes Objekt interpretiert.

Eine zusätzliche Problematik zeigt Abb. 51. Die Bewegung der Extremitäten führt dazu, dass der entsprechende Bereich als primäre Partition angesehen wird. Streng genommen ist die Lokalisation nicht fehlgeschlagen, da gemäß dem gewählten Vorgehen die Distanzdaten, welche durch die Bewegung der Extremität generiert werden, einem zulässigen Distanzcluster entsprechen. Im gewählten Ansatz wird der obere Teil eines Objekts, was im Falle einer Person dem Torso entspricht, favorisiert.

Durch eine, wie bereits im vorigen Beispiel beschriebene Betrachtung der Plausibilität, kann eine Umentscheidung hervorgerufen werden. Der Grund für die gewählte Bevorzugung des oberen Teiles ist darin zu sehen, dass die Fokussierung auf den oberen Bereich der Objekte im Falle von Personen, bedingt durch den im Vergleich zu den unteren Extremitäten geringere Bewegungsvariation, eine stabilere Positionsbestimmung ergibt.

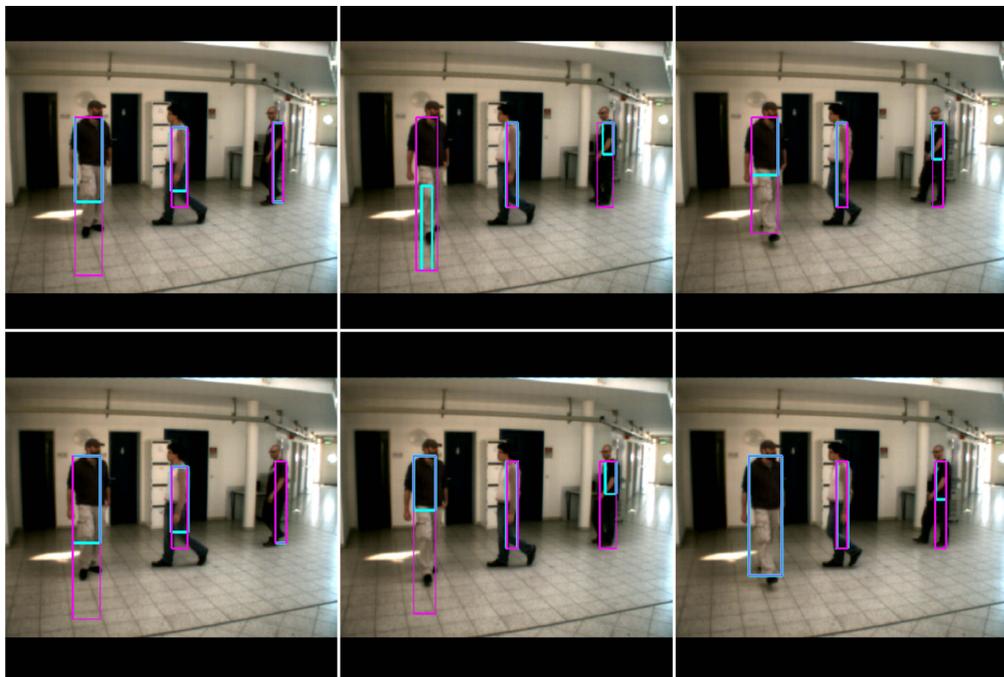


Abbildung 51: Lokalisationsproblematik III.

Oben: Durch Extremitätenbewegung induzierte Fehllokalisierung. Unten: Korrektur durch Größenanalyse der Modellgewichte sowie Positionsanalyse der primären Partition. Parameter:  $w_{min}=0.25$ .

<sup>66</sup>Als unplausibel wird der untere Bereich des Aktivitätsclusters angenommen.

#### 4 Agentenbasierte Detektion, Lokalisation und Klassifikation von Objekten

Ein weiterer Fall, welcher zu einer Fehllokalisierung führt, tritt auf, falls die durch Personen mit unterschiedlicher Entfernung zur Kamera erzeugten Vordergrundinformationen zu einem Aktivitätscluster verschmelzen (Abb. 52, obere Zeile).

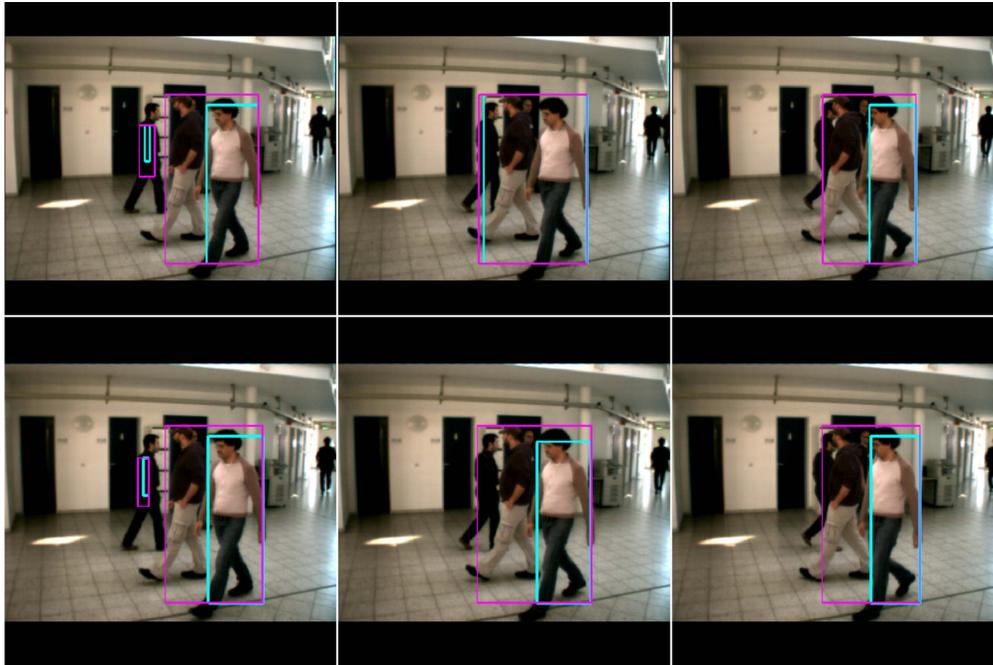


Abbildung 52: Lokalisationsproblematik IV.

Oben: Fehllokalisierung bedingt durch ein großes aufgespanntes Distanzgebiet des Aktivitätsclusters sowie einer zu breiten primären Partition. Unten: Korrektur durch Erhöhung der Clusteranzahl. Parameter:  $\Delta z_{max} = 150\text{cm}$ ,  $\Delta x_{max} = 85\text{cm}$ .

Die Folge davon ist, dass aufgrund des durch die Tiefeninformationen des Aktivitätsclusters  $C_k^j$  aufgespannten Distanzbereichs<sup>67</sup> eine Unterteilung in einen Vorder- und Hintergrundcluster nicht mehr ausreichend ist. Die Wahl von lediglich zwei Clustern führt dazu, dass der durch die primäre Partition abgedeckte Bereich in gewissen Fällen so groß wird, dass mehrere Objekte, welche sich in unterschiedlicher Entfernung zur MultiCam befinden, in die primäre Partition eingebettet werden (Abb. 52, mittleres Bild der oberen Zeile). Als Folge dieser ungewollten Einbettung ergibt sich unter Umständen eine Vergrößerung der primären Partition über Schwellenwerte für die maximal durch eine Person abdeckbare Breite  $\Delta x_{max}$  und Tiefe  $\Delta z_{max}$ <sup>68</sup> hinaus.

Der Ansatz zur Lösung dieser Problematik besteht in der erneuten Modellierung des Hintergrunds durch den EM-Algorithmus unter iterativer Erhöhung der Clusteranzahl, wobei die initialen Werte der Clusterzentren äquidistant aus dem Intervall  $[\mu_{min,t}^j, \mu_{max,t}^j]$  entnommen werden. Abgebrochen wird die iterative Prozedur, falls der Vordergrundcluster eine Breite unterhalb von  $\Delta x_{max}$  vorweisen kann oder der durch die Distanzdaten der primären Partition aufgespannte Distanzbereich den Wert  $\Delta z_{max}$  unterschreitet (Abb. 52, unten).

Den prinzipiellen Ablauf zur Bestimmung der primären Partition liefert Abb. 53.

<sup>67</sup>Definiert wird der aufgespannte Distanzbereich durch den Betrag der Differenz zwischen den Pixeln des Aktivitätsclusters mit der geringsten sowie der größten Entfernung zur MultiCam.

<sup>68</sup>Die Schwellenwerte  $\Delta x_{max}$ ,  $\Delta z_{max} \in \mathbb{R}^+$  entsprechen dabei Größen, welche auf Grundlage von Messungen empirisch ermittelt wurden (Abb. 52).

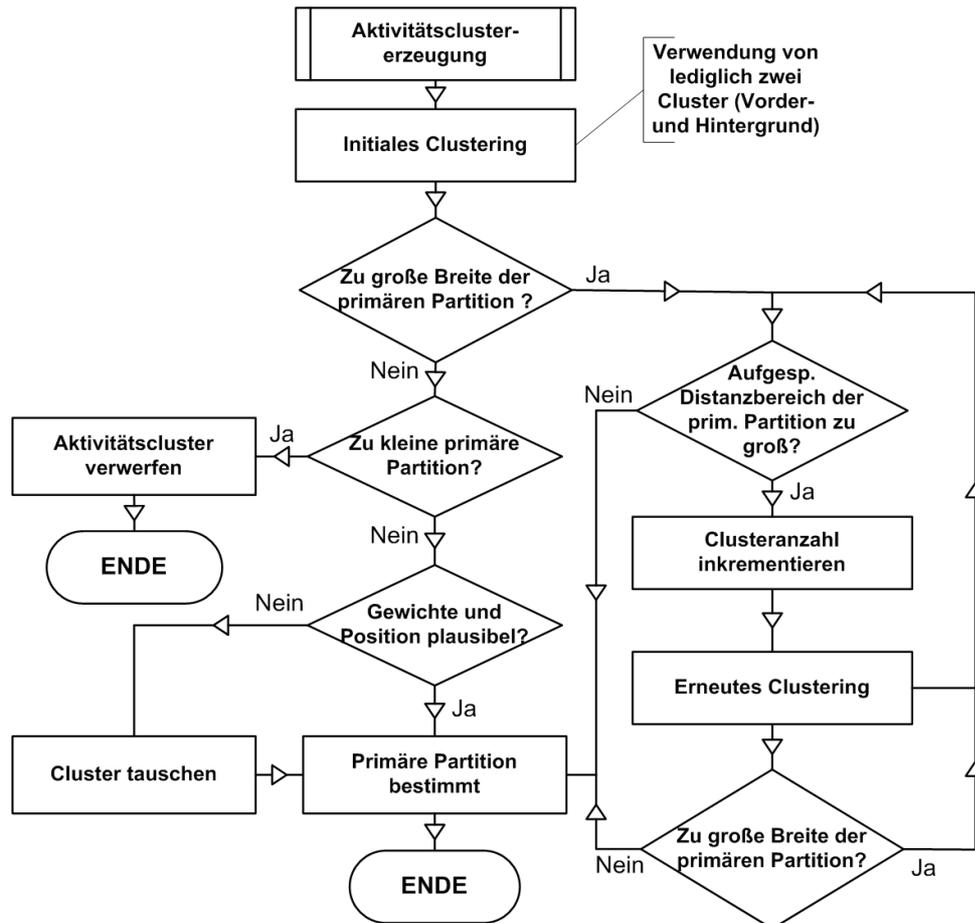


Abbildung 53: Methodik zur Bestimmung der primären Partition.

### 4.3 OBJEKTKLASSIFIKATION

Aufgrund der geringen lateralen Auflösung sowie der extrem starken Variation der Form des Objekts in Abhängigkeit der Entfernung ist eine Klassifikation der Objekte als Personen auf Basis der Tiefeninformationen nur bedingt möglich. Weiterhin bieten die Intensitätsinformationen des PMD-Chips, welche als Grauwertbilder interpretiert werden können, aufgrund der geringen Auflösung eine zu geringe Informationsdichte, um eine intensitätsdatenbasierte, zuverlässige Klassifikation lokalisierter Objekte durchführbar erscheinen zu lassen. Naturgemäß liefert die höhere Informationsdichte des vorhandenen 2D-Chips hierzu die Möglichkeit einer weitaus effektiveren Objektbestimmung.

Der Ansatz zur Klassifikation lokalisierter Objekte (Abschnitt 4.2.4) basiert auf der Verwendung zweier State-of-the-Art-Algorithmen, welche auf den hoch aufgelösten 2D-Intensitätsdaten der MultiCam operieren. Die Wahl des zum Einsatz kommenden Verfahrens wird dabei durch die Entfernungen der Objekte zur MultiCam bestimmt. Während für den Nahbereich eine von Viola und Jones modifizierte Version des AdaBoost-Algorithmus zur Gesichtsdetektion verwendet wird, wird im mittleren und entfernten Bereich das von Dalal 2006 eingeführte Verfahren, basierend auf der Verwendung von Histogrammen orientierter Gradienten (HOG) zur Personendetektion, eingesetzt.

Erläuterungen der Boosting-Methode sind u. a. zu finden in [163]-[167]. Eine Diskussion der Eigenschaften sowie möglicher Optimierungen findet sich in [160]-[162] sowie [168]-[171]. Eine Beschreibung der von Viola und Jones modifizierten Version des Boostings (AdaBoost) liefert [156], [157] und [188]. Der in dieser Arbeit zum Einsatz kommende Algorithmus basiert auf einer

## 4 Agentenbasierte Detektion, Lokalisation und Klassifikation von Objekten

von Lienhart und Maydt leicht modifizierten Version der Viola-Jones Methode, die in [159] diskutiert wird.

Eine ausführliche Beschreibung der Funktionalität des HOG-Algorithmus findet sich in den Publikationen von Dalal [175], [176]. Das zugrunde liegende Verfahren zur Erzeugung der Gradienten wird zusätzlich in [177] diskutiert.

Im Gegensatz zur Objektlokalisierung entsprechen beide verwendeten Methoden einem Top-Down Ansatz. Sämtliche verfügbare Daten werden auf das Vorhandensein eines zu klassifizierenden Objekts untersucht. Aufgrund der Operation auf hoch aufgelösten Intensitätsbildern ist die Bearbeitung jedoch als relativ zeitaufwendig bzw. rechenintensiv einzustufen.

Durch die zusätzliche Verwendung der Informationen des PMD-Sensors im Rahmen einer kooperativen Fusion von 2D- und 3D-Daten bietet sich jedoch die Möglichkeit, den zu untersuchenden Bereich zu verringern und führt zu einer deutlichen Beschleunigung des Klassifikationsprozesses (Abschnitt 4.3.3).

In den folgenden beiden Abschnitten wird zunächst eine Analyse der Performanz beider im Rahmen dieser Arbeit verwendeten Algorithmen auf den von der MultiCam gelieferten Intensitätsdaten ohne die Hinzunahme zusätzlicher Informationen durchgeführt. Die diesbezüglichen Ergebnisse dienen als Vergleichsbasis für den Fall eines durch 3D-basierten Informationen unterstützten Ansatzes, welcher in den Abschnitten 4.3.3.1 und 4.3.3.2 diskutiert wird.

### 4.3.1 OBJEKTKLASSIFIKATION MITTELS ADABOOST

Die Ergebnisse einer mittels AdaBoost durchgeführten Klassifikation unter Verwendung einer Klassifikator-kaskade, welche auf die Erkennung frontale Gesichter trainiert wurde, sind in Abb. 54 zu finden. Neben dem bereits erwähnten hohen Rechenbedarf stellen die durch den Algorithmus durchgeführten Fehlklassifikationen ein zusätzliches Problem dar und sind im Allgemeinen das Ergebnis eines Klassifikatortrainings unter Verwendung einer zu geringen Anzahl von Negativbeispielen.



Abbildung 54: AdaBoost-Gesichtsklassifikation I.

*Gesichts- und Fehlklassifikationen des verwendeten AdaBoost-Algorithmus auf einem vollständigen Intensitätsbild der MultiCam.*

Die zusätzlich in Abb. 55 gezeigten Aufnahmen repräsentieren Situationen, in denen eine richtige Klassifikation nicht durchgeführt werden konnte. Die relativ einfach identifizierbaren Gründe hierfür sind ungenügende Informationen bedingt durch zu geringe Objektgröße, falsche Ausrichtung des Objekts sowie Verdeckung des Objekts bzw. Gesichtes durch Bart und Mütze.

Fehlende Klassifikation aufgrund ungenügender Informationen ließe sich prinzipiell durch die Verwendung höher aufgelöster 2D-Sensoren minimieren. Die Auflösung der MultiCam (640x480 Pixel) stellt hier eine grundsätzliche Einschränkung dar.



Abbildung 55: AdaBoost-Gesichtsklassifikation II.

*Fehlende Klassifikation aufgrund falscher Ausrichtung, zu geringer Objektgröße sowie zusätzlicher Verdeckung. Analog zu Abb. 54 ergeben sich Fehlklassifikationen aufgrund des Trainings des Klassifikators mit einer unzureichenden Anzahl an Negativbeispielen.*

Einer nicht durchführbaren Klassifikation durch eine falsche Ausrichtung kann auf zwei Arten begegnet werden. Die erste Möglichkeit ist die Verwendung einer größeren Anzahl besser trainierter Klassifikatoren, welche jeweils für unterschiedliche Ausrichtungen des Objekts verwendet werden und parallel auf den Daten operieren. Der Nachteil ist jedoch, dass durch die erhöhte Berechnungsdauer eine echtzeitfähige Operation zusätzlich erschwert wird. Der zweite Ansatz liegt in der Verwendung mehrerer Kameras mit unterschiedlichen Positionen und partieller Überlappung des Sichtbereiches. Durch Zuhilfenahme mehrerer, simultaner Ansichten auf das Objekt erhöht sich die Wahrscheinlichkeit, das zu detektierende Objekt in einer für die Klassifikation passenden Ausrichtung vorzufinden.

Um eine Abschätzung im Hinblick auf eine echtzeitfähige Operation erhalten zu können, wurden die notwendigen Bearbeitungszeiträume der verwendeten Implementierung der Viola-Jones-Methode unter Verwendung unterschiedlicher Bildgrößen untersucht.

Begonnen wurde zunächst mit der Bestimmung der zur Detektion des Gesichts notwendigen Bearbeitungsdauer unter Verwendung des kompletten Bildes (Bildnummer  $k=1$ ) des CMOS-Chips. In den jeweils 21 folgenden Schritten wurde die zu analysierende Bildgröße, unter Einsatz einer sich um einen festen Betrag pro Bild verkleinernden *Region of Interest* (ROI), verringert<sup>69</sup> und erneut dem Algorithmus zur Berechnung zugeführt. Verwendet wurden somit 22 Bilder ( $k=1, \dots, 22$ ) mit jeweils unterschiedlicher ROI. Die Ergebnisse diesbezüglich sind in Abb. 56 sowie Tab. 2 enthalten.

Während Abb. 56 die pro Bildnummer notwendige Berechnungsdauer in grafischer Form enthält, findet sich in Tab. 2 zusätzlich die für jede Bildnummer zugehörige Größe der ROI. Gut zu erkennen ist die leicht exponentielle Verringerung der Berechnungsdauer pro Bild bedingt durch eine Verringerung der ROI gemäß Tab. 2. Dieses ist nicht verwunderlich, da sich durch eine Verringerung der ROI eines Bildes naturgemäß eine Reduktion der vom Algorithmus zu betrachtenden Pixelmenge ergibt, was wiederum zu einer beschleunigten Ausführung der Gesichtsdetektion führt.

Weiterhin lässt sich aus den Informationen von Abb. 56 bzw. Tab. 2 entnehmen, dass der AdaBoost-Algorithmus für eine Operation auf einem kompletten Bild des CMOS-Sensors (Bildnummer  $k=1$ ) im Durchschnitt ca. 180 ms<sup>70</sup> benötigt und damit eine Systemoperation unter der anvisierten Grenze von  $t_{max}^{Akt}$  von 50 ms pro Bild verhindert. Erst eine Reduktion der Bildgröße bzw. ROI auf Abmessungen, wie sie ab Bildnummer  $k=14$  in Tab. 2 angegeben sind, verringert die notwendige Bearbeitungszeit auf ein verwendbares Ausmaß.

<sup>69</sup>Die Verringerung der Bildweite betrug pro Schritt 25 Pixel. Die Bildhöhe wurde pro Schritt um 17,5 Pixel verringert. Bei der Reduktion wurde sichergestellt, dass sich das zu detektierende Objekt weiterhin in der ROI befand.

<sup>70</sup>Um etwaige Schwankungen zu minimieren, wurde zur Bestimmung der Bearbeitungsdauer pro Bild jeweils der Mittelwert von 100 Berechnungen benutzt.

## 4 Agentenbasierte Detektion, Lokalisation und Klassifikation von Objekten

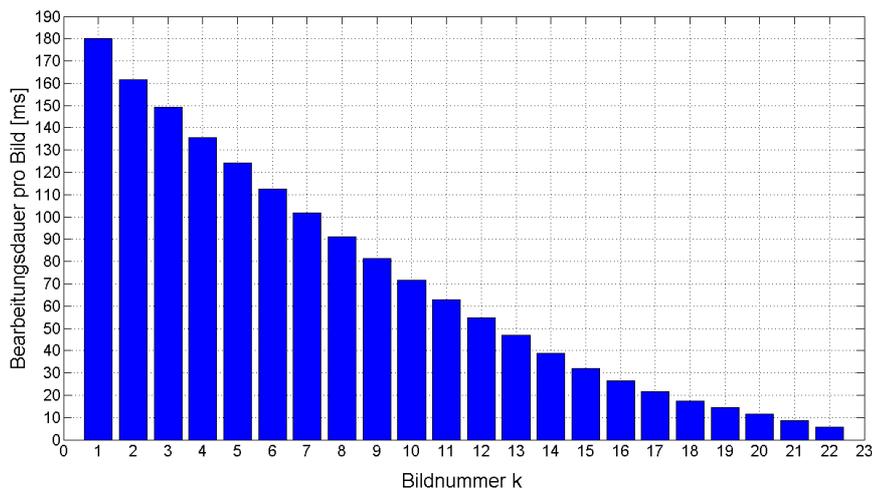


Abbildung 56: Bearbeitungsdauer der Gesichtserkennung pro Bild unter Verwendung unterschiedlicher ROI (AdaBoost).

Die der jeweiligen Bildnummer ( $k=1, \dots, 22$ ) zugeordnete ROI findet sich in Tab. 2.

Tabelle 2: Bearbeitungsdauer pro Bild in Bezug auf verwendete ROI (AdaBoost).

Bildnummer k	ROI-Breite [Pixel]	ROI-Höhe [Pixel]	Berechnungsdauer [ms]	Bildnummer k	ROI-Breite [Pixel]	ROI-Höhe [Pixel]	Berechnungsdauer [ms]
1	640	480	180.2	12	365	287	54.7
2	615	462	161.7	13	340	270	47.0
3	590	445	149.1	14	315	252	38.9
4	565	427	135.6	15	290	235	31.9
5	540	410	124.3	16	265	217	26.4
6	515	392	112.6	17	240	200	21.7
7	490	375	101.7	18	215	182	17.4
8	465	357	91.2	19	190	165	14.4
9	440	340	81.3	20	165	147	11.4
10	415	322	71.6	21	140	130	8.5
11	390	305	62.7	22	115	112	5.6

### 4.3.2 OBJEKTKLASSIFIKATION MITTELS HOG

Der HOG-Klassifikator benötigt zur verlässlichen Klassifikation eine bis auf leichte Verdeckungen vollständige Ansicht des zu detektierenden Objekts. Im Falle der Personendetektion bedeutet dieses, dass Objekte, die sich in der Nähe zur MultiCam befinden, nicht klassifiziert werden können (Abb. 57).



Abbildung 57: HOG-Personenklassifikation I.

*Bedingt durch die zu geringe Distanz zur MultiCam findet lediglich eine partielle Abbildung der Person statt und mündet in einer nicht oder nur fehlerhaft durchgeführten Klassifikation.*

Erst eine komplette Abbildung des zu klassifizierenden Objekts ergibt eine zuverlässige Beurteilung durch den Algorithmus. Klassifikationsbeispiele für den Fall mehrerer Personen finden sich in Abb. 58. Einer der herausragenden Vorteile dieses Verfahrens ist, dass die Personenerkennung für eine Vielzahl unterschiedlicher Ansichten auf eine Person funktioniert und somit die Hinzunahme zusätzlicher Klassifikatoren unnötig werden lässt.



Abbildung 58: HOG-Personenklassifikation II.

*Im Falle einer Personenerkennung durch den HOG-Algorithmus liefert dieser ein Rechteck zurück, welches das klassifizierte Objekt zentriert enthält.*

Zwecks Abschätzung der zur Klassifikation notwendigen Berechnungsdauer durch den HOG-Algorithmus wurden zunächst die notwendigen Bearbeitungszeiträume unter Verwendung verschiedener ROI untersucht (Abb. 59 sowie Tab. 3). Die Vorgehensweise entspricht dabei der aus Abschnitt 4.3.1<sup>71</sup>. Wie auch im Falle von AdaBoost führt eine Verringerung der Pixelmenge durch Verkleinerung der ROI zu einer direkten Beschleunigung der Personendetektion. Während der in Abb. 59 enthaltene Balkengraph die pro Bild notwendige Berechnungsdauer durch den HOG-Algorithmus grafisch wiedergibt, findet sich in Tab. 3, anlog zum AdaBoost-Algorithmus, die Information bezüglich der pro Bild verwendeten ROI.

Die zur Klassifikation auf einem vollständigen CMOS-Bild (Bildnummer  $k=1$ ) notwendige Bearbeitungszeit von ca. 690ms macht deutlich, dass der Einsatz des HOG-Algorithmus eine drastische Reduktion der erreichbaren Zeitauflösung zur Folge hat. Selbst eine Reduktion auf eine Bildgröße von ca. 320x240 Pixel führt immer noch zu Bearbeitungszeiten weit oberhalb von  $t_{max}^{Akq}$ . Erst eine Bildgröße bzw. ROI wie ab Bildnummer  $k=21$  in Tab. 3 angegeben, reduziert die

<sup>71</sup>Wie auch im Falle von AdaBoost, wurde jeweils der Mittelwert aus 100 Berechnungen verwendet.

#### 4 Agentenbasierte Detektion, Lokalisation und Klassifikation von Objekten

notwendige Bearbeitungszeit auf ein akzeptables Maß und erlaubt eine Systemoperation mit hinreichender Zeitauflösung.

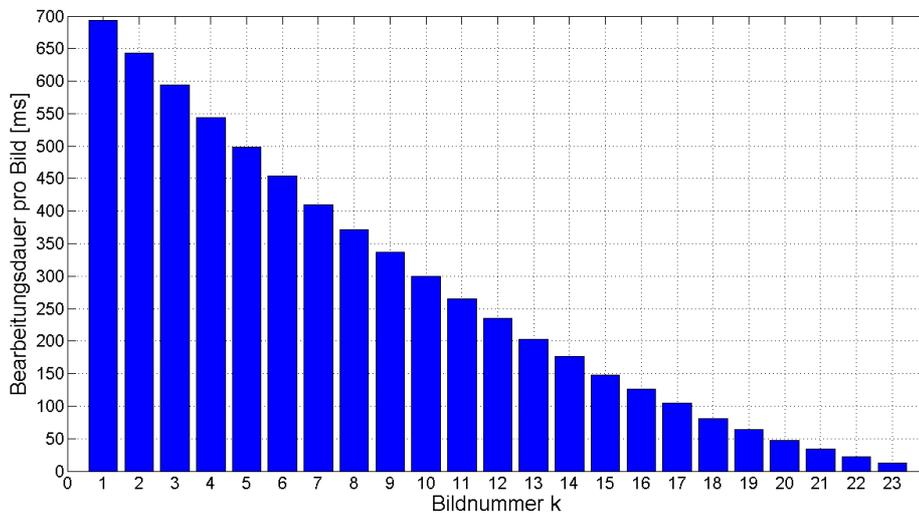


Abbildung 59: Bearbeitungsdauer der Personenerkennung pro Bild unter Verwendung unterschiedlicher ROI (HOG).

Die der jeweiligen Bildnummer ( $k=1, \dots, 23$ ) zugeordnete ROI findet sich in Tab. 3.

Tabelle 3: Bearbeitungsdauer pro Bild in Bezug auf verwendete ROI (HOG).

Bildnummer k	ROI-Breite [Pixel]	ROI-Höhe [Pixel]	Berechnungsdauer [ms]	Bildnummer k	ROI-Breite [Pixel]	ROI-Höhe [Pixel]	Berechnungsdauer [ms]
1	640	480	694.0	12	371	316	235.6
2	615	465	643.5	13	345	300	202.3
3	591	450	593.8	14	321	286	176.3
4	566	435	544.1	15	296	270	147.5
5	542	420	497.9	16	272	256	126.0
6	518	405	453.6	17	248	241	104.4
7	492	390	409.5	18	222	225	81.0
8	468	375	371.6	19	198	211	64.0
9	444	361	337.2	20	174	196	46.7
10	419	345	299.4	21	149	181	34.6
11	395	330	264.7	22	125	166	22.2
				23	101	121	12.4

Die Forderung des HOG-Algorithmus, dass das zu klassifizierende Objekt vollständig in den Bilddaten vorhanden sein muss, impliziert, dass eine Personenklassifikation mittels dieser

## 4 Agentenbasierte Detektion, Lokalisation und Klassifikation von Objekten

Methode lediglich für Objekte ab einer minimalen Distanz anwendbar ist<sup>72</sup>. In Abb. 60 sind die Detektionsleistungen beider Verfahren in Abhängigkeit der Objektdistanz abgebildet. Erkennbar ist, dass der HOG-Algorithmus erst ab einer Entfernung von ca. 3.5 m als zuverlässig klassifizierend einzuschätzen ist. Erkennbar ist weiterhin, dass die Performanz der HOG-Methode zu größeren Entfernungen stark abnimmt. Ein Umstand, der nicht verwundert, da das Verfahren prinzipiell auf der Analyse von Bildgradienten basiert und die relativ geringe Auflösung des 2D-Sensors ab einer gewissen Distanz nicht mehr genügend Informationen für eine verlässliche Klassifikation liefert.

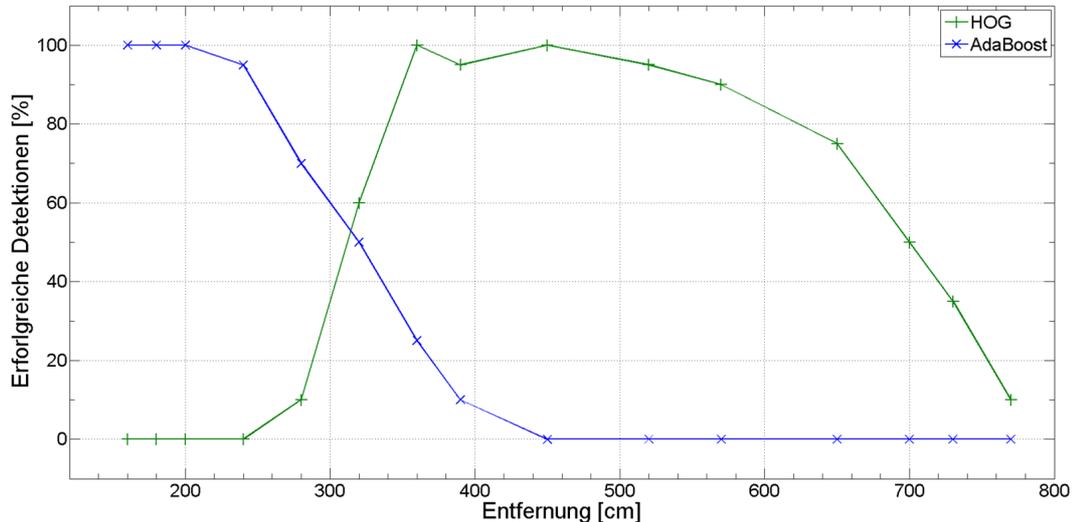


Abbildung 60: Detektionsleistung in Abhängigkeit der Objektentfernung.

Für den AdaBoost-Algorithmus gilt, dass er zu einer zuverlässigen Gesichtsklassifikation eine gewisse Nähe der Objekte zur MultiCam benötigt. Der drastische Abfall der Klassifikationsleistung bereits ab ca. 3 m ist vor allem durch die geringe Datenmenge bedingt. Die VGA-Auflösung der MultiCam bietet nicht genug Informationen, um eine zuverlässige Gesichtserkennung ab dem mittleren Distanzbereich der MultiCam durchführen zu können.

Der im Rahmen dieser Arbeit verfolgte Ansatz zur Personenerkennung basiert auf der Verwendung zweier unterschiedlicher Klassifikationsmethoden, deren Auswahl auf Grundlage der Entfernung des Objekts bestimmt wird. Während für den Bereich zwischen 3.5 m und 7.5 m der HOG-Algorithmus Verwendung findet, wird in dem Nahbereich zwischen 1.5 m und 2.5 m das AdaBoost-Verfahren zur Gesichtserkennung eingesetzt. Der Bereich zwischen 2.5 m und 3.5 m sowie 7.5 m und 9 m bleibt ungedeckt, was unter der Annahme mehrerer simultan operierender und unterschiedlich positionierter MultiCams jedoch keine gravierende Einschränkung bedeutet.

### 4.3.3 PERSONENERKENNUNG UNTER VERWENDUNG KOOPERIERENDER FUSION

Grundsätzlich bieten sich zwei Methoden zur Beschleunigung der Objektklassifikation an. Zum Einen durch die Erhöhung der Rechenleistung unter Ausnutzung eventuell verfügbarer Parallelisierungen (Multikern-CPU, Verlagerung der Berechnung auf Grafikkarte) zum Anderen die Reduktion der Datenmenge unter Verwendung von Vorwissen.

Durch die Verwendung der Informationen basierend auf den Aktivitätscluster-Parameterungen (45) kann der zu betrachtende Bildbereich von vornherein reduziert werden, was naturgemäß in einer beschleunigten Ausführung des Klassifikationsalgorithmus resultiert [179].

<sup>72</sup>Um eine erfolgreiche Klassifikation durchzuführen, benötigt die im Rahmen dieser Arbeit gewählte Implementierung des HOG-Algorithmus neben der vollen Präsenz des Objektes im Bild zusätzlich einen das Objekt vollständig umgebenden Bereich (Abb. 64, Abb. 65).

## 4 Agentenbasierte Detektion, Lokalisation und Klassifikation von Objekten

### 4.3.3.1 UNTERSTÜTZTER ADABOOST-ALGORITHMUS (A-ADABOOST)

Aufgrund der notwendigen Objektgröße findet die Gesichtsdetektion im Nahbereich der MultiCam statt (Abb. 60). In diesem Distanzbereich kann davon ausgegangen werden, dass die Distanzdaten des Aktivitätsclusters sich durch eine entsprechend hohe Modulationsamplitude auszeichnen. Dieses führt dazu, dass Objekte als Ganzes zuverlässig lokalisiert werden. Basierend auf den Informationen der Parametermenge  ${}^{2D3D}c_i^j$  (45) eines Aktivitätsclusters lautet die Zielsetzung, ein passendes, in der Größe reduziertes Suchfenster zu erzeugen, welches das Zielobjekt enthält.

Sei zunächst  $R$  das den kompletten Aktivitätscluster einhüllende Rechteck,  $\tilde{R}$  das entsprechende Rechteck für die Elemente der primären Partition (Abb. 48). Eine Generation eines Suchfensters basiert dann auf folgenden Annahmen:

1. Das Zielobjekt befindet sich in  $R$  horizontal zentriert.
2. Das Zielobjekt befindet sich in einem oberen Abschnitt von  $R$ .

Wie bereits angesprochen, kann davon ausgegangen werden, dass in dem betrachteten Distanzbereich alle objektbezogene Distanzwerte (somit alle Elemente des Aktivitätsclusters) in Betracht gezogen werden können. Eine Beachtung lediglich der Elemente der primären Partition führt aufgrund des vorausgegangenen Clusterings und der damit verbundenen Reduzierung der Objekt beschreibenden Informationen sowie einer nichtdeterministischen Größe<sup>73</sup> von  $\tilde{R}$  zu einer unzuverlässigen Bestimmung des Suchfensters und ist der Grund, warum nicht auf die Parameter von  $\tilde{R}$  zurückgegriffen wird. Unter Verwendung der Tiefeninformation des Punktes  ${}^{2D3D}P_i^j$  aus der zugehörigen Parametermenge  ${}^{2D3D}c_i^j$  (45) und Verwendung plausibler Werte bezüglich der Objektgröße<sup>74</sup> (Kopf) wird zunächst mittels (16) die Breite des Suchfensters berechnet. Der horizontale Mittelpunkt des Suchfensters wird aufgrund der ersten Annahme durch  $R$  bestimmt.

Für die Bestimmung der Höhe des Suchfensters wird der Quotient der Tiefenpositionen der beiden Vorder- und Hintergrundcluster berechnet, wobei vorausgesetzt wird, dass die Distanzwerte der Clusterzentren in aufsteigender Reihenfolge geordnet vorliegen. Unterschreitet der Quotient einen empirisch ermittelten Schwellenwert, ist das ein Hinweis darauf, dass die einzelnen Clusterzentren eine gewisse Distanz besitzen, was wiederum darauf schließen lässt, dass Hintergrundpixel in der Elementemenge des Aktivitätsclusters vorhanden sind (Abb. 61). In diesem Falle wird die Höhe des Suchfensters gleich der Höhe von  $R$  gesetzt. Die Position der oberen Kante des Suchfensters stimmt dabei mit der oberen Kante von  $R$  überein.

Im Falle nahe beieinanderliegender Clusterzentren gilt die Annahme, dass lediglich die Distanzdaten des Objekts in  $C_i^j$  vorhanden sind und durch  $R$  die korrekte Umrandung des Zielobjekts stattgefunden hat. Die Höhe des Suchfensters wird, wie auch im Falle der Breite, auf eine plausible Größe gesetzt, wobei als obere Begrenzung des Suchfensters eine um einen geringen Faktor reduzierte Höhe der obersten Kante von  $R$  verwendet wird (Abb. 61).

Das Vorhandensein von ungewollten Hintergrundinformationen ist durch die Nähe des Objekts zur MultiCam bedingt. Die relativ hohe Intensität des reflektierten Lichts führt aufgrund von Multipfad-Propagationen (Abschnitt 2.2) zu einer Beeinflussung der Hintergrunddaten. Erkennbar wird dieser Umstand in der mittleren Spalte von Abb. 61. In gewissen Positionen zur Beleuchtung wechselwirkt das reflektierte Licht mit den Reflexionen des Hintergrunds derart, dass Teile davon fälschlicherweise als gültige Distanzwerte erscheinen.

<sup>73</sup>Aufgrund der Fluktuation der Distanzdaten des PMD-Sensors erzeugt der EM-Algorithmus grundsätzlich eine variierende Größe der primären Partition. Entsprechend wird auch das die primäre Partition einhüllende Rechteck in seiner Größe schwanken. Dessen Abmessung stellt somit keine zuverlässige Ausgangsposition dar, um ein in der Größe reduziertes Suchfenster zu erzeugen.

<sup>74</sup>Als plausible Größe ist hierbei eine Fläche von 45cm x 45cm angesetzt worden. Eine Umrechnung in Pixel ergibt sich durch (16).

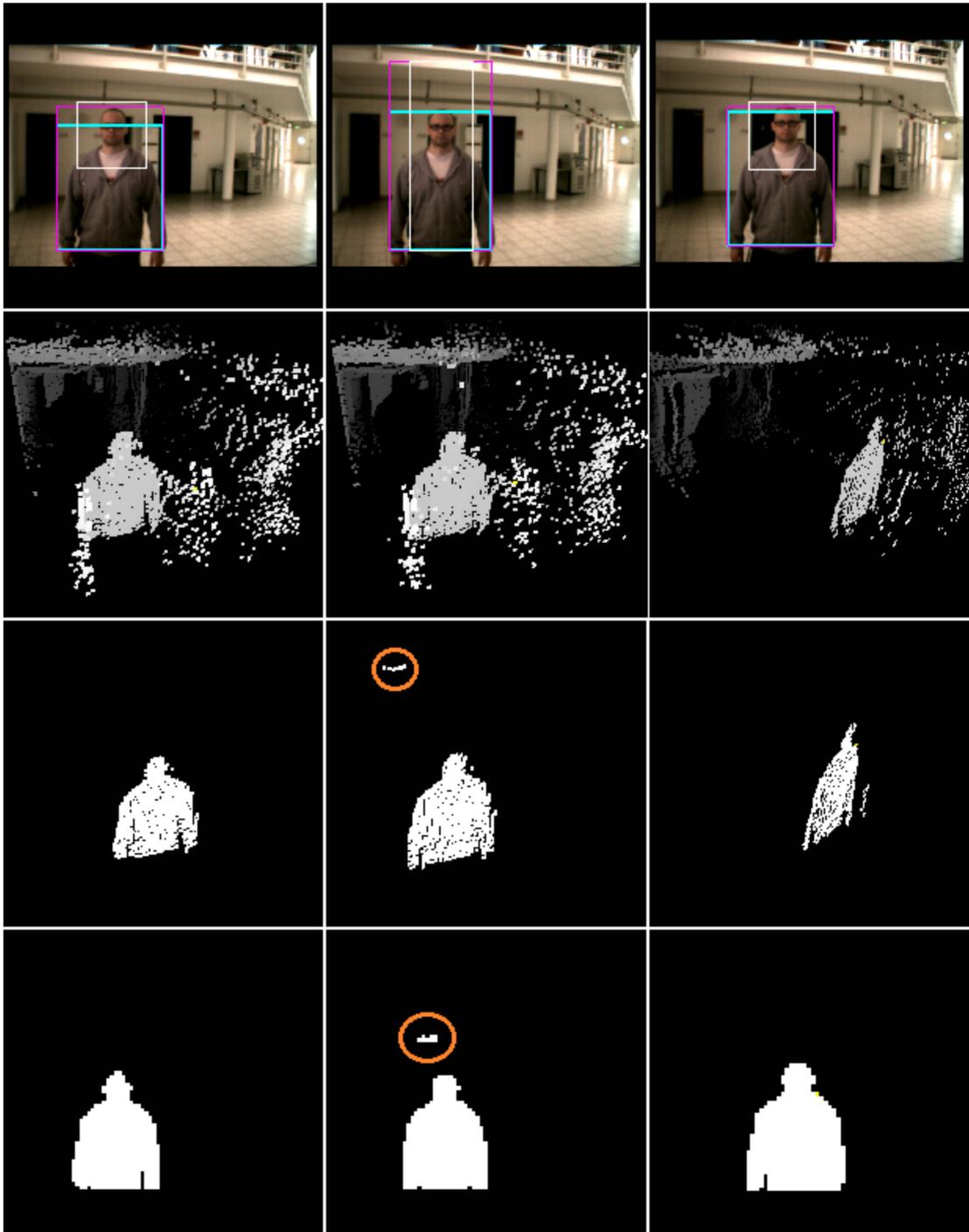


Abbildung 61: Gesichtsdetektion I.

1. Zeile: Ergebnisse der vorgeschalteten Lokalisationsstufe, wobei das violette Rechteck der Einhüllenden des kompletten Aktivitätsclusters entspricht, das türkise Rechteck der Umrandung der primären Partition. Das weiße Rechteck repräsentiert das erzeugte Suchfenster. 2. Zeile: PMD-Basisdaten nach durchgeführter Präprozessierung in perspektivischer Projektion. 3. Zeile: Elemente des Aktivitätsclusters nach abgeschlossener Lokalisation in orthogonaler Projektion. Mittels des in der Parametermenge vorhandenen Vorwissens findet eine Reduktion der Datenmenge statt. 4. Zeile: Daten analog zu obiger Zeile, jedoch mit veränderter Sichtrichtung frontal auf das Objekt. Zu erkennen sind in der mittleren Spalte die Effekte der Multipropagation des Modulationslichts im Nahbereich der MultiCam, welche zu veränderten Hintergrunddaten führen und in dem Aktivitätscluster zusätzlich auftauchen.

#### 4 Agentenbasierte Detektion, Lokalisation und Klassifikation von Objekten

Ausgewählte Aufnahmen aus einer aus 50 aufeinanderfolgenden Bildern bestehenden Sequenz liefert Abb. 62. Darin ersichtlich sind die jeweils zum entsprechenden Zeitpunkt erzeugten Fenster, welche dem unterstützten Algorithmus als zu benutzendes Suchgebiet als Parameter zur Verfügung gestellt werden. Die Anzahl der eventuell nicht erfolgreich durchgeführten Gesichtserkennung, aufgrund eines fehlerhaften Suchgebietes, können prinzipiell durch eine Vergrößerung Letzteres minimiert werden. Es stellt sich aber die Frage, inwieweit dieses sinnvoll ist, da eine Hochskalierung des Suchgebietes sowohl eine direkte Auswirkung auf die Berechnungsdauer als auch auf die Anzahl der durchgeführten Fehlklassifikationen hat. Im Allgemeinen ist eine Abwägung zwischen Klassifikationsleistung und Berechnungsdauer in Abhängigkeit der zugrunde liegenden Anwendung notwendig.

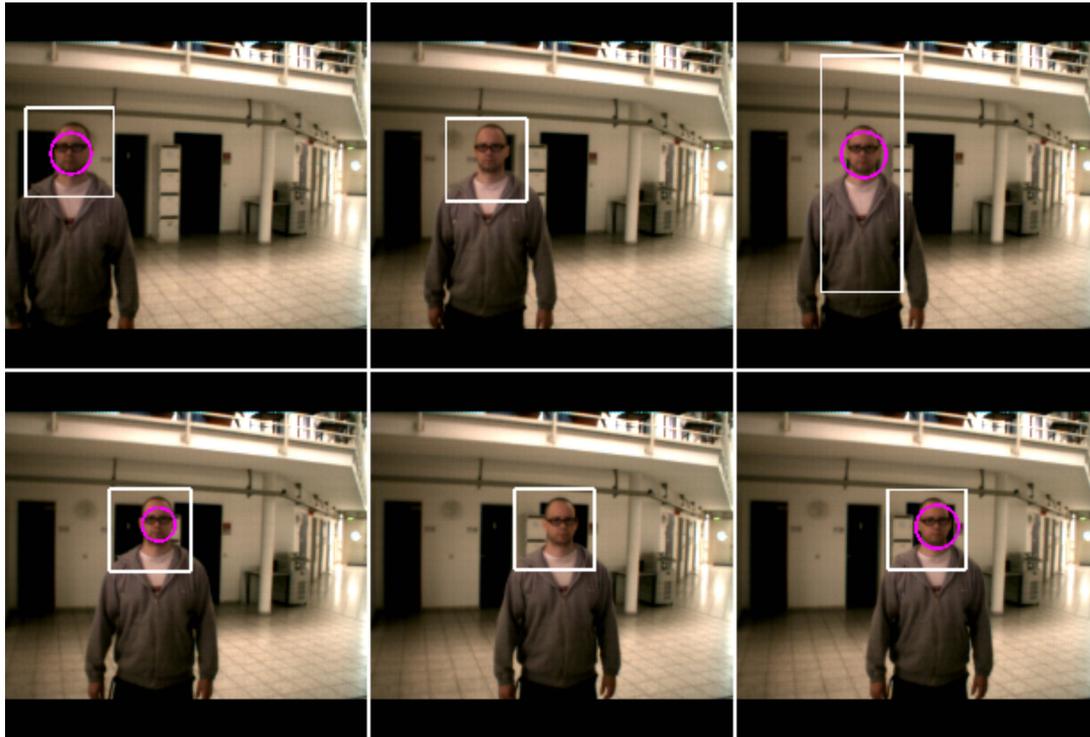


Abbildung 62: Klassifikationsbeispiele A-AdaBoost.

*Während die erste nicht durchgeführte Klassifikation (mittleres Bild, oben) durch ein fehlerhaft platziertes Suchgebiet bedingt ist, liegt der Grund für eine nicht durchgeführte Erkennung im zweiten Fall (mittleres Bild, unten) in dem Versagen des Klassifikators an sich (die unveränderte Methode lieferte hier ebenfalls kein Klassifikationsergebnis).*

Eine Analyse jedes einzelnen Bildes der Videosequenz enthält Abb. 63. Der obere Graph zeigt die durch die Verringerung der Datenmenge erreichte Beschleunigung der Bearbeitungsdauer pro Bild. Während eine Gesichtserkennung mit einem kompletten Intensitätsbild der MultiCam ca. 175 ms in Anspruch nimmt, wird durch die Reduktion des Suchfensters mittels kooperativer Datenfusion eine Beschleunigung um den Faktor 7-10 erreicht.

Der mittlere Graph von Abb. 63 enthält die Ergebnisse der Lokalisation sowie Klassifikation des in Abb. 62 abgebildeten Objekts. Um einer Fehlinterpretation der in Abb. 63 enthaltenen Informationen vorzubeugen, soll an dieser Stelle noch einmal bekräftigt werden, dass im Rahmen dieser Arbeit eine valide Objektlokalisierung als eine erfolgreiche, allein auf PMD-Daten basierte Erzeugung eines Aktivitätsclusters angesehen wird (Abschnitt 4.2.4.2). Eine gültige Objektklassifikation wird hingegen als eine erfolgreich durchgeführte Detektion eines Gesichts bzw. Körpers durch die lediglich auf den 2D-Daten operierenden Algorithmen interpretiert.

#### 4 Agentenbasierte Detektion, Lokalisation und Klassifikation von Objekten

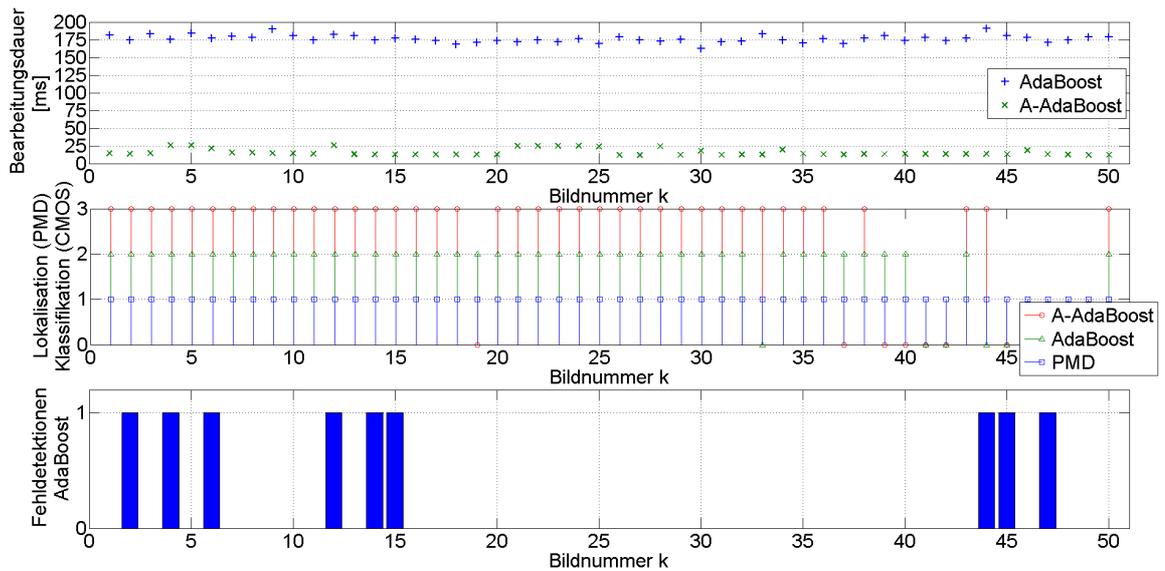


Abbildung 63: Berechnungsdauer, Lokalisation und Klassifikation, Fehldetektionen (AdaBoost).

Ergebnisse der Detektion und Klassifikation auf Basis der in Abb. 62 angegebenen Klassifikationssequenz. Oben: Berechnungsdauer des ursprünglichen und des unterstützten Algorithmus. Mitte: Lokalisations- und Klassifikationsleistung der entsprechenden Prozesse. Jede erfolgreich durchgeführte Lokalisation bzw. Klassifikation des Objekts ist durch einen entsprechenden Wert größer 0 gekennzeichnet (1 für Lokalisation, 2 und 3 für Klassifikation). Zu beachten ist, dass die Informationen in diesem Graphen separat zu betrachten sind. Die Zusammenlegung der Daten wurde aus Übersichtlichkeitsgründen gewählt. Unten: Bilder, in denen durch den ursprünglichen Algorithmus zusätzliche Fehldetektionen von Gesichtern auftraten, sind durch von 0 verschiedene Werte markiert.

Die blauen Punkte oberhalb der 0 im mittleren Graphen von Abb. 63 markieren eine erfolgreich durchgeführte Lokalisation mithilfe der Daten des PMD-Chips, grün gibt eine erfolgreiche Gesichtsdetektion durch den originalen Algorithmus (Klassifikation eines Objekts als Person), basierend auf den kompletten Intensitätsdaten, wider und rot repräsentiert die Klassifikationsleistung des unterstützten Algorithmus durch Angabe eines reduzierten Suchgebietes. Man erkennt, dass für jedes Bild aus der Sequenz eine Objektlokalisierung erfolgreich durchgeführt werden konnte.

Im Vergleich dazu liefern die beiden intensitätsbasierten Algorithmen zur Gesichtserkennung eine verminderte Performanz. Bei Betrachtung der Klassifikationsrate beider intensitätsbasierten Algorithmen fällt auf, dass der unterstützte Algorithmus (A-AdaBoost) leicht abfällt. Begründet wird dieses dadurch, dass in gewissen Aufnahmen ein schlecht platziertes Suchgebiet erzeugt wird, welches dem A-AdaBoost-Algorithmus als Vorinformation zugetragen wird. Dem kann im Prinzip durch eine Vergrößerung des Suchgebietes begegnet werden, was aber wiederum einen direkten negativen Einfluss auf die zur Klassifikation notwendige Berechnungsdauer zur Folge hat (Abb. 56, sowie Tab. 2).

Die in dem mittleren Graphen von Abb. 63 enthaltenen Daten bestätigen die Erfahrung, dass mithilfe der Informationen des 3D-Sensors zu jedem Bild<sup>75</sup> Distanzinformationen erzeugt werden können, die eine zuverlässige Lokalisation pro Aufnahme erlauben. Dieses unterstreicht die Zuverlässigkeit der distanzbasierten Informationen im Vergleich zu dem Wissen, welches auf Grundlage der intensitätsbasierten Algorithmen erzeugt wird.

Ein zusätzlicher Vorteil, welcher sich durch die Hinzunahme von 3D-basiertem Vorwissen ergibt, ist, dass durch die Reduktion des durch den Klassifikator zu untersuchenden Gebietes gleichzeitig eine Verringerung der Fehlklassifikationen erreicht wird. Ersichtlich wird dieses in dem untersten

<sup>75</sup>Gemeint ist hierbei der Nahbereich der Kamera bis 2.5m.

## 4 Agentenbasierte Detektion, Lokalisation und Klassifikation von Objekten

Graphen aus Abb. 63. Werte, verschieden von 0, entsprechen Bildindizes, in denen durch den ursprünglichen Klassifikator eine oder mehrere Fehlklassifikationen durchgeführt worden sind. Im Vergleich dazu hat der unterstützte Algorithmus in der betrachteten Sequenz keine Einzige vorzuweisen.

### 4.3.3.2 UNTERSTÜTZTER HOG-ALGORITHMUS (A-HOG)

Das Vorgehen im Falle des unterstützten HOG-Algorithmus basiert auf den Informationen des Punktes  ${}^{2D3D}P_k^j$  aus der Parametermenge (45) des Aktivitätsclusters. Auf Grundlage der Tiefeninformation des Punktes sowie seiner Position in der  $u-v$ -Ebene von  $K^{2D3D}$  wird, unter Verwendung von plausiblen Personengrößen, ein entsprechendes Suchrechteck erzeugt, welches dem HOG-Algorithmus als Vorwissen zur Verfügung gestellt wird. Erkennbar ist (Abb. 64), dass die berechneten Suchfenster die Oberfläche der zu klassifizierenden Person wesentlich überschreitet. Der Grund für das vergrößerte Suchgebiet liegt in der Arbeitsweise des HOG-Algorithmus, welcher zur Klassifikation eines Objekts eine im Vergleich zum Objekt größere Fläche benötigt. Die zur Klassifikation von Personen notwendigen Größenordnungen des Suchfensters sind im Vorfeld durch eine Analyse der durch den Algorithmus im Falle einer durchgeführten Klassifizierung zurückgelieferten, den Zielobjekten umrandenden Rechtecken, bestimmt worden (Abb. 58).

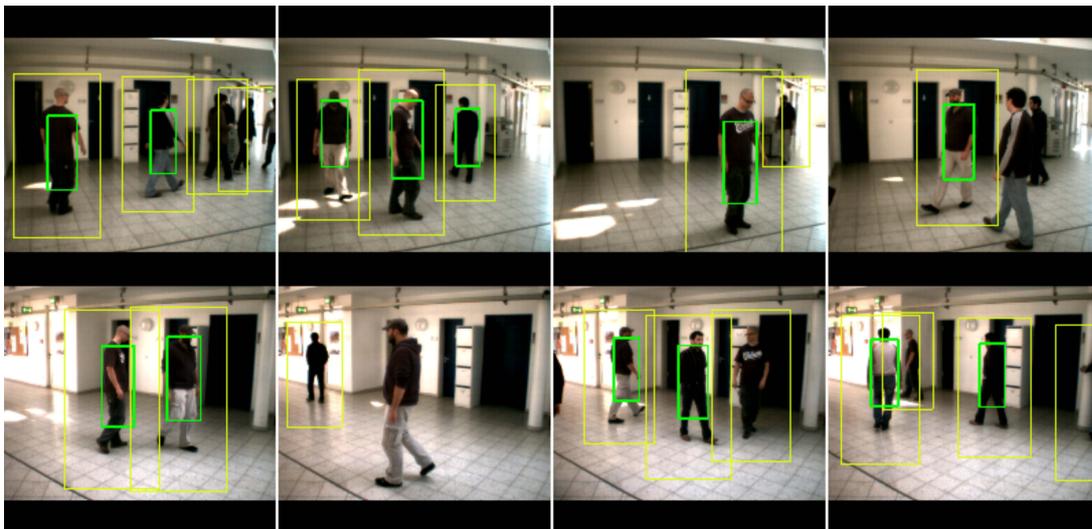


Abbildung 64: Klassifikationen des A-HOG-Algorithmus.

*Während die gelben Rechtecke den reduzierten Suchfenstern entsprechen, indizieren die grünen Rechtecke eine erfolgreiche Klassifikation des Objekts als Person. Zwecks Fokussierung auf das Objekt ist das vom HOG-Algorithmus zurückgelieferte Gebiet verkleinert worden. Die fehlenden Suchfenster für Personen in dem vierten Bild der oberen Zeile sowie dem zweiten Bild der unteren Zeile sind bedingt durch eine zu geringe Nähe zur Kamera. Aufgrund der Verletzung der Forderung nach einer minimalen Distanz zum Sensor, befinden sich die Objekte in einem Gebiet, für welches der HOG-Algorithmus nicht zuständig ist.*

Untersuchungen im Vorfeld der Implementierung haben ergeben, dass eine zuverlässige Klassifikation durch den HOG-Algorithmus ebenfalls auf 2D-Bilder, mit im Vergleich zu den Intensitätsbildern der MultiCam verringerten Auflösung, möglich ist. Durch eine Reduktion der von der MultiCam gelieferten 2D-Informationen auf eine Auflösung von 320x240 Pixel konnte, bei vergleichbarer Klassifikationsleistung, eine drastische Reduktion der zur Berechnung notwendigen Zeiträume von ca. 690 ms auf ca. 120 ms pro Intensitätsbild erreicht werden (Tab. 3). Die Möglichkeit einer bereits im Vorfeld durchführbaren Reduktion der Information ohne Verlust an

#### 4 Agentenbasierte Detektion, Lokalisation und Klassifikation von Objekten

Klassifikationsleistung ist im Rahmen dieser Arbeit als Ausgangspunkt zur zusätzlichen Beschleunigung des HOG-Verfahrens verwendet worden.

Wie auch im Falle der Klassifikation im Nahbereich der Kamera mittels Gesichtserkennung (Abschnitt 4.3.3.1) wurde zur Untersuchung ein aus 50 Bildern bestehendes Video verwendet (Abb. 65).

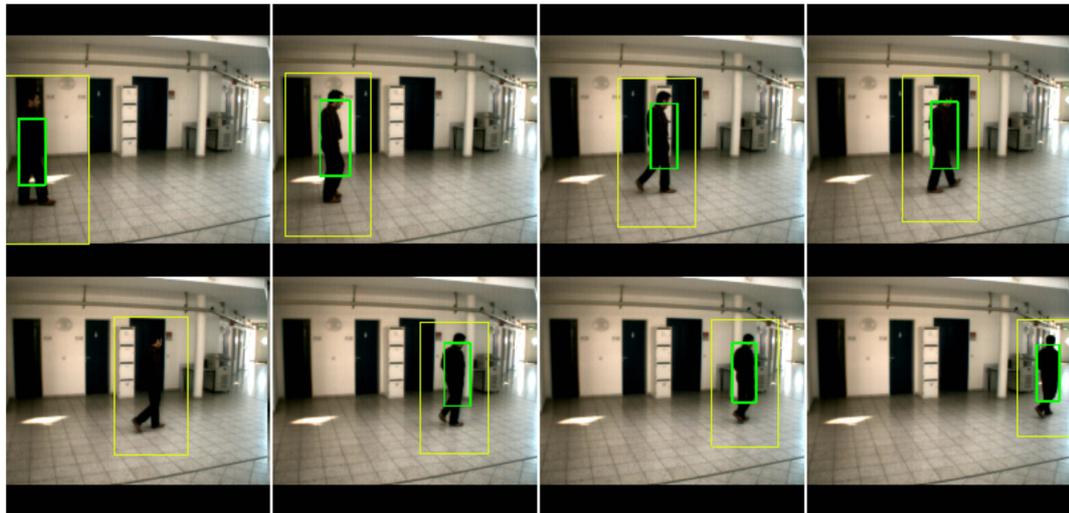


Abbildung 65: Klassifikationssequenz des HOG-Algorithmus.

Aufnahmesequenz (50 Bilder) einer sich durch das Detektionsfeld der ersten MultiCam quer bewegendes Person. Erkennbar ist die Verringerung des Suchgebietes in Abhängigkeit der Distanz, was sich naturgemäß in einer reduzierten Berechnungsdauer niederschlägt (Abb. 66, oberster Graph).

Die sich durch Verringerung des Suchgebiets ergebende Beschleunigung des HOG-Klassifikators, auf Grundlage einer Operation auf einem reduzierten Intensitätsbild von 320x240 Pixel, liefert der obere Graph der Abb. 66. Während das ursprüngliche Verfahren eine relativ konstante Berechnungszeit von ca. 120 ms benötigt, ergibt sich für den unterstützten Algorithmus eine von der Distanz des Objekts abhängige Berechnungsperiode. Je nach Nähe der Objekte zur MultiCam beträgt die notwendige Berechnungsdauer zwischen 20 ms und 45 ms. Zu beachten ist, dass die reduzierten Berechnungsperioden pro zu untersuchendes Objekt anfallen. Eine Analyse der zur Klassifikation notwendigen Zeiträume auf Grundlage von Szenen mit unterschiedlicher Personenzahl erfolgt in Abschnitt 5.2.7.

Der untere Graph von Abb. 66 liefert die Informationen bzgl. der Lokalisationsleistung durch den PMD-Sensor sowie der Detektion einer Person durch den ursprünglichen und den unterstützten Algorithmus. Erkennbar ist die im Vergleich zur originalen HOG-Methode leicht verringerte Klassifikationsleistung. Begründet wird diese Reduktion mit der Tatsache, dass in manchen Fällen das erzeugte Suchfenster zu klein ist und/oder nicht exakt positioniert wird und wodurch dem HOG-Algorithmus unzureichende Informationen zur Verfügung gestellt werden.

Wie auch im Falle der Gesichtsdetektion ließe sich eine Erhöhung der Klassifikationsleistung durch Vergrößerung des Suchfensters erreichen. Als Folge davon ergäbe sich aber wiederum eine direkte Erhöhung der Berechnungsdauer pro Bild, was höchst unerwünscht ist.

Die verringerte Klassifikationsleistung ist jedoch unkritisch, da im Falle des Vorhandenseins mehrerer parallel operierender MSK eine, je nach Objektposition, gleichzeitige und mehrfache Klassifikation desselben Objekts stattfindet. Bedingt durch die Tatsache, dass die Verfolgung der lokalisierten Objekte nicht auf Grundlage der Ergebnisse der Personenklassifikatoren stattfindet, ist eine durchgängige Klassifikation der Objekte nicht von entscheidender Bedeutung.

## 4 Agentenbasierte Detektion, Lokalisation und Klassifikation von Objekten

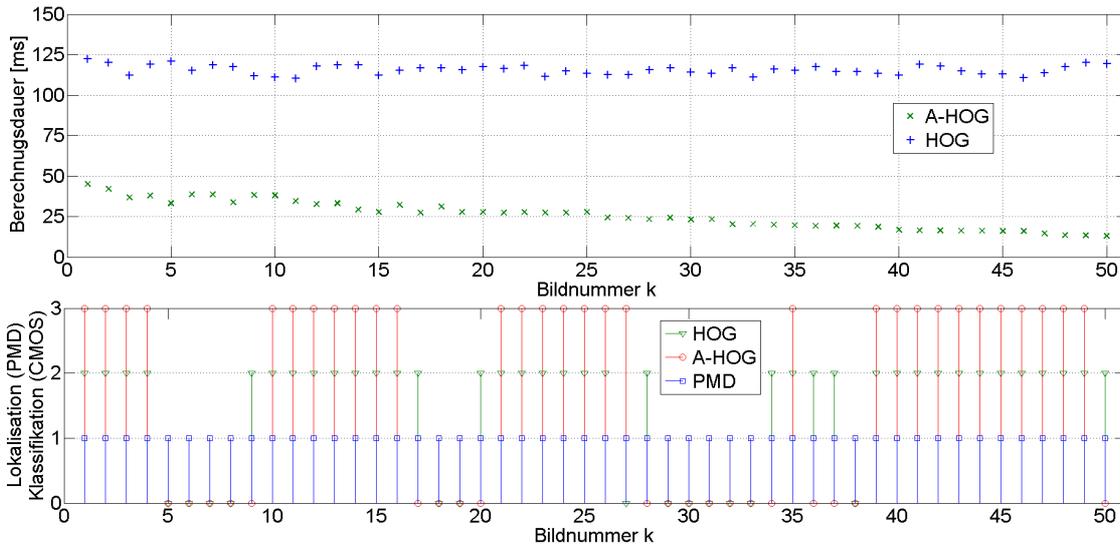


Abbildung 66: Berechnungsdauer, Lokalisation, Klassifikation (HOG).

Ergebnisse der Lokalisation und Klassifikation auf Basis der in Abb. 65 angegebenen Videosequenz. Oben: Dauer der Berechnung pro Bild des HOG-Klassifikators mit und ohne Unterstützung auf Grundlage eines reduzierten MultiCam-Intensitätsbildes von 320x240 Pixel. Unten: Lokalisations- und Klassifikationsergebnisse pro Bild. Von 0 verschiedene Werte entsprechen einer erfolgreich durchgeführten Operation (analog zu dem mittleren Graphen aus Abb. 63, 1 für erfolgreiche Lokalisation, 2 und 3 für erfolgreiche Klassifikation).

Um die Ergebnisse einer durchgeführten Klassifikation beider Algorithmen in die Parametermenge  $c_k^j$  (45) des Aktivitätsclusters  $C_k^j$  einzubinden, wird diese um eine binäre Indikatorvariable  $\beta_k^j$  erweitert. Sie gibt an, ob auf Basis der Informationen des Aktivitätsclusters  $C_k^j$  das Objekt zum Zeitpunkt  $k$  als Person klassifiziert werden konnte. Die finale Form von  $c_k^j$  (45) ergibt sich damit zu:

$$c_k^j = \{ {}^{2D3D}P_k^j, {}^{2D3D}P_{1,k}^{j,PP}, {}^{2D3D}P_{2,k}^{j,PP}, {}^{2D3D}P_{1,k}^{j,AC}, {}^{2D3D}P_{2,k}^{j,AC}, \mathbf{q}_k^j, \mathbf{t}_k^j, \beta_k^j \}. \quad (46)$$

### 4.4 ZUSAMMENFASSUNG

In den vorigen Abschnitten dieses Kapitels wurden die der Objektverfolgung vorgeschalteten Bearbeitungsschritte der JDL-Prozessebene 0 und 1 diskutiert. Begonnen wurde zunächst mit der Beschreibung der im Rahmen dieser Arbeit durchgeführten Präprozessierung der Sensordaten des PMD-Sensors (Abschnitt 4.1.2), welche sich der JDL-Prozessebene 0 zuordnen lassen (Abb. 27). Durch Anwendung einer rechnerisch effektiven Entfernungskorrektur (17), (18) sowie ebenfalls schnell durchführbaren Entfernungs- (19), (21) und Modulationsdatenfilterung (20) konnte die Information des 3D-Sensors auf ein für die Weiterverarbeitung notwendiges Format und Menge transformiert bzw. reduziert werden (Abb. 35).

Im Anschluss dazu erfolgte die Diskussion der Prozesse der JDL-Prozessebene 1. Begonnen wurde zunächst mit der Beschreibung der implementierten Bewegungsdetektion (Abschnitt 4.2.2.1) sowie mit der Modellierung des Hintergrunds (Abschnitt 4.2.2.2) auf Basis der präprozessierten PMD-Daten. Die Vorteile der dargelegten Vorgehensweise (Abb. 39, Abb. 40) sind neben einer schnellen Ausführung eine Unabhängigkeit von den Tageslichtverhältnissen. Zusätzlich dazu besitzt die

## 4 Agentenbasierte Detektion, Lokalisation und Klassifikation von Objekten

---

Modellierung des Hintergrunds die Fähigkeit zur Adaption, wobei die Adaptionsgeschwindigkeit bzw. Lernrate selbst parametrisierbar ist.

Auf Grundlage der vorigen Berechnungsschritte erfolgte in Abschnitt 4.2.3 die Diskussion der Methode der Objektlokalisierung unter Verwendung von Aktivitätsclustern (Abb. 53), welche wiederum durch den Einsatz probabilistischen Clusterns (33) erzeugt wurden. Wie auch im Falle der vorangegangenen Prozessschritte basiert die Objektlokalisierung lediglich auf der Verwendung von Informationen des PMD-Sensors und besitzt damit ebenfalls eine weitgehende Unabhängigkeit von den Tageslichtverhältnissen. Weiterhin erlaubt die Methode der Objektlokalisierung auf Basis von Aktivitätsclustern, aufgrund der Operation auf den niedrig aufgelösten Bilddaten des 3D-Sensors, eine rechnerisch effektive Detektion und Lokalisation von mehreren Objekten im Sichtfeld der MultiCam (Abb. 47).

Der Methode der Objektklassifikation widmete sich Abschnitt 4.3. Neben einer kurzen Vorstellung der zwei eingesetzten Algorithmen zu Klassifikation von Objekten auf Grundlage der Daten des hoch aufgelösten CMOS-Chips erfolgte zusätzlich eine Analyse der zur Klassifikation von Objekten notwendigen Bearbeitungsdauer unter Verwendung variierender ROI (Abb. 56, Tab. 2 und Abb. 59, Tab. 3) sowie eine Bestimmung der Detektionsleistung in Abhängigkeit der Objektentfernung (Abb. 60). Durch Hinzunahme der Distanzinformationen der PMD konnte gewährleistet werden, dass jeweils der für die entsprechende Objektentfernung geeignetste klassifizierende Algorithmus eingesetzt worden ist. Aufgrund der Tatsache, dass die Wahl des Algorithmus zur Objektklassifikation lediglich durch die Objektentfernung bestimmt wird, entfallen zusätzliche Entscheidungen und damit Berechnungen.

Weiterhin konnte gezeigt werden, dass durch die Hinzunahme von durch Tiefeninformationen erzeugtem Vorwissen eine Reduktion der 2D-Intensitätsdaten vorgenommen werden kann (Abb. 63, 66), welche zu einer teilweise erheblichen beschleunigten Berechnung der Personendetektion führt und somit ermöglicht, die verwendeten Klassifikationsalgorithmen im Umfeld eines echtzeitfähigen Systems einzusetzen. Bedingt durch die prinzipielle Unabhängigkeit der pro Bild erzeugten Aktivitätscluster ergibt sich, zwecks beschleunigter Berechnung, die Möglichkeit der parallelen Objektklassifikation.

Weiterhin ist durch die strikte Separation der Lokalisation und Klassifikation gewährleistet, dass eine Lokalisierung der Objekte unabhängig von der jeweiligen Umgebungsbeleuchtung zuverlässig durchgeführt werden kann und dem System somit eine solide Grundfunktionalität verleiht, vor allem im Falle sich beleuchtungsmäßig stark verändernden Umgebungen.



## 5 OBJEKT- UND PERSONENVERFOLGUNG

Der Begriff des Trackings wird im Kontext des maschinellen Sehens im Allgemeinen als die Verfolgung eines Objekts über eine Sequenz von Bildern verstanden. Eine diesbezügliche Definition liefert Bar Shalom et al. in [186]:

*Tracking ist die Estimation des Zustands eines bewegten Objekts basierend auf entfernte Messungen.*

Durch Betrachtung des bewegten Objekts als ein dynamisches System, dessen Entwicklung in der Zeit mittels eines *Systemmodells* (47) [180], [183], [191] beschrieben werden kann<sup>76</sup> und welches sowohl die Systemstör- als auch die Steuergrößen beinhaltet, lässt sich jeder Objektzustand durch einen Zustandsvektor des Systemmodells wiedergeben, dessen Entwicklung in der Zeit durch

$$\mathbf{x}_k = \mathbf{f}_{k-1}(\mathbf{x}_{k-1}, \mathbf{u}_{k-1}, \mathbf{v}_{k-1}) \quad (47)$$

beschrieben wird<sup>77</sup>, wobei  $\mathbf{x}_k \in \mathbb{R}^{n_x}$  den Zustandsvektor des Zeitpunkts  $k$  darstellt, und  $\mathbf{f}_{k-1}$  eine möglicherweise nichtlineare Funktion des vorigen Zustands  $\mathbf{x}_{k-1} \in \mathbb{R}^{n_x}$ , des Steuervektors  $\mathbf{u}_{k-1} \in \mathbb{R}^{n_u}$  sowie des Prozessrauschens, welches durch den Zufallsvariablenvektor  $\mathbf{v}_{k-1} \in \mathbb{R}^{n_v}$  repräsentiert wird, ist.

Im Allgemeinen sind es die Trajektorien der Objekte im Zustandsraum, die im Rahmen einer Objektverfolgung von besonderem Interesse sind, da sie alle notwendigen Informationen bezüglich des Zielobjekts beinhalten.

Der Zustandsvektor an sich ist dabei eine nicht direkt erfassbare Größe. Er muss mithilfe vorhandener Sensorik gemessen werden, wobei der Messprozess selbst durch ein zugrunde liegendes *Messmodell* bzw. *Beobachtungsmodell* (48) [180] beschrieben wird:

$$\mathbf{y}_k = \mathbf{h}_k(\mathbf{x}_k, \mathbf{w}_k), \quad (48)$$

mit  $\mathbf{y}_k \in \mathbb{R}^{n_y}$  als dem Messvektor des Zeitpunkts  $k$ . Die Funktion  $\mathbf{h}_k$  entspricht dabei einer möglicherweise nichtlinearen Funktion des Zustandsvektors  $\mathbf{x}_k$  und  $\mathbf{w}_k \in \mathbb{R}^{n_w}$  als dem mit der Messung assoziierten Messfehler.

Aufgrund der nicht idealen Operation der Sensoren besitzen die Messungen naturgemäß Messfehler und repräsentieren somit eine fehlerbehaftete Abbildung des Systemzustandes. Die Aufgabe des Estimationsprozesses ist es, auf Basis der zur Verfügung stehenden, fehlerbehafteten Informationen (Messungen), eine Schätzung des Systemzustands zu erzeugen, wobei vorhandenes Vorwissen mit einbezogen wird (Abb. 67).

<sup>76</sup>Die im Rahmen dieser Arbeit verwendete zeitdiskrete Formulierung wird durch die ebenfalls zeitdiskrete Form der vorhandenen Messungen begründet. Eine eingängige Einführung in die Modellierung dynamischer Systeme mittels in der Zeit kontinuierlicher sowie diskreter System- und Messmodelle findet sich in [180].

<sup>77</sup>Gleichung (47) entspricht dabei einem vektoriellen Markov-Prozess erster Ordnung.

## 5 Objekt- und Personenverfolgung

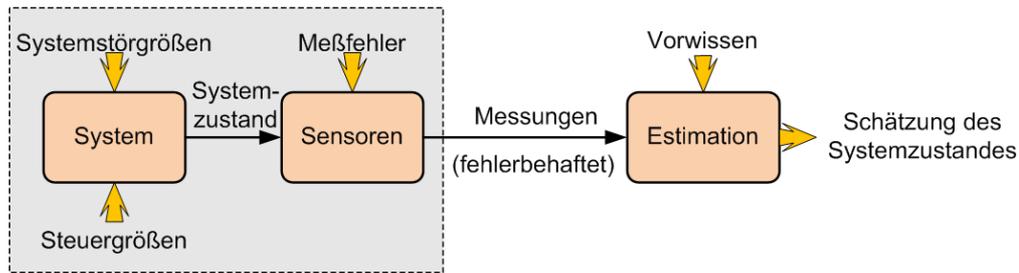


Abbildung 67: Prozess der Estimation des Systemzustands [180].

Den Zusammenhang zwischen System- und Messmodell enthält Abb. 68.

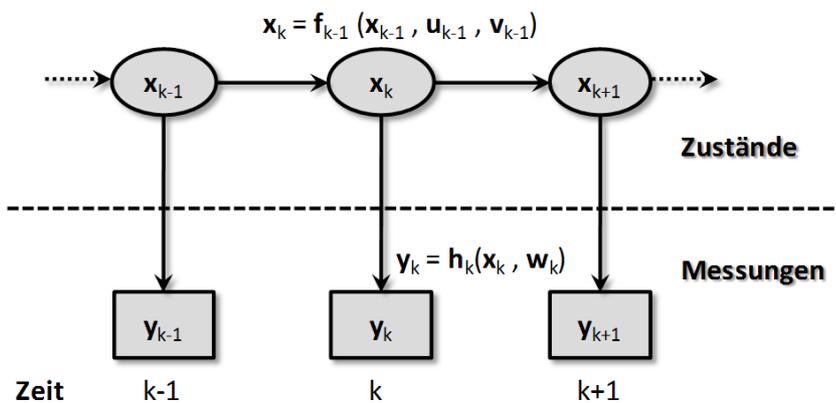


Abbildung 68: Zusammenhang zwischen System- und Messmodell.

Die Entwicklung des Systems in der Zeit erfolgt gemäß des zugrunde liegenden Systemmodells unter Verwendung lediglich der Information des vorangegangenen Zeitpunkts. Für jeden Zeitpunkt ergibt sich mittels des Messmodells eine fehlerbehaftete Abbildung des Systemzustands in Form von Messungen, welche als Grundlage der Aufbereitung der Systemzustände durch den Estimator verwendet werden. Für die Messungen der unterschiedlichen Zeitpunkte wird angenommen, dass sie voneinander unabhängig sind.

Der Anfang dieses Kapitels (Abschnitt 5.1) dient zunächst der Betrachtung der Estimation von Systemzuständen basierend auf bayes'scher Filterung.

Im Anschluss dazu (Abschnitt 5.2) erfolgt die Behandlung der lokal auf einem Multisensorknoten durchgeführten Objektverfolgung auf Basis eines linearen Kalman-Filters [181]-[184]. Einleitend dazu wird der verwendete Ansatz des lokalen Objektmanagements diskutiert. Durch das lokale Objektmanagement werden die notwendigen Informationen erzeugt, auf welche der nachgeschaltete Prozess der globalen Objektverfolgung beruht.

Abschnitt 5.3 diskutiert die Objektverfolgung unter Fusionierung der Informationen der an der Systemoperation teilnehmenden MSK. Nach anfänglicher Beschreibung des globalen Objektmanagements erfolgt anschließend daran die Erläuterung des gewählten Ansatzes zur Objektverfolgung unter Verwendung des CONDENSATION-Algorithmus [192], [193].

Abgeschlossen wird das Kapitel durch die Diskussion der Ergebnisse, herrührend durch die Verwendung zweier MultiCams.

## 5.1 PROBABILISTISCHE ESTIMATION

Grundsätzlich basiert die rekursive bayes'sche Estimation von Systemzuständen auf der Verwendung zweier sukzessiver Schritte:

### 1. SCHRITT: PRÄDIKTION

Ausgehend von der Wahrscheinlichkeitsdichtefunktion  $f_{x_k|Y_k}(\xi_k|Z_k)$  des Zustands zum Zeitpunkt  $k-1$  sowie der Übergangswahrscheinlichkeitsdichte  $f_{x_k|x_{k-1}, Y_{k-1}}(\xi_k|\xi_{k-1}, Z_{k-1})$ , welche die zugrunde liegende Dynamik des Systemmodells (47) enthält, wird zunächst unter Verwendung der Chapman-Kolmogorov-Gleichung [183], [199] eine Prädiktion des Systemzustands

$$\begin{aligned} f_{x_k|Y_{k-1}}(\xi_k|Z_{k-1}) &= \int f_{x_k|x_{k-1}, Y_{k-1}}(\xi_k|\xi_{k-1}, Z_{k-1}) f_{x_{k-1}|Y_{k-1}}(\xi_{k-1}|Z_{k-1}) d\xi_{k-1} \\ &= \int f_{x_k|x_{k-1}}(\xi_k|\xi_{k-1}) f_{x_{k-1}|Y_{k-1}}(\xi_{k-1}|Z_{k-1}) d\xi_{k-1} \end{aligned} \quad (49)$$

berechnet.  $x_k$  entspricht dabei dem Systemzustand des Zeitpunkts  $k$ ,  $Y_{k-1} = \{y_1, \dots, y_{k-1}\}$  der Menge der Messungen bis einschließlich  $k-1$ .  $\xi_k$  und  $Z_{k-1} = \{\zeta_1, \dots, \zeta_{k-1}\}$  repräsentieren die zugehörigen Dummyvariablen bzw. Menge der Dummymessvektoren [180]. Eine in (49) getroffene Annahme ist dabei, dass die Übergangswahrscheinlichkeitsdichte einen Markov-Prozess erster Ordnung darstellt, womit sich die Vereinfachung  $f_{x_k|x_{k-1}, Y_{k-1}}(\xi_k|\xi_{k-1}, Z_{k-1}) = f_{x_k|x_{k-1}}(\xi_k|\xi_{k-1})$  ergibt.

### 2. SCHRITT: AKTUALISIERUNG

Unter Zuhilfenahme einer zum Zeitpunkt  $k$  vorhandenen Messung  $y_k$  erfolgt eine Aktualisierung der im vorigen Schritt durchgeführten Zustandsprädiktion (49) durch

$$f_{x_k|Y_k}(\xi_k|Z_k) = \frac{f_{y_k|x_k}(\zeta_k|\xi_k) f_{x_k|Y_{k-1}}(\xi_k|Z_{k-1})}{f_{y_k|Y_{k-1}}(\zeta_k|Z_{k-1})}, \quad (50)$$

wobei das Messmodell (48) durch den Term  $f_{y_k|x_k}(\zeta_k|\xi_k)$  in die Berechnung Eingang findet. Eine Annahme, welche zusätzlich getroffen wird, ist, dass die Wahrscheinlichkeitsdichte des initialen Systemzustands  $f_{x_0|Y_0}(\xi_0|Z_0) = f_{x_0}(\xi_0)$  gegeben ist.

Die Simplität der bayes'schen Estimation, bestehend aus Prädiktion (49) und Aktualisierung (50), was einer rekursiven Propagation der posterioren Wahrscheinlichkeitsdichte entspricht [183], täuscht darüber hinweg, dass es sich hierbei lediglich um einen konzeptionellen Ansatz handelt. Im Allgemeinen sind die Prädiktions- und Aktualisierungsgleichungen jedoch nur unter der Annahme relativ restriktiver Bedingungen analytisch lösbar.

Einen solchen Fall, welcher im Sinne der bayes'schen Estimation als optimale Lösung betrachtet wird [183], [184], stellt das Kalman-Filter dar (Abschnitt 5.2). Ist eine Lösung von (49) und (50) nicht gegeben, muss auf eine Näherung ausgewichen werden. Ein Repräsentant dieser Klasse ist der CONDENSATION-Algorithmus, dessen Diskussion im Rahmen der globalen Objektverfolgung (Abschnitt 5.3) stattfindet.

### 5.2 LOKALE OBJEKT- UND PERSONENVERFOLGUNG (JDL-PROZESSEBENE 1)

Unter lokaler Objekt- bzw. Personenverfolgung wird im Rahmen dieser Arbeit die Verfolgung von Objekten bzw. Personen unter Verwendung lediglich der Informationen eines MSK verstanden. Ziel dieses Prozesses ist die zu jedem Zeitpunkt erneut stattfindende Erzeugung eines lokalen Zustandsvektors, welcher die Sicht der Welt bzw. Umgebung aus der Perspektive des entsprechenden MSK wiedergibt.

#### 5.2.1 LOKALES OBJEKTMANAGEMENT

Die Erzeugung, Erhaltung sowie Entfernung von lokalen Objekten geschieht im Rahmen des lokalen Objektmanagements. Als Informationsgrundlage dienen dabei die Daten aus den Parametermengen (46) der vorhandenen Aktivitätscluster sukzessiver Zeitpunkte sowie einem auf diesen Daten operierenden Ähnlichkeitsmaß (82). Aktivitätscluster zweier aufeinanderfolgender Zeitpunkte, welche die höchste Ähnlichkeit besitzen, werden dabei als zeitlich zusammenhängend interpretiert. Gelingt es, einen zeitlichen Zusammenhang zwischen Aktivitätsclustern über einer durch  $O^{\text{in}} \in \mathbb{N}$  definierten Anzahl aufeinanderfolgender Bilder (Zeitpunkten) aufrechtzuerhalten, führt das zur Erzeugung eines eindeutig gekennzeichneten lokalen Objekts (Abb. 69).



Abbildung 69: Lokale Objekterzeugung.

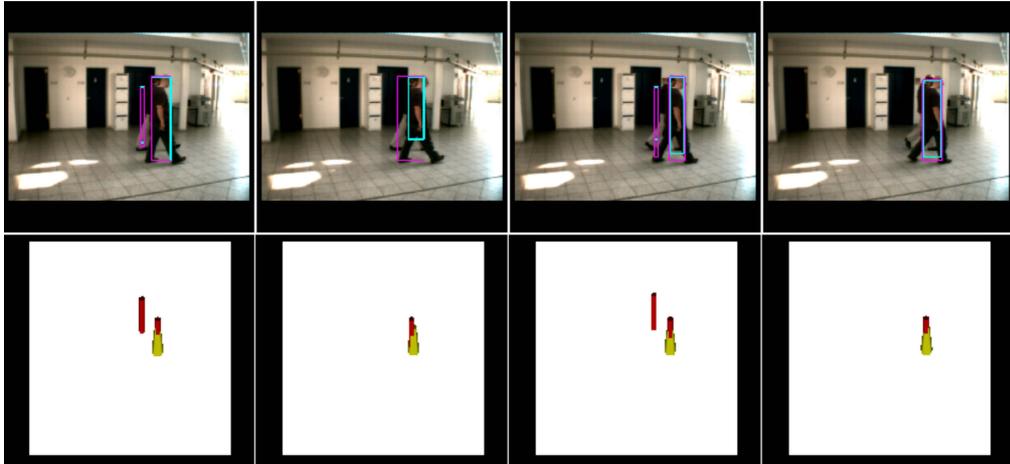
*Oben: Eintritt eines Objekts in den Detektionsbereich der MultiCam. Unten: Positionen der lokalisierten bzw. generierten Objekte im Koordinatensystem  $K^{2D3D}$ . Rote Stäbe entsprechen Positionen des Aktivitätsclusters, gelbe Kreise den Positionen des erzeugten lokalen Objekts. Die Erzeugung eines lokalen Objekts erfolgt aufgrund eines zeitlichen Zusammenhangs von Aktivitätsclustern über einer Länge von  $O^{\text{in}}=2$  aufeinanderfolgenden Bildern.*

Durch die Forderung eines zeitlichen Zusammenhangs von Aktivitätsclustern über eine zusammenhängende Bildfolge der Länge  $O^{\text{in}}$  hinweg, wird die Erzeugung lokaler Objekte auf Grundlage flüchtiger Aktivitätscluster verhindert (Abb. 70).

Lokal erzeugte Objekte bedürfen ihrerseits einer Erhaltung durch Aktivitätscluster, welche ihnen im Rahmen der Objekterhaltung zugewiesen werden. Gelingt es nicht, einem lokalen Objekt in einer durch  $O^{\text{out}} \in \mathbb{N}$  bestimmten Anzahl aufeinanderfolgender Bilder einen es erhaltenden Aktivitätscluster zuzuweisen, wird es gelöscht.  $O^{\text{out}}$  wird als das maximale Alter eines Objekts interpretiert und entspricht der Zeitspanne, welcher es ohne Informationen durch einen es erhaltenden Aktivitätscluster überleben kann. Aufgrund der Objekalterung über einen durch  $O^{\text{out}}$  bestimmten Zeitrahmen hinweg wird sichergestellt, dass kurzzeitige Fluktuationen in der Erzeugung von erhaltenden Aktivitätsclustern aufgrund sporadischer gegenseitiger Verdeckung der

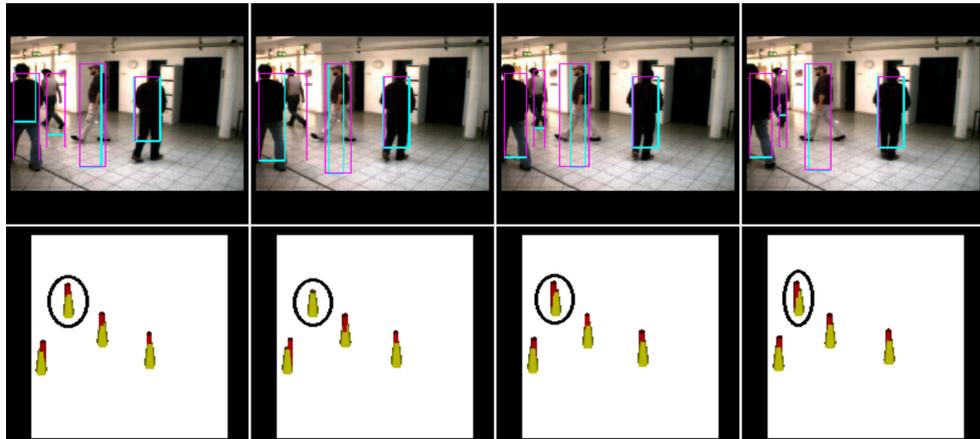
## 5 Objekt- und Personenverfolgung

realen Objekte (Abb. 71) oder unzureichender Daten (Abb. 72) nicht zu einer Entfernung des lokalen Objekts führen.



*Abbildung 70: Unterdrückte Objekterzeugung.*

*Obere Zeile: Sequenz aus vier aufeinanderfolgenden Aufnahmen mit flüchtiger, bewegungsbedingter Erzeugung von Aktivitätsclustern (rote Stäbe). Untere Zeile: Durch die Forderung eines zeitlichen Zusammenhangs von Aktivitätsclustern über einer Länge von  $O^n=2$  aufeinanderfolgenden Bilder hinweg wird die Erzeugung von Objekten (gelbe Konen) auf Grundlage flüchtiger Informationen unterdrückt.*



*Abbildung 71: Überbrückung fehlender Objekterhaltung.*

*Bedingt durch den Alterungsprozess der Objekte kann das Fehlen erhaltender Aktivitätscluster kurzzeitig überbrückt werden. Parameter:  $O^{ult}=6$ .*

Aufgrund der Bewegung der realen Objekte durch unterschiedliche Intensitätsbereiche der Beleuchtung sind die Positionsinformationen der Aktivitätscluster Schwankungen unterworfen. Im Falle einer direkten Verwendung dieser Informationen zur Positionsbestimmung lokaler Objekte wird die Fluktuation ungehindert übertragen und ist unerwünscht. Weiterhin ergibt sich im Falle einer direkten Verwendung der Aktivitätsclusterpositionen das Problem, dass auftretende Erzeugungsflektuationen von Aktivitätscluster (Abb. 72) dazu führen, dass einem lokalen Objekt keine Informationen bezüglich seiner momentanen Position zugeführt werden können.

Eine detaillierte Beschreibung des lokalen Prozesses zur Erzeugung, Erhaltung und Eliminierung von lokalen Objekten findet in Anhang A statt. Die Reduktion der Positionsfluktuationen der lokalen Objekte sowie die Überbrückung von kurzen Perioden, in denen lokalen Objekten keine erhaltenden Aktivitätscluster zugewiesen werden können, findet auf Grundlage der Verwendung eines linearen Kalman-Filters statt, dessen kurze Diskussion der folgende Abschnitt gewidmet ist.

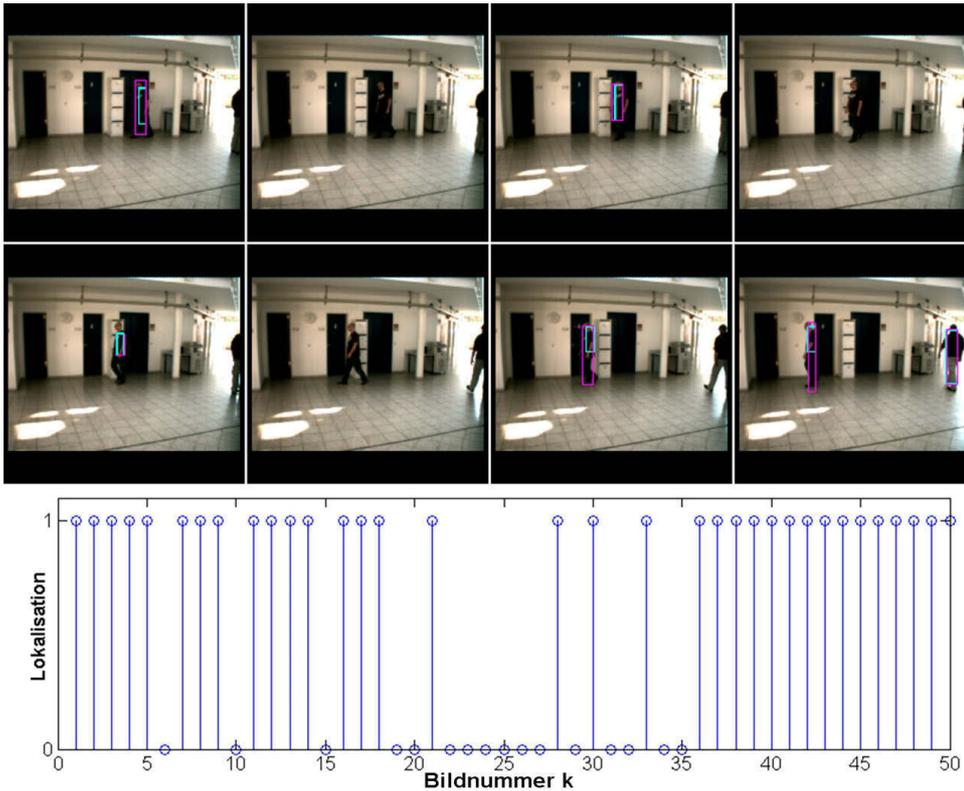


Abbildung 72: Fluktuation der Aktivitätsclustergeneration.

Obere Bildfolge: Ausgewählte Bilder zur Demonstration fehlgeschlagener Lokalisation des Objekts aufgrund unzureichender Informationen. Unterer Graph: Per Bild aufgetragene Objektlokalisierung der aus 50 Aufnahmen bestehenden Videosequenz. Jeder Bildnummer mit einer 0 entspricht einer Aufnahme, in welcher eine Objektlokalisierung aufgrund zu geringer Informationen nicht durchgeführt werden konnte.

### 5.2.2 KALMAN-FILTER

Eine im Sinne der bayes'schen Filterung optimale Bestimmung eines Systemzustands ermöglicht das Kalman-Filter für den Fall, dass folgende restriktive Annahmen erfüllt werden [183]:

1. Die Zustandsübergangsfunktion  $f_{k-1}(\underline{x}_{k-1}, \underline{u}_{k-1}, \underline{v}_{k-1})$  ist eine bekannte lineare Funktion in allen Parametern.
2. Das Messmodell  $h_k(\underline{x}_k, \underline{w}_k)$  ist bekannt und linear in  $\underline{x}_k \in \mathbb{R}^{n_x}$ .
3.  $\underline{v}_{k-1} \in \mathbb{R}^{n_v}$  und  $\underline{w}_k \in \mathbb{R}^{n_w}$  werden durch Normalverteilungen mit bekannten Parametern erzeugt.
4. Die Beschreibung des Systemzustands jedes Zeitpunkts ist durch eine monomodale normalverteilte Wahrscheinlichkeitsdichtefunktion gegeben.

Unter Verwendung obiger Annahmen kann das Systemmodell (47) umgeformt werden zu [180]

$$\underline{x}_k = A_{k-1} \underline{x}_{k-1} + \underline{u}_{k-1} + \underline{v}_{k-1}, \quad (51)$$

## 5 Objekt- und Personenverfolgung

---

mit  $A_{k-1} \in \mathbb{R}^{n_y \times \mathbb{R}^{n_x}}$  als der potenziell zeitvarianten Zustandsübergangsfunktion. Das Messmodell (48) ergibt sich zu [180]

$$\underline{y}_k = C_k \underline{x}_k + \underline{w}_k, \quad (52)$$

mit  $C_k \in \mathbb{R}^{n_y \times \mathbb{R}^{n_x}}$  als der ebenfalls potenziell zeitvarianten Beobachtungsmatrix.

Die stochastischen Eingangsgrößen  $\underline{v}_{k-1}$  und  $\underline{w}_k$  werden aufgrund der getroffenen Annahmen als normalverteilt, weiß, nullzentriert und gegenseitig unabhängig angenommen. Weiterhin sind  $Q_{k-1} = (\underline{v}_{k-1} \underline{v}_j^T) \delta_{k-1,j} \in \mathbb{R}^{n_y \times \mathbb{R}^{n_y}}$  und  $R_k = (\underline{w}_k \underline{w}_j^T) \delta_{k,j} \in \mathbb{R}^{n_w \times \mathbb{R}^{n_w}}$  die zugehörigen Kovarianzmatrizen. Da im Rahmen dieser Arbeit die betrachteten dynamischen Systeme keine externe Steuerung erfahren, kann der entsprechende Eingangsvektor  $\underline{u}_{k-1} \in \mathbb{R}^{n_u}$  aus (51) vernachlässigt werden.

Die Verbindung zwischen der Prädiktions- und Aktualisierungsgleichung (90, 93) ist im Kontext des Kalman-Filters gegeben durch [181]:

### PRÄDIKTION

$$\hat{\underline{x}}_k = A_{k-1} \underline{x}_{k-1}, \quad \underline{x}_0 \text{ bekannt.} \quad (53)$$

Die Kovarianz des Fehlers der Prädiktion wird bestimmt durch:

$$\hat{P}_k = A_{k-1} \cdot P_{k-1} \cdot A_{k-1}^T + Q_{k-1}, \quad P_0 \text{ bekannt.} \quad (54)$$

### AKTUALISIERUNG

$$\underline{x}_k = \hat{\underline{x}}_k + K_k \underline{r}_k \quad (55)$$

$$K_k = \hat{P}_k C_k^T [C_k \hat{P}_k C_k^T + R_k]^{-1} \quad (56)$$

mit  $K_k$  als dem Kalmangain und

$$\underline{r}_k = \underline{y}_k - C_k \hat{\underline{x}}_k \quad (57)$$

als dem Residuum. Die Fehlerkovarianz des Schätzwertes ist definiert als:

$$P_k = (I - K_k C_k) \hat{P}_k. \quad (58)$$

## 5 Objekt- und Personenverfolgung

### 5.2.3 ESTIMATION DER OBJEKTZUSTÄNDE

Im verfolgten Ansatz werden lokale Objekte als sich auf der  $u-z$ -Ebene von  $K^{2D3D}$  bewegendende Zylinder mit Radius  $\Delta^{max}$  und der Höhe des letzten zugewiesenen Aktivitätsclusters modelliert. Sei

$$\underline{x}_{q,k-1} = \begin{bmatrix} \underline{\rho}_{q,k-1} \\ \dot{\underline{\rho}}_{q,k-1} \end{bmatrix} \quad (59)$$

der Zustandsvektor des lokalen Objekts  $\Omega_q$  im Koordinatensystem  $K^{2D3D}$  zum Zeitpunkt  $k-1$  und  $\underline{\rho}_{q,k-1}$ ,  $\dot{\underline{\rho}}_{q,k-1}$  als der entsprechenden Position und Geschwindigkeit in der  $u-z$ -Ebene von  $K^{2D3D}$ . Verwendet wird das lineare Systemmodell (51)

$$\underline{x}_{q,k} = \underline{f}_{k-1}(\underline{x}_{q,k-1}, \underline{v}_{q,k-1}) = A_{k-1} \underline{x}_{q,k-1} + \underline{v}_{q,k-1} = \begin{bmatrix} \underline{\rho}_{q,k-1} + \Delta t_k \dot{\underline{\rho}}_{q,k-1} + \underline{v}_{1,q,k-1} \\ \dot{\underline{\rho}}_{q,k-1} + \underline{v}_{2,q,k-1} \end{bmatrix} \quad (60)$$

mit  $\underline{v}_{q,k-1} = [\underline{v}_{1,q,k-1}, \underline{v}_{2,q,k-1}]^T$ . Der Term  $\Delta t_k = t_k - t_{k-1}$  entspricht dabei der Zeitdifferenz zweier aufeinanderfolgender Bilder<sup>78</sup>. Der Messvektor  $\underline{y}_k$  ist definiert als:

$$\underline{y}_{q,k} = \begin{bmatrix} {}^{2D3D}\hat{\underline{P}}_k^i \\ \Delta {}^{2D3D}\hat{\underline{P}}_k^j \Delta t_k^{-1} \end{bmatrix} = \begin{bmatrix} {}^{2D3D}\hat{\underline{P}}_k^i \\ ({}^{2D3D}\hat{\underline{P}}_k^j - {}^{2D3D}\hat{\underline{P}}_{k-1}^j) \Delta t_k^{-1} \end{bmatrix}, \quad (61)$$

mit  ${}^{2D3D}\hat{\underline{P}}_k^i$  als der Position der primären Partition des erhaltenden Aktivitätsclusters  $C_k^i$  in der  $u-z$ -Ebene von  $K^{2D3D}$  und  $\Delta t_k^{-1} \in \mathbb{R}^+$  als der Zeitdifferenz der Bilder zum Zeitpunkt  $k$  sowie  $k-1$ , welche mittels der Zeitstempel der akquirierten Bilder berechnet wird. Das verwendete Messmodell (48) ist definiert vermöge

$$\underline{y}_{q,k} = \underline{h}_k(\underline{x}_{q,k}, \underline{w}_{q,k}) = C_k \underline{x}_{q,k} + \underline{w}_{q,k} = \mathbf{1} \underline{x}_{q,k} + \underline{w}_{q,k}. \quad (62)$$

Sei weiterhin  $\Gamma_q$  der dem Objekt  $\Omega_q$  zugewiesene Kalman-Filter. Für die Berechnung des Zustandsvektors des Folgezeitpunkts existieren zwei Möglichkeiten. Zunächst wird angenommen, dass für den Zeitpunkt  $k$  der Nachfolger  $C_k^i$  von  $C_{k-1}^j$  existiert, dann gilt:

$$\underline{x}_{q,k} = \Gamma_q(\underline{y}_{q,k}). \quad (63)$$

Für den Fall eines nicht vorhandenen erhaltenden Aktivitätsclusters sowie gültigen Alters des Objekts wird der vorige filtergenerierte Zustand als neuer Messwert interpretiert:

$$\underline{x}_{q,k} = \Gamma_q(\underline{x}_{q,k-1}). \quad (64)$$

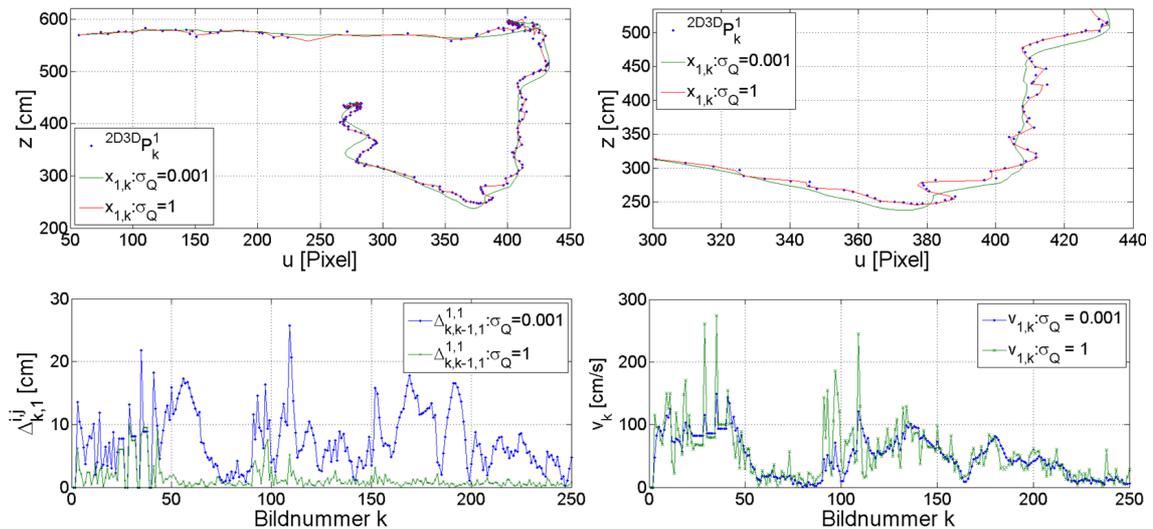
<sup>78</sup>Jede Aufnahme wird automatisch mit einem Zeitstempel versehen.

## 5 Objekt- und Personenverfolgung

Die Kovarianzmatrizen des System- sowie Messrauschens werden als diagonal, zeitlich konstant und mit gleichen Diagonalelementen angenommen:  $Q_k=Q=\sigma_Q \mathbf{I}$  und  $R_k=R=\sigma_R \mathbf{I}$ , mit  $\sigma_R, \sigma_Q \in \mathbb{R}^+$ . Eine Anpassung der Filterdynamik wird durch Veränderung des Wertes  $\sigma_Q$  erreicht, welcher den Einfluss neuer Messwerte im Filterprozess bestimmt [181].

### 5.2.4 ANALYSE DER DYNAMIK

Während zu große Werte von  $\sigma_Q$  zu einer zu strikten Verfolgung des Objekts führen (Abb. 73), wird durch die Wahl zu geringer Werte die Trägheit des Filters erhöht. Änderungen der Geschwindigkeit durch Beschleunigen, plötzliches Umkehren oder ändern der Bewegungsrichtung haben zur Folge, dass das Filter das Objekt verliert. Als Ergebnis altert das verlorene Objekt und wird entfernt.



*Abbildung 73: Objektverfolgung bei hohem und moderatem Systemrauschen.*

*Analyse durchgeführt auf einer Videosequenz bestehend aus 250 Bildern. Oben links: Positionen eines lokal erzeugten Objekts auf Basis der Bewegungen einer Person (Abb. 75) sowie zwei durch das Kalman-Filter erzeugte Objekttrajektorien unter Verwendung zweier unterschiedlicher Varianzen in der  $u-z$ -Ebene des Koordinatensystems  $K^{2D3D}$ . Zu erkennen ist die strenge ( $\sigma_Q=1$ ) bzw. gedämpfte ( $\sigma_Q=0.001$ ) Verfolgung des Objekts. Oben rechts: Ausschnitt der Objekttrajektorie. Deutlich zu erkennen ist die durch die Trägheit der Filters reduzierte Positionsfluktuation aufgrund geringerer Varianz. Unten links: Distanz des lokalen Objekts zu dem generierenden und erhaltenden Aktivitätscluster. Auch hier ist der Einfluss der unterschiedlichen Varianzen deutlich zu erkennen. Während durch hohe Werte von  $\sigma_Q$  eine gewisse Nähe zu dem Aktivitätscluster erzwungen wird, führen moderate Werte erwartungsgemäß zu einer steigenden Entfernung des Objekts. Unten rechts: Geschwindigkeiten des Objekts. Während für den moderaten Varianzwert die Geschwindigkeit des Objekts plausibel erscheint, ergeben sich im Falle hoher Varianzen Geschwindigkeiten, die teilweise drastisch oberhalb der entsprechenden Objektbewegung liegen. Erklärt werden diese Spitzen dadurch, dass im Falle einer strengen Verfolgung die Fluktuation stärkeren Einfluss auf die Bestimmung der Geschwindigkeiten nimmt. Für den Fall relativ hoher Distanzvariationen, was besonders durch zunehmende Entfernung in Erscheinung tritt, ergeben sich Geschwindigkeiten in aufeinanderfolgenden Zeitpunkten, die sich wesentlich von den Objektgeschwindigkeiten unterscheiden. Durch Ausnutzung der dämpfenden Eigenschaft des Filters kann dieser Effekt minimiert werden, sodass die aus den Messungen berechneten Geschwindigkeiten des Objekts der tatsächlichen Geschwindigkeit der realen Person in größerem Maße entspricht.*

## 5 Objekt- und Personenverfolgung

Für die Aktivitätscluster, welche aufgrund des realen Objektes weiterhin erzeugt werden, gilt, dass im Falle einer hinreichend langen Anzahl zeitlich zusammenhängender Aktivitätscluster ein neues Objekt erzeugt wird (Abb. 74). Diese zusätzliche lokale Objekterzeugung ist deshalb unerwünscht, weil sie die Periode der eindeutigen Zuweisbarkeit zwischen einem realen und einem im System erzeugten Objekt verringert, und führt dazu, dass dem realen Objekt mehrere, lokal erzeugte Objekte zugewiesen werden, was letztendlich zur Folge hat, dass die Trajektorien lokaler Objekte verkürzt werden und ihr Informationsgehalt vermindert wird.

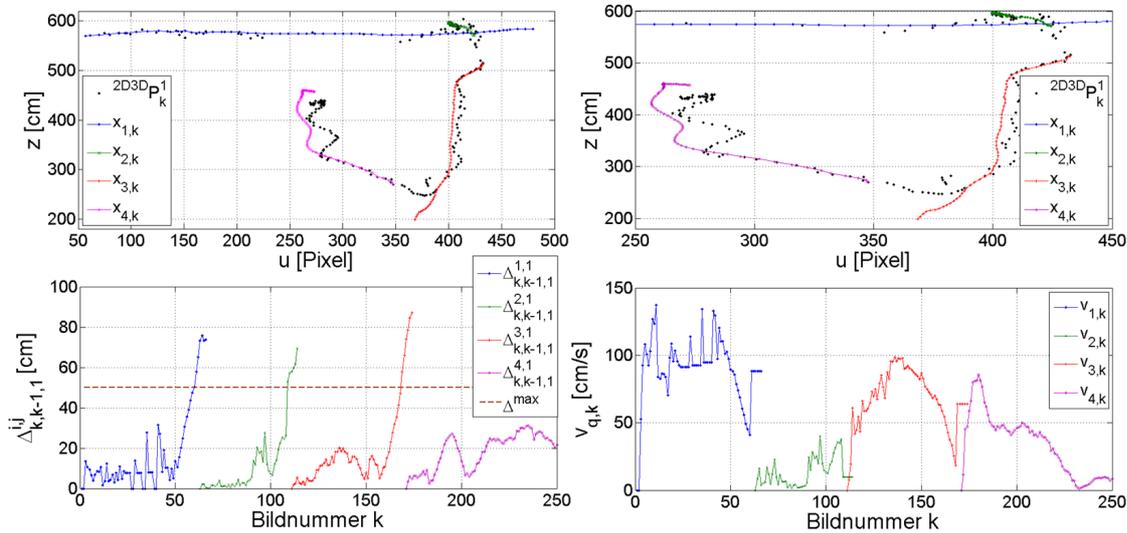


Abbildung 74: Objektverfolgung bei geringem Systemrauschen.

Objektverfolgung unter Verwendung niedriger Varianzen ( $\sigma_{\varrho}=0.00001$ ). Bedingt durch die Trägheit des Filters kommt es, aufgrund des mehrmaligen Objektverlustes, zu zusätzlichen, ungewünschten Objektgenerationen. Oben links: Wie auch Abb. 73 enthält dieser Graph zunächst die Positionen des generierenden bzw. erhaltenden Aktivitätsclusters. Im Unterschied zum Fall hoher bzw. moderater Varianzen führt ein zu niedriger Wert dazu, dass das Kalman-Filter träge auf Änderungen der Bewegung reagiert. Bereits moderate Objektgeschwindigkeiten bzw. deren Änderungen haben zur Folge, dass die Distanz zwischen dem Objekt und dem es erhaltenden Aktivitätscluster eine Schwellgröße  $\Delta^{\max}$  überschreitet und damit den Alterungsprozess des Objekts auslöst. Eine fehlende Neuzuweisung des ursprünglichen Objekts aufgrund zu großer Distanz führt zum Einen zur Entfernung dessen sowie einer Erzeugung eines neuen lokalen Objekts zum Anderen. Im Gegensatz zu dem Fall hoher bzw. moderater Varianzen, wo es lediglich zu einer Objekterzeugung kommt, zeigt sich in diesem Fall, dass im Verlauf der Personenbewegung vier unterschiedliche lokale Objekte erzeugt werden und somit die Trajektorieninformation deutlich reduzieren. Oben rechts: Trajektorienausschnitt. Erkennbar sind hier die Abbrüche der jeweiligen Bewegungslinien der assoziierten Objekte bedingt durch die Filterträgheit. Die Abstände zwischen den jeweiligen Trajektorien ist durch zwei Faktoren bedingt. Zunächst bedarf es einiger Schritte, bis das Objekt sich hinreichend weit vom Aktivitätscluster fortbewegt hat, um als isoliert zu gelten. Erst danach setzt der Prozess der erneuten Objektgeneration ein, welcher wiederum, bedingt durch die notwendige Länge der Assoziationssequenz, nicht sofort erfolgt. Unten links: Distanzen der lokalen Objekte zu den jeweils mit ihnen assoziierten Aktivitätsclustern. Bei einer zu langen Überschreitung des Maximums der zu einer Assoziation notwendigen Distanz  $\Delta^{\max}$  reagiert das System durch Erzeugung eines neuen Objekts. Unten rechts: Geschwindigkeiten der im Laufe der Bewegung erzeugten Objekte. Der Graph zeigt, dass es nicht notwendigerweise hoher Geschwindigkeiten bedarf, um den Kontakt zwischen erhaltenden Aktivitätsclustern und Objekt zu trennen. Abrupte Richtungsänderungen bei relativ moderaten Geschwindigkeiten führen aufgrund der Filterträgheit ebenfalls zum Verlust einer Objektzuweisung.

## 5 Objekt- und Personenverfolgung

Zwecks Verdeutlichung der Objektbewegungen im realen Raum sind in Abb. 75 ausgewählte Szenen der analysierten Videosequenz abgebildet worden.

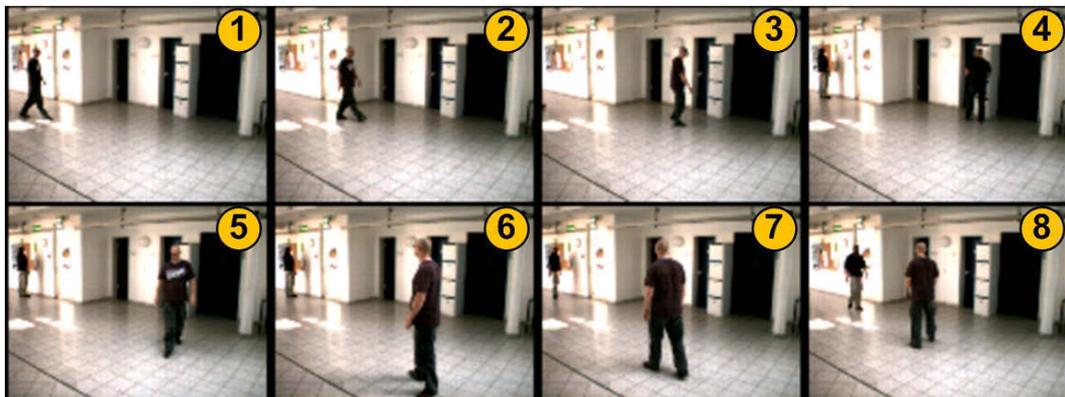


Abbildung 75: Ausgewählte Bilder der untersuchten Verfolgungsszene.

### 5.2.5 POSITIONSWICHTUNG

Eine der fundamentalen Aussagen, die in Bezug auf der Verwendung der MultiCam gemacht werden kann, ist, dass die Varianz der Entfernungsdaten direkt von der Modulationsamplitude () [36], [62] abhängig ist, welche wiederum direkt mit der distanzabhängigen Lichtintensität sowie Lichtverteilung der verwendeten Beleuchtung zusammenhängt. Besonders eine inhomogene Ausleuchtung des Sichtbereichs führt zu Bildbereichen, deren Distanzdaten unterschiedliche Verlässlichkeit besitzen.

Um dieser positionsabhängigen Qualität der Entfernungsmessung Rechnung zu tragen, werden die Positionsinformationen der Objekte mit zwei zusätzlichen Gütefaktoren versehen. Als Grundlage der Berechnung des Ersten, welche der nichtlinearen Distanzabhängigkeit der Intensität Rechnung trägt, dient eine Funktion, welche das Distanzintervall der MultiCam auf das Einheitsintervall  $[z_{min}, z_{max}] \rightarrow [1, 0]$  abbildet. Zur Bestimmung der Funktion sind die Modulationsamplituden eines Pixels mit maximaler Ausleuchtung in Abhängigkeit der Distanz zu einer Fläche aufgetragen. Genähert wurde der Graph durch eine Exponentialfunktion<sup>79</sup> (Abb. 76).

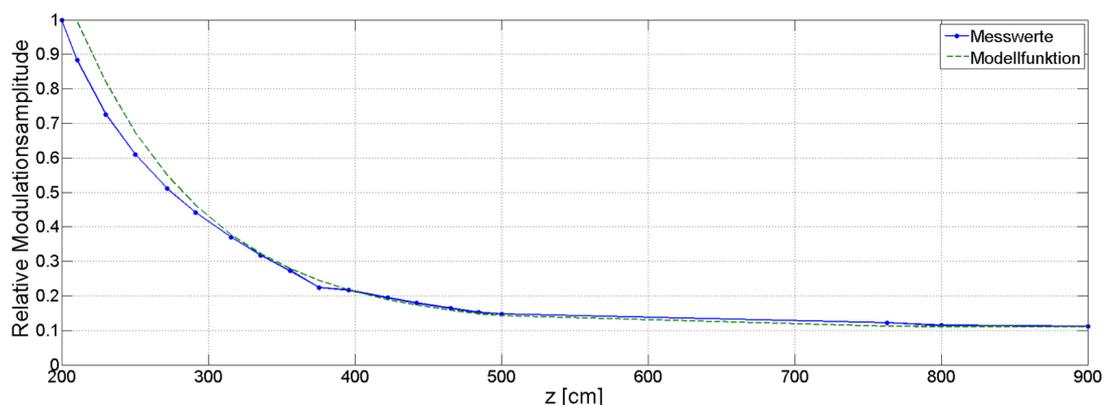


Abbildung 76: Relative Modulationsamplitude in Abhängigkeit der Distanz.

Charakteristische Abnahme der Modulationsamplitude in Abhängigkeit der Distanz. Für die Entfernungen zwischen 1.5m und 2m ist die relative Modulationsamplitude auf 1 gesetzt worden.

<sup>79</sup>Der Befund einer mit der Distanz exponentiell abfallenden Modulationsamplitude deckt sich mit den Ergebnissen der in [62] durchgeführten Untersuchungen.

## 5 Objekt- und Personenverfolgung

Normiert wurden die Modulationsamplituden in Bezug auf den im Abstand von 200 cm zur MultiCam gemessenen Wert.

Ein zusätzlicher Faktor wird in Abhängigkeit der Position des Punktes in der  $u-v$ -Ebene des PMD-Chip-Koordinatensystems  $K^{3D}$  berechnet. Modelliert wird dazu die Lichtverteilung der verwendeten Beleuchtung durch die Superposition zweier multivariater, rotationsymmetrischer Gaußfunktionen:

$$G_{1,2}(u, v) = \exp\left(-\frac{1}{2} \frac{(u - \mu_u^{1,2})^2 + (v - \mu_v^{1,2})^2}{\sigma^2}\right). \quad (65)$$

Die Parameter  $\sigma, \mu_{u,v}^{1,2}$  von (65) sind durch im Vorfeld durchgeführte Untersuchungen bestimmt worden (Abb. 77).

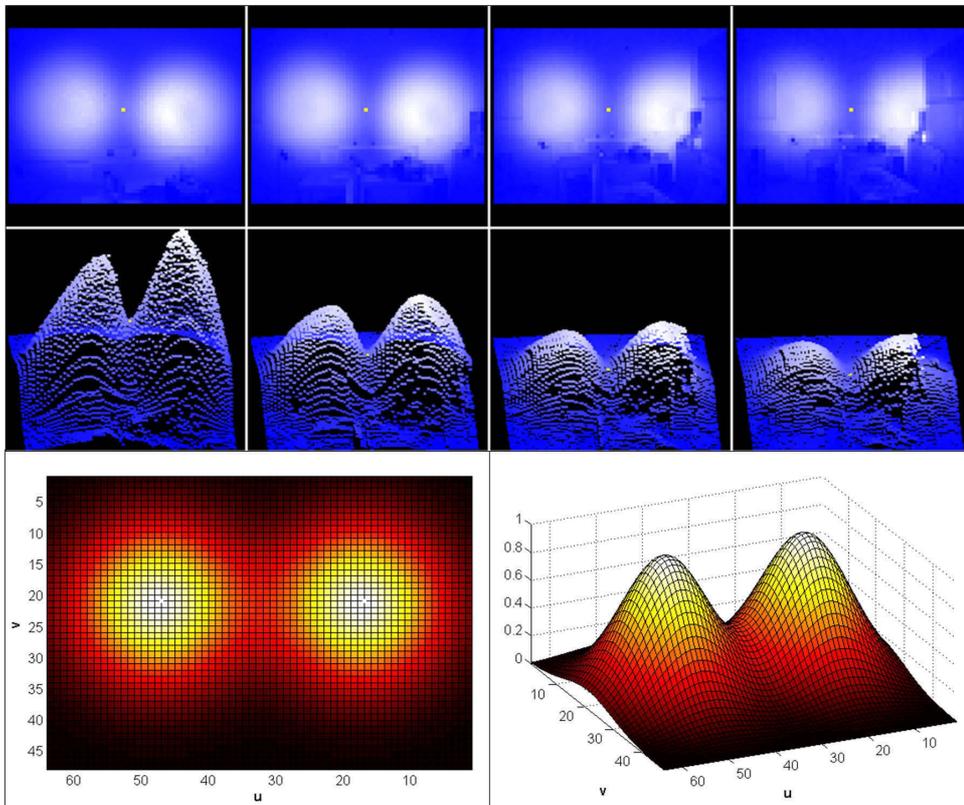


Abbildung 77: Modellierung der Beleuchtungsintensität.

*Obere Zeile: Amplitude der Modulation in orthogonaler Projektion in einer Entfernung von 250, 335, 420 und 500cm. Mittlere Zeile: Modulationsamplitude unter verändertem Blickwinkel. Die Form der Modulationsamplituden lässt sich mit dem Umstand erklären, dass per MultiCam jeweils zwei Beleuchtungsmodulare, welche links und rechts neben der Kamera montiert wurden, verwendet worden sind (siehe dazu auch Abb. 6). Die Ausrichtung der Beleuchtungen wurde dabei so gewählt, dass ein möglichst großer Teil des Sichtbereichs der MultiCam abgedeckt wurde. Untere Zeile: Modellierung der Intensitätsverteilung durch zwei multivariate, rotationsymmetrische Gaußfunktionen (65). Parametersatz:  $\mu_u^1 = 17, \mu_u^2 = 47, \mu_v^1 = \mu_v^2 = 21, \sigma = 8.5$ .*

## 5 Objekt- und Personenverfolgung

Je nach Position des Punktes in  $K^{3D}$  ergibt sich der Faktor  $\gamma^{uv}$  durch

$$\gamma^{uv} = 0.5(G_1(u, v) + G_2(u, v)). \quad (66)$$

Der endgültige Gütefaktor  $\gamma$ , der die Position des Objektes wichtet, ergibt sich aus dem Produkt von (65) und (66):

$$\gamma = \gamma^z \gamma^{uv}. \quad (67)$$

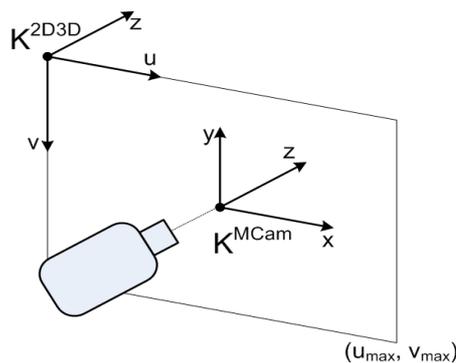
Zur Berechnung des Gütefaktors eines Objekts wird auf die vom zugehörigen Kalman-Filter erzeugte Positionsinformation, unter Rücktransformation der Position in das Koordinatensystem  $K^{3D}$ , zurückgegriffen.

### 5.2.6 ERZEUGUNG DES LOKALEN ZUSTANDSVEKTORS

Die Zusammensetzung des lokalen Zustandsvektors eines Multisensorknotens findet auf Grundlage der durch die Schritte der Klassifikation, Verfolgung sowie Positionswichtung gefundenen lokalen Objekte und deren erzeugten Informationen statt. Für einen beliebigen Zeitpunkt  $k$  gilt, dass der lokale Zustandsvektor des MSK  $n$

$$\underline{L}_k^n = [x_{1,k}^n, \gamma_{1,k}^n, \beta_{1,k}^n, \dots, x_{N_n,k}^n, \gamma_{N_n,k}^n, \beta_{N_n,k}^n, t_k^n]^T \quad (68)$$

sich zusammensetzt aus den Vektoren  $x_{j,k}^n$ ,  $j=1, \dots, N_n$ , die die Positionen und Geschwindigkeiten im lokalen Koordinatensystem<sup>80</sup>  $K_n^{MCam}$  enthalten (Abb. 78), den  $\gamma_{j,k}^n$  Positionswichtungen der Objekte sowie den binären Informationen  $\beta_{j,k}^n \in [0,1]$ , welche angeben, ob das Objekt zum Zeitpunkt  $k$  als Person klassifiziert worden ist. Die Variable  $t_k^n$  in (68) repräsentiert den Zeitstempel des Zeitpunktes  $k$ <sup>81</sup>.



*Abbildung 78: Zusammenhang zwischen den Koordinatensystemen  $K^{2D3D}$  und  $K^{MCam}$ .*

*Beide Koordinatensysteme besitzen übereinstimmende Tiefeninformationen.*

<sup>80</sup>Die Positions- und Geschwindigkeitsinformationen des Kalman-Filters werden aus dem gemeinsamen Koordinatensystem der Sensoren  $K^{2D3D}$  im Verlauf der Erzeugung des lokalen Zustandsvektors in das lokale Koordinatensystem  $K_n^{MCam}$  transformiert (16).

<sup>81</sup>Die Methode der Zeitsynchronisation des Systems wurde in Abschnitt 3.1.5 diskutiert.

## 5 Objekt- und Personenverfolgung

Die in (68) enthaltenen Informationen spiegeln damit die komplette Weltansicht des MSK  $n$  auf Grundlage seines lokalen Koordinatensystems  $K_n^{MCam}$  zum Zeitpunkt  $k$  wider und werden im Anschluss an ihrer Erzeugung an den zur Fusion der Daten zuständigen FSK gesendet.

### 5.2.7 ANALYSE DES ZEITBEDARFS ZUR ERZEUGUNG LOKALER ZUSTANDSVEKTOREN

Der zur Erstellung des lokalen Zustandsvektors (68) notwendige Zeitraum<sup>82</sup>, unter Variation der Anzahl von im Detektionsgebiet gleichzeitig vorhandenen Personen sowie Verwendung unterschiedlicher MultiCam-Informationen, liefert Abb. 79.

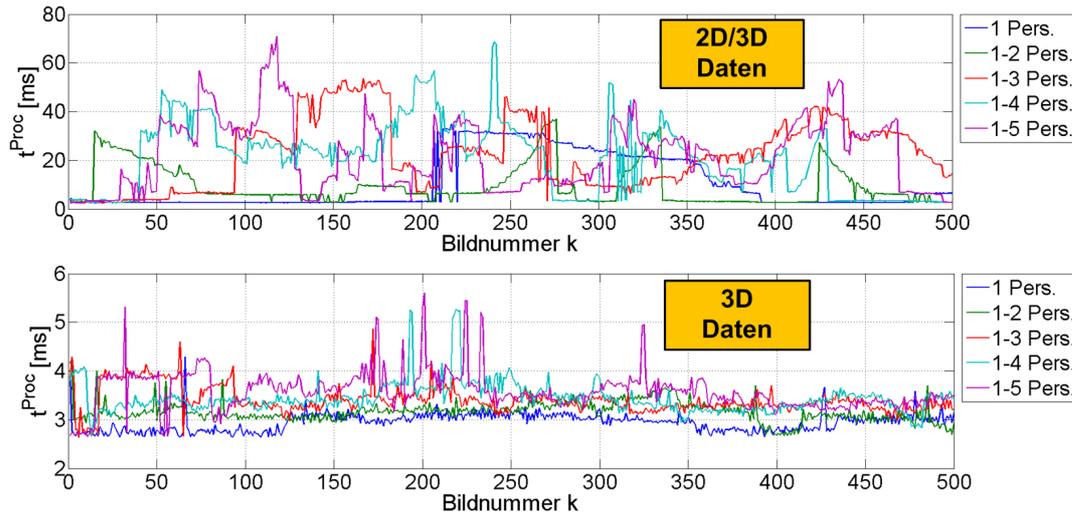


Abbildung 79: Berechnungsdauer der Erzeugung lokaler Zustandsvektoren eines Bildes. Analysesequenz, bestehend aus 500 Bildern. Oberer Graph: Berechnungsdauer unter Verwendung sämtlicher Informationen (2D/3D). Unterer Graph: Berechnungsdauer unter Verwendung von lediglich der 3D-Daten (Distanz, Modulationsamplitude, Intensität).

Unter Einsatz sämtlicher Sensordaten (2D, 3D) kann die maximale Dauer der Berechnung bei einer maximalen Anzahl von bis zu fünf gleichzeitig vorhandenen Personen im Sichtbereich auf ca. 70 ms angesetzt werden. Für Szenen bis zu drei Personen liegt die Grenze der Berechnungsdauer in etwa bei der maximalen Akquisitionszeit von  $t_{Max}^{Akq}$  von 50 ms und ermöglicht eine zur Akquisition parallelen Bearbeitung ohne eine Reduktion der durch die Hardware definierten Zeitauflösung. Zusätzliche Objekte führen zur Erzeugung einer naturgemäß vermehrten Anzahl von Suchfenstern (Abb. 64), welche dem HOG-Algorithmus zugeführt werden, woraus sich eine entsprechende Verlängerung der Berechnungszeit ergibt.

Für den Fall, dass ausschließlich die Daten des PMD-Sensors verwendet werden, liegt die notwendige Berechnungsdauer zur Erzeugung eines lokalen Zustandsvektors für alle betrachteten Szenen unter 6 ms. Die durch Hinzunahme zusätzlicher Personen im Sichtfeld bedingte Verlängerung der Berechnungszeit ist ausgesprochen moderat, sodass der verfolgte Ansatz auch zur Lokalisation und Verfolgung einer Vielzahl von Objekten verwendbar ist.

## 5.3 GLOBALE OBJEKT- UND PERSONENVERFOLGUNG (JDL-PROZESSEBENE 2)

Gemäß des verfolgten Ansatzes werden sämtliche Operationen, die der JDL-Prozessebene 2 zuzuordnen sind und welche im Rahmen dieser Arbeit als globale Operationen interpretiert werden, auf dem FSK durchgeführt (Abb. 29). Ziel der globalen Objekt- und Personenverfolgung ist es, mithilfe der vorhandenen Informationen der partizipierenden MSK eine Steigerung des

<sup>82</sup>Verwendet wurde ein Standard-PC (Core2Duo 3MHz).

## 5 Objekt- und Personenverfolgung

Informationsgehalts herbeizuführen. Der erste Schritt diesbezüglich besteht dabei in der Angleichung der lokal erzeugten Informationen (Abb. 25).

### 5.3.1 DATENANGLEICHUNG

Sei zunächst  $L_k = \{L_k^1, \dots, L_k^{N_L}\}$  die Menge der Zustandsvektoren der  $N_L$  MSK zum Zeitpunkt  $k$  (68). Für jedes Element aus  $L_k$  gilt, dass zwecks Angleichung der Daten seine auf dem zugehörigen lokalen Koordinatensystem  $K_n^{MCam}$  basierten Informationen, wie Position und Geschwindigkeit, in das globale Koordinatensystem  $K^{Glob}$  transformiert werden müssen.

Mit  $\rho^n \in \mathbb{R}^3$  als ein auf dem lokalen Koordinatensystem  $K_n^{MCam}$  basierender Vektor des MSK  $n$  erfolgt eine Transformation in das globale Koordinatensystem  $K^{Glob}$  durch eine Rotation mittels der Quaternionenmultiplikation

$${}^G p^n = f_n(p^n) = q_n p^n q_n^{-1} \quad (69)$$

mit  $q_n, q_n^{-1}, p^n, {}^G p^n \in H$  als Elemente der Menge der Hamilton-Zahlen<sup>83</sup> (Anhang E). Für die Komponente  $q_n$  gilt, dass sie im Vorfeld der Systemoperation zu berechnen ist<sup>84</sup>. Der ins globale Koordinatensystem transformierte Positionsvektor ist dabei in  ${}^G p^n = [0, {}^G \rho^n]^T$  enthalten<sup>85</sup>. Für das zu transformierende Quaternion gilt:  $p^n = [0, T_n(\rho^n)]^T$  mit  $T_n$  als dem Translationsoperator, welcher die Ursprünge der Koordinatensysteme  $K_n^{MCam}$  und  $K^{Glob}$  zur Deckung bringt<sup>86</sup>. Für jedes Element aus  $L_k$  ergibt sich seine Repräsentation nach seiner Transformation in  $K^{Glob}$  zu

$${}^G L_k^n = [{}^G x_{1,k}^n, \gamma_{1,k}^n, \beta_{1,k}^n, \dots, {}^G x_{N_n,k}^n, \gamma_{N_n,k}^n, \beta_{N_n,k}^n, t_k^n]^T, \quad n=1, \dots, N_L. \quad (70)$$

Die Positionswichtungen  $\gamma_{1,k}^n, \dots, \gamma_{N_n,k}^n$  und Indikatorvariablen  $\beta_{1,k}^n, \dots, \beta_{N_n,k}^n$  des ursprünglichen Zustandsvektors werden unverändert übernommen. Nach Beendigung der Transformation aller Größen enthält die Menge der transformierten Zustandsvektoren

$${}^G L_k = \{{}^G L_k^1, \dots, {}^G L_k^{N_L}\} \quad (71)$$

somit sämtliche verfügbaren Informationen bezogen auf das globale Koordinatensystem  $K^{Glob}$  zum Zeitpunkt  $k$  und bildet die Basisinformation für das globale Objektmanagement.

<sup>83</sup>Streng genommen handelt es sich im Falle der  $q_n$  aus (69) um Einheitsquaternionen (Versoren).

<sup>84</sup>Wie bereits in Abschnitt 1.2 diskutiert, wird angenommen, dass die zum globalen Koordinatensystem relativen Positionen und Drehungen der lokalen Systeme zu diesem Zeitpunkt vorhanden sind.

<sup>85</sup>Siehe dazu Anhang E.

<sup>86</sup>Analog zu  $q_n$  ist der Translationsoperator  $T_n$  ebenfalls im Vorfeld zu bestimmen.

## 5 Objekt- und Personenverfolgung

### 5.3.2 GLOBALES OBJEKTMANAGEMENT

Aufgabe des globalen Objektmanagements ist es, auf Basis der durch die MSK erzeugten, lokalen Zustandsinformationen (71) sogenannte globale Objekte zu erzeugen, zu erhalten sowie zu entfernen. Der wesentliche Unterschied zum lokalen Objektmanagement besteht darin, dass globale Objekte, im Vergleich zu den lokal durch die MSK erzeugten Objekte, keinen Alterungsprozess besitzen. Dieses führt dazu, dass globale Objekte zu jedem Zeitpunkt eine Erhaltung durch lokale Objekte benötigen. Findet diese nicht statt, werden sie entfernt.

Als globale Objekte werden Einheiten bezeichnet, welche auf Basis der Positionsinformationen lokaler MSK-Objekte im globalen Koordinatensystem  $K^{Glob}$  erzeugt werden und einer Verfolgung unterliegen. Örtlich eng benachbarte oder sich teilweise überlappende MSK-Objekte unterschiedlicher MSK entsprechen dabei den Konstituenten eines globalen Objekts und werden in einer zeitlich variierenden Konstituentenmenge zusammengefasst. Die Berechnung von Abständen zwischen zwei MSK-Objekten unterschiedlicher MSK zum Zeitpunkt  $k$  im Koordinatensystem  $K^{Glob}$  erfolgt dabei mittels

$$\lambda_{i,j,k}^{n,m} = \left\| \rho_{i,k}^n - \rho_{j,k}^m \right\|_2, \quad n \neq m. \quad (72)$$

Während die Indizes  $n, m$  in (72) die eindeutigen Identitäten der MSK bestimmen, definieren die Indizes  $i, j$  jeweils die eindeutige Identität des lokalen Objekts auf dem entsprechenden MSK. Die Vektoren  $\rho_{i,k}^n$  und  $\rho_{j,k}^m$  enthalten die Positionen der lokalen Objekte in der  $x-z$ -Ebene von  $K^{Glob}$ . Eine Annahme, welche im Rahmen der globalen Objektverfolgung getroffen wird, ist, dass die Bewegungen der Objekte in einer Ebene verlaufen. Ein Objekt wird dabei als ein sich in der  $x-z$ -Ebene des globalen Koordinatensystems  $K^{Glob}$  bewegender Zylinder mit parametrisierbarem Radius  $\Delta^{max}$  und Höhe interpretiert. Die Position in  $y$ -Richtung ergibt sich aus dem Mittel der entsprechenden Werte der Konstituenten.

Im Falle von  $N_L$  MSK existieren maximal  $N_L$  unterschiedliche Repräsentationen bzw. Konstituenten, da davon ausgegangen wird, dass jedes reale Objekt in einem MSK durch lediglich ein lokales Objekt repräsentiert wird. Für ein globales Objekt  $\Omega_{a,k}$  zum Zeitpunkt  $k$  gilt, dass es durch den Tupel

$${}^G\Omega_{a,k} = \{ {}^G\underline{\Xi}_{a,k}, {}^G O_{a,k}, {}^G\beta_a, \Pi_a \} \quad (73)$$

dargestellt wird.  ${}^G O_{a,k}$  entspricht der Konstituentenmenge des globalen Objekts. Die Variable  ${}^G\beta_a \in \mathbb{N}$  enthält den ersten Zeitindex, in dem konstituierende MSK-Objekte lokal als Person klassifiziert worden sind<sup>87</sup>. Die Position des globalen Objekts in der  $x-z$ -Ebene von  $K^{Glob}$  enthält der Vektor  ${}^G\underline{\Xi}_{a,k}$ . Berechnet wird diese mittels des Partikelfilters  $\Pi_a$ , welcher dem globalen Objekt bei seiner Erzeugung zugewiesen wird. Eine ausführliche Beschreibung des Prozesses der Erzeugung, Erhaltung sowie Entfernung globaler Objekte wird in Anhang B gegeben.

Aus der Definition eines globalen Objekts ergibt sich, dass sein Aufenthaltsort aus den Positionsinformationen einer zeitlich variierenden Anzahl  $|{}^G O_{a,k}|$  von Konstituenten berechnet wird. Dieses entspricht einer Informationsgewinnung aus einer ebenfalls in ihrer Modenanzahl variierenden multimodalen Verteilungsdichte.

---

<sup>87</sup>Initialisiert wird  ${}^G\beta_a$  mit dem Standardwert 0.

## 5 Objekt- und Personenverfolgung

Der im Rahmen dieser Arbeit verfolgte Ansatz zur Berechnung der Position globaler Objekte auf Grundlage einer solchen Verteilungsdichte basiert auf der Verwendung des CONDENSATION-Algorithmus, dessen Ablauf im Folgenden kompakt dargestellt wird.

### 5.3.3 CONDENSATION-ALGORITHMUS

Im Allgemeinen führen nichtlinearer System- und Messmodelle (47, 48) dazu, dass die Prädiktions- und Aktualisierungsgleichungen (49, 50) nicht mehr lösbar sind [183], sodass auf im bayes'schen Sinne nähernde Algorithmen ausgewichen werden muss. Beispiele hierfür sind das Erweiterte-Kalman-Filter (EKF) oder das Unscented-Kalman-Filter (UKF), welche eine Linearisierung der zugrunde liegenden Modelle analytisch (EKF) oder statistisch (UKF) durchführen [182]-[184]. Grundsätzlich bleibt jedoch bei Verwendung des Kalman-Filters die Festlegung auf monomodale, normale Verteilungsdichten als Beschreibungen des Systemzustands erhalten.

Einen Ansatz zur Behandlung multimodaler Verteilungsdichten, bei gleichzeitiger Unabhängigkeit der Form der verwendeten System- und Messmodelle [194], [195], bietet der zur Familie der sequenziellen Monte-Carlo-Methoden [183] zugehörige CONDENSATION-Algorithmus. Die getroffene Grundannahme ist, dass sich jede, einen Systemzustand  $\mathbf{x}_k$  repräsentierende Dichte  $f_{\mathbf{x}_k|Y_k}(\xi_k|Z_k)$ , durch einen Satz von Partikeln bzw. Hypothesen  $s_k = \{\mathbf{s}_k^{(n)}: n=1, \dots, N\}$  mit zugehörigen Partikel- bzw. Hypothesengewichte aus der Menge  $\pi_k = \{\pi_k^{(n)}: n=1, \dots, N\}$  mit  $\sum_{n=1}^N \pi_k^{(n)} = 1$  durch

$$f_{\mathbf{x}_k|Y_k}(\xi_k|Z_k) \approx \sum_{n=1}^N \pi_k^{(n)} \delta(\mathbf{x}_k - \mathbf{s}_k^{(n)}) \quad (74)$$

nähern lässt [183], [184]. Eine Erhöhung der Näherungsgenauigkeit durch (74) ergibt sich mittels Steigerung der Hypothesenanzahl, sodass für den Grenzwert  $N \rightarrow \infty$  die Dichte  $f_{\mathbf{x}_k|Y_k}(\xi_k|Z_k)$  durch den Tupel  $\{s_{k-1}, \pi_{k-1}\}$  exakt beschreibbar ist [183], [184].

#### 5.3.3.1 ESTIMATIONSPROZESS

Die Bestimmung der Wahrscheinlichkeitsdichte  $f_{\mathbf{x}_k|Y_k}(\xi_k|Z_k)$  des Systemzustands eines Zeitpunkts  $k$  durch den CONDENSATION-Algorithmus erfolgt iterativ. Gegeben sei zunächst die Wahrscheinlichkeitsdichte des vorausgegangenen Zeitpunkts<sup>88</sup>  $f_{\mathbf{x}_{k-1}|Y_{k-1}}(\xi_{k-1}|Z_{k-1})$ , dargestellt durch den Tupel  $\{s_{k-1}, \pi_{k-1}\}$ . Als erster Schritt wird aus der Menge  $s_{k-1}$  durch zufällige Entnahme mit Rücksetzen eine neue Menge  $s_k'$  gebildet, wobei die Anzahl  $N_{\Pi}$  der Partikel dabei konstant gehalten wird. Die Entnahmewahrscheinlichkeit eines Partikels  $\mathbf{s}_{k-1}^{(n)}$  aus  $s_{k-1}$  wird dabei durch sein zugehöriges Gewicht  $\pi_{k-1}^{(n)}$  bestimmt.

Für die Partikelgewichte der neu gebildeten Menge  $s_k'$ , welche in  $\pi_k'$  sind, gilt, dass sie zunächst alle gleich der reziproken Partikelanzahl gesetzt werden. Anschließend erfahren alle Partikel aus  $s_k'$  eine durch das zugrunde liegende Systemmodell  $f_{\mathbf{x}_k|X_{k-1}, Y_{k-1}}(\xi_k|\Xi_{k-1}, Z_{k-1})$  festgelegte Änderung und bilden die Partikelmenge  $s_k$ . Das zu diesem Zeitpunkt vorhandene Tupel  $\{s_k, \pi_k'\}$  entspricht dabei der Prädiktion des Systemzustands  $f_{\mathbf{x}_k|Y_{k-1}}(\xi_k|Z_{k-1})$  (*dynamic prior*) [183].

<sup>88</sup>Der Anfangszustand  $f_{\mathbf{x}_0|Y_0}(\xi_0|Z_0) = f_{\mathbf{x}_0}(\xi_0)$  bzw. die Elemente der Mengen  $s_0, \pi_0$  werden dabei entweder anhand vorhandenen Vorwissens oder zufällig initialisiert. Für den Fall einer zufälligen Initialisierung der Partikel werden alle Partikelgewichte üblicherweise gleich der reziproken Partikelanzahl gesetzt.

## 5 Objekt- und Personenverfolgung

Im abschließenden Schritt werden mittels der Observationsdichte  $f_{\underline{y}_k|\underline{y}_k}(\underline{\xi}_k|\underline{\xi}_k)$  unter Anwendung des Observationsschrittes des faktorisierten Entnehmens (*factored sampling*) [196]-[198] die Gewichte der Hypothesen aus  $s_k$  bestimmt wobei sich eine neue Menge  $\pi_k$  ergibt. In Verbindung mit der Partikelmenge  $s_k$  repräsentiert das Tupel  $\{s_k, \pi_k\}$  unter der Annahme (74) damit der gesuchten posterioren Dichte  $f_{\underline{y}_k|Y_k}(\underline{\xi}_k|Z_k)$ . Für eine detailliertere Einführung in die Funktionsweise des CONDENSATION-Algorithmus sei der interessierte Leser auf die Publikationen von Isard und Blake [192]-[195] sowie auf Anhang D verwiesen.

### 5.3.4 ESTIMATION DER OBJEKTZUSTÄNDE

Der verwendete Ansatz zur Verfolgung globaler Objekte basiert auf der Verwendung des CONDENSATION-Algorithmus mit diffusionsgetriebenem Systemmodell ohne deterministischen Drift. Die Berechnung des dynamischen Priors ergibt sich damit durch

$$\underline{s}_k^{(n)} = A \underline{s}_{k-1}^{(n)} + B \underline{v}_{k-1}^{(n)} = \underline{s}_{k-1}^{(n)} + B \underline{v}_{k-1}^{(n)} \quad \text{mit} \quad B = \begin{pmatrix} \sigma_{\Pi} & 0 \\ 0 & \sigma_{\Pi} \end{pmatrix}. \quad (75)$$

Die Partikel des Filters  $\underline{s}_k^{(n)} \in \mathbb{R}^2$ ,  $n=1, \dots, N_{\Pi}$ , entsprechen Positionen in der  $x-z$ -Ebene des globalen Koordinatensystems. Die Komponenten des Vektors  $\underline{v}_{k-1}^{(n)}$  entstammen einer nullzentrierten Gleichverteilung aus dem Intervall  $[-1,1]$ . Der Wert von  $\sigma_{\Pi}$  gilt für sämtliche den globalen Objekten zugeordneten Partikelfiltern. Durch die Verwendung der isotropen Diffusion besteht eine grundsätzliche Unabhängigkeit bzgl. des Bewegungsmusters des zu verfolgenden Objekts. Weiterhin bietet die geringe Komplexität von (75) die Möglichkeit einer extrem schnellen Berechnung.

Für die Messungen gelten, dass sie durch die jeweiligen Konstituenten globaler Objekte bestimmt werden. Sei  ${}^G O_{i,k}$  die Konstituentenmenge eines beliebigen globalen Objekts  ${}^G \Omega_{i,k}$  zum Zeitpunkt  $k$ . Der zugehörige Messvektor von  ${}^G \Omega_{i,k}$  ist definiert als

$${}^G \underline{y}_{i,k} = [{}^G \underline{x}_{1,k}, \underline{y}_{1,k}, \dots, {}^G \underline{x}_{|{}^G O_{i,k}|,k}, \underline{y}_{N_{i,k},k}]^T. \quad (76)$$

${}^G \underline{y}_{i,k}$  ist eine Zusammensetzung der Positions- und Geschwindigkeitsinformationen der konstituierenden MSK-Objekte sowie ihrer lokal auf den jeweiligen Multisensorknoten bestimmten Positionswichtungen<sup>89</sup>.

#### 5.3.4.1 FILTERINITIALISIERUNG

Die Initialisierung des dem Objekt bei dessen Erzeugung zugewiesenen Partikelfilters wird ebenfalls zum Zeitpunkt der Objekterzeugung durchgeführt. Sei zunächst  ${}^G \underline{\Xi}_{i,k}$  der initiale Positionsvektor eines globalen Objekts  ${}^G \Omega_{i,k}$ , welcher mittels

$${}^G \underline{\Xi}_{i,k} = H \left( \frac{1}{\underline{y}_{i,k}} \sum_{j=1}^{|{}^G O_{i,k}|} \underline{y}_{j,k} {}^G \underline{x}_{j,k} \right) \quad \text{mit} \quad \underline{y}_{i,k} = \sum_{j=1}^{|{}^G O_{i,k}|} \underline{y}_{j,k}, \quad H = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix} \quad (77)$$

<sup>89</sup>Zu beachten ist, dass die Indizierung der Elemente in (76) im Vergleich zu denen in (70) unterschiedlich ist. Die Indizierung in (76) basiert auf der vorhandenen Anzahl der Konstituenten, von denen angenommen wird, dass sie von unterschiedlichen MSK stammen.

## 5 Objekt- und Personenverfolgung

berechnet wird. Für alle  $N_\Pi$  Partikel des Filters  $\Pi_i$  von  ${}^G\Omega_{i,k}$  gilt zum Zeitpunkt der Initialisierung, dass sie durch

$$\underline{s}_{i,k}^{(n)} = {}^G\underline{\Xi}_{i,k} + B \underline{y}_k^{(n)} \quad (78)$$

mittels der stochastischen Diffusion  $B \underline{y}_k^{(n)}$  eine Verteilung um die initiale Position  ${}^G\underline{\Xi}_{i,k}$  des Objekts  ${}^G\Omega_{i,k}$  in der  $x-z$ -Ebene von  $K^{Glob}$  erfahren. Die anfänglichen Partikelgewichte entsprechen dem Kehrwert der Partikelanzahl:  $\pi_{i,k}^{(n)} = 1/N_\Pi$ . Die zum Zeitpunkt der Initialisierung bestimmten Partikel und Gewichte, welche in dem Tupel  $\{s_{i,k}, \pi_{i,k}\}$  zusammengefasst werden können, repräsentieren damit den Anfangszustand  $f_{x_0|y_0}(\underline{\xi}_0 | \underline{Z}_0) = f_{x_0}(\underline{\xi}_0)$  des Partikelfilters.

### 5.3.4.2 OBJEKTVERFOLGUNG

Im Falle der Objekterhaltung wird die Position des globalen Objekts durch das zugehörige Partikelfilter bestimmt. Sei dazu  ${}^G\Omega_{i,k}$  ein globales Objekt, welches aufgrund der Informationen des momentanen Zeitpunkts  $k$  eine Auffrischung seiner Konstituentenmenge  ${}^G\mathcal{O}_{i,k}$  erhalten hat (Anhang B). Die Neuberechnung der Partikelgewichte basiert auf der Position des zugehörigen Partikels sowie den Informationen des in (76) definierten Messvektors  ${}^G\underline{y}_{i,k}$ :

$$\pi_{i,k}^{(n)} = f(\underline{s}_{i,k}^{(n)}, {}^G\underline{y}_{i,k}, \Phi) = \left( \frac{1}{\gamma_{i,k}} \sum_{m=1}^{|\mathcal{O}_{i,k}|} \gamma_{m,k} \left\| \underline{s}_{i,k}^{(n)} - H {}^G\underline{x}_{m,k} \right\|_2 \right)^{-\frac{1}{\Phi}} \quad \text{mit } \gamma_{i,k} = \sum_{m=1}^{|\mathcal{O}_{i,k}|} \gamma_{m,k}. \quad (79)$$

Der Parameter  $\Phi$  dient der Variation des Gradienten von (79). Wichtig ist an dieser Stelle anzumerken, dass die Partikel des Filters zum Zeitpunkt der Neubestimmung der Gewichte bereits eine stochastische Diffusion erfahren haben. Die Berechnung der Position von  ${}^G\Omega_{i,k}$  ergibt sich im Anschluss mittels:

$${}^G\underline{\Xi}_{i,k} = \hat{\pi}_{i,k}^{-1} \sum_{q=1}^{N_\Pi} \pi_{i,k}^{(q)} \underline{s}_{i,k}^{(q)} \quad \text{mit } \hat{\pi}_{i,k} = \sum_{q=1}^{N_\Pi} \pi_{i,k}^{(q)}. \quad (80)$$

### 5.3.5 ANALYSE DER FILTERDYNAMIK

Die Partikeldynamik des Filters wird aufgrund der Wahl des Systemmodells (75) durch zwei Faktoren bestimmt. Zunächst ergibt sich durch die stochastische Diffusion eine Variation der Partikelpositionen, was zu einer Vergrößerung des Gebietes führt, in welchem das zu verfolgende Objekt einen nennenswerten Einfluss auf die Partikelmenge ausübt.

Durch die Neubestimmung der Partikelgewichte wird ein Bewegungszwang auf die Menge der Partikel ausgeübt, welcher von dem Gradienten der gewählten Funktion (79) abhängt. Ein zu geringer Gradient führt zu einem unzureichenden Bewegungszwang der Partikelmenge in Richtung des zu verfolgenden Objekts, was selbst im Falle relativ großer Diffusionen eine hohe Distanz zum Zielobjekt erzeugt und letztendlich zum Verlust dessen führt.

Durch eine hinreichend große Wahl von  $\sigma_\Pi$  sowie einer Gewichtsauffrischungsfunktion (79) mit ausreichend hohem Gradienten wird der Filter in die Lage versetzt, selbst Objekte mit beliebig

## 5 Objekt- und Personenverfolgung

pathologischen Bewegungsmustern<sup>90</sup> erfolgreich zu verfolgen. Untersucht wird im Folgenden die Verfolgungsleistung des Partikelfilters in Bezug auf eine Szene mit zwei sich bewegenden Personen (Abb. 80).



Abbildung 80: Ausgewählte Momentaufnahmen der untersuchten Szene.

Videosequenz bestehend aus 350 Bildern. Obere Zeile: Aufnahmen aus der Sicht von MultiCam 1. Untere Zeile: Perspektive von MultiCam 2. Erste Spalte: Globales Koordinatensystem  $K^{Glob}$  aus Sicht der beiden verwendeten Multisensorsysteme  $K_1^{MCam}$  und  $K_2^{MCam}$ . Die in den Graphen der Abb. 81, 82, 83 und 84 enthaltenen Informationen beziehen sich auf die von Obj. 1 generierte Trajektorie in  $K^{Glob}$ .

Abb. 81 enthält die Ergebnisse der globalen Verfolgung auf Basis des Parametersatzes  $\sigma_{\Pi}=10$  und  $\Phi=2$ . Die Wahl eines zu geringen Diffusionswertes zum Einen sowie einer Neuberechnung der Partikelgewichte zum Anderen resultiert in einer zu großen Filterträgheit, die schnell zu wachsenden Entfernungen zwischen dem globalen Objekt und seinen Konstituenten führt und die Verfolgung ineffektiv werden lässt. Die Entfernungen des globalen Objekts zu den Konstituenten korrespondieren ebenfalls mit der Positionsvarianz der Filterpartikel.

Eine große Distanz aller Partikel zu den MSK-Objekten mündet in einer tendenziellen Gleichverteilung der Partikelgewichte und führt dazu, dass mehr Partikel eine längere Überlebensdauer besitzen. Aufgrund seiner Isotropie der Diffusion breiten sich die Partikel in der Ebene aus, wobei die Geschwindigkeit der Ausbreitung durch  $\sigma_{\Pi}$  bestimmt wird.

Abb. 82 gibt die Informationen auf Basis des Parametersatzes  $\sigma_{\Pi}=10$  und  $\Phi=1$  wieder. Durch die Verringerung von  $\Phi$  ergibt sich ein steileres Abfallen der Partikelgewichte in Abhängigkeit ihrer Distanz zum Zielobjekt und führt dazu, dass der Bewegungsdruck auf die Partikelmenge im Falle kürzerer Entfernungen zu den Konstituenten stärker wird. Erkennbar wird der Effekt der Verringerung von  $\Phi$  an der im Vergleich zum vorigen Beispiel verringerten Distanz zu den Konstituenten sowie reduzierter Positionsvarianz der Filterpartikel. Weiterhin führt die Erhöhung des Bewegungszwangs dazu, dass im Falle einer geringer werdenden Geschwindigkeit des Zielobjekts das Filter in der Lage ist, aufzuholen<sup>91</sup>.

Die Ergebnisse einer erhöhten Diffusion mittels der Verwendung des Parametersatzes  $\sigma_{\Pi}=15$  und  $\Phi=2$  münden in dem in Abb. 83 dargestellten Systemverhalten. Erkennbar ist die gesteigerte Verfolgungsleistung des Partikelfilters. Eine Eigenart dieser Parametrierung ist, dass die Verfolgungsgeschwindigkeit des Filters im Falle eines sich bewegenden globalen Objekts leicht oszilliert (s. Abb. 83 untere Zeile rechts, Distanzgraph). Erklärt wird dieses durch das Zusammenspiel der Partikeldiffusion sowie der Amplitude der Partikelgewichte. Während eine zu schnelle Bewegung des zu verfolgenden Objekts zu einer anwachsenden Distanz führt, kann durch die erhöhte Diffusion die Reichweite des Filters vergrößert werden, was wiederum den Einfluss auf den durch (79) bestimmten Bewegungszwang erhöht. Reduziert sich die Varianz der Partikelpositionen aufgrund der Bündelung der Gewichte durch (79), wobei sich das zu verfolgende Objekt weiterhin gemäß seiner anfänglichen Geschwindigkeit bewegt, ergibt sich ein

<sup>90</sup>Z. B. häufiges und abruptes Ändern der Bewegungsrichtung.

<sup>91</sup>Man betrachte diesbezüglich die dritte Spalte in Abb. 80. In dieser Phase der Szene befand sich das zu verfolgende, reale Objekt vergleichsweise lange auf gleicher Position bzw. hatte eine geringe Bewegungsgeschwindigkeit.

## 5 Objekt- und Personenverfolgung

erneutes Anwachsen der Distanz, was wiederum zu einer Erhöhung der Reichweite des Filters und dem oben beschriebenen Verhalten führt. Im Falle einer relativ gleichmäßigen Personenbewegung ergibt sich somit eine oszillierende Geschwindigkeit des globalen Objekts und bedingt eine, im Vergleich zum vorigen Fall, verbesserte Verfolgungsleistung des Filters.

Eine Verbesserung dessen ist in Abb. 84 durch die Verwendung des Parametersatzes  $\sigma_{\Pi}=15$  und  $\Phi=1$  wiedergegeben. Durch Vergrößerung sowohl der Partikeldiffusion als auch der Erhöhung des Gradienten von (79), was besonders in geringer Nähe des globalen Objekts zu seinen Konstituenten zu einem hohen Bewegungsdruck auf die Partikelmenge führt, kann eine ausreichende Verfolgungsleistung des Partikelfilters sichergestellt werden. Die durch die verwendeten Parameter induzierte Dynamik des Filters führt zu einer permanenten Verfolgung der Zielobjekte unter Beibehaltung einer geringen Distanz und einer dem Zielobjekt angepassten Geschwindigkeit.

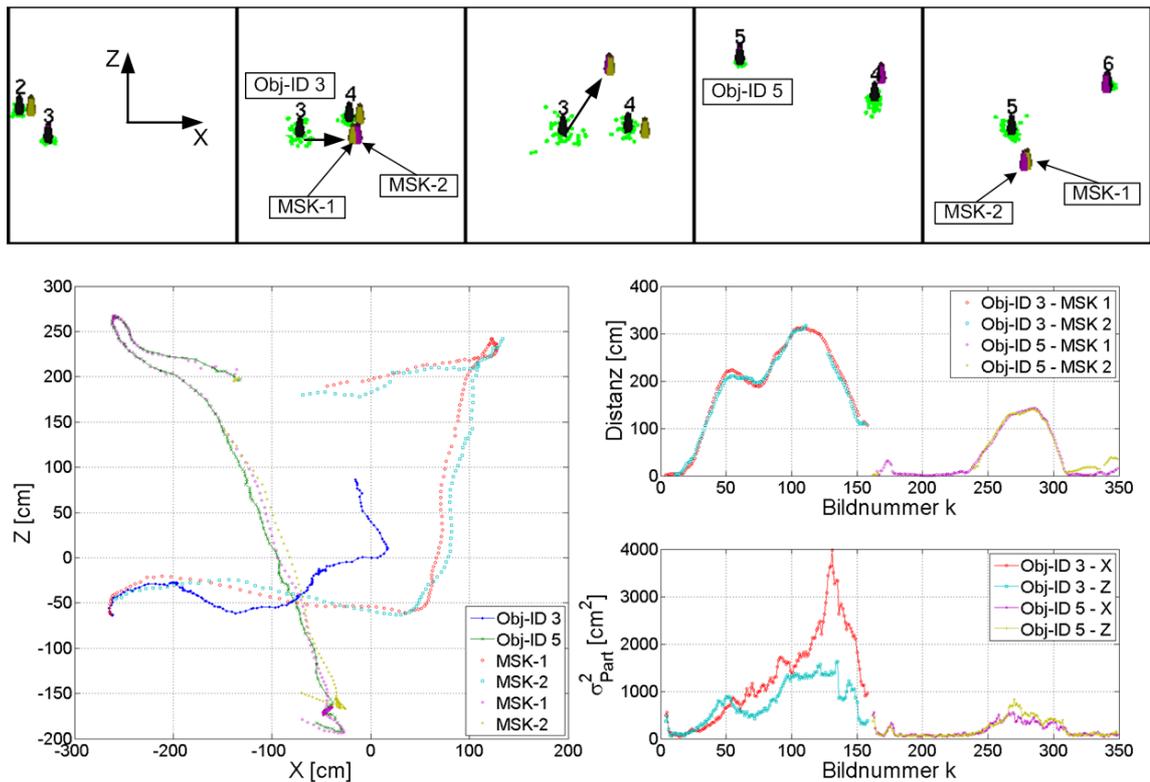


Abbildung 81: Objektverfolgung CONDENSATION-Algorithmus I.

*Parametersatz:  $\sigma_{\Pi}=10$ ,  $\Phi=2$ . Aufgrund des Bewegungsmusters und den damit entstandenen Verdeckungen des realen Objekts (Obj. 1, Abb. 80) wurden in der betrachteten Szene mehrere zugehörige, globale Objekte generiert (Obj-ID 3,5). Die Erzeugung mehrerer globaler Objekte, welche demselben realen Objekt zugeordnet werden können, ist durch die Diskontinuität der Daten beider MSKs bedingt. Obere Zeile: Die Unterkanten der schwarzen, mit einer Objekt-ID versehenen Kegel, entsprechen den durch den Partikelfilter berechneten Positionen globaler Objekte in  $K^{Glob}$ . Die gelben und violetten Kegel repräsentieren jeweils die Positionen der durch die beiden verwendeten Multisensorsysteme (MSK-1, MSK-2) lokal erzeugten MSK-Objekte. Die Partikelpositionen sind durch die grünen Punkte repräsentiert. Die Bilder dieser Zeile korrespondieren mit denen aus Abb. 80. Untere Zeile, linker Graph: Trajektorien der generierten globalen Objekte (Obj-ID 3,5) sowie die von den Multisensorknoten übermittelten Positionsinformationen. Die Nähe der Trajektorie von Obj-ID 5 zu den Positionen von MSK-1 am Ende der Szene ist bedingt durch die hohe Positionswichtung seiner Daten, da sich in diesem Fall das zu verfolgende reale Objekt in relativ kurzer Distanz zu ihm befindet (s. Spalte fünf, Abb. 80). Untere Zeile, rechts oben: Die Graphen zeigen jeweils die Distanzen der Konstituenten zu den zugehörigen globalen Objekten. Untere Zeile, rechts unten: Positionsvarianz der Filterpartikel jeweils in x- und z-Richtung.*

## 5 Objekt- und Personenverfolgung

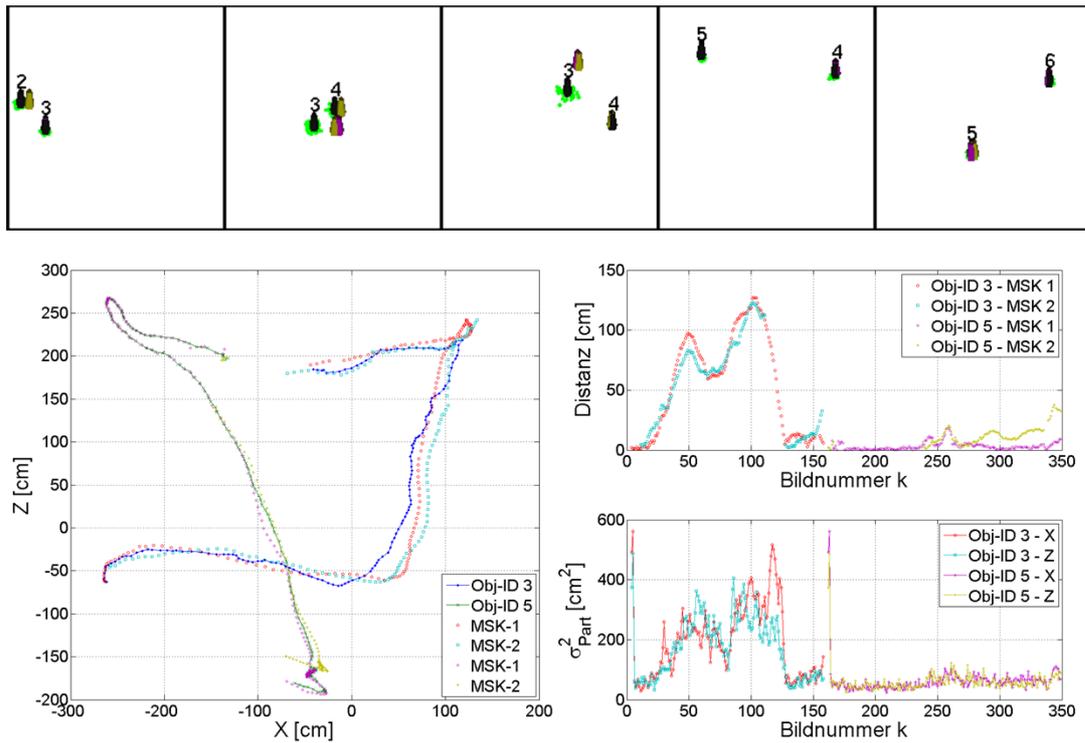


Abbildung 82: Objektverfolgung CONDENSATION-Algorithmus II.

Verwendeter Parametersatz:  $\sigma_{\Pi}=10$ ,  $\Phi=1$ .

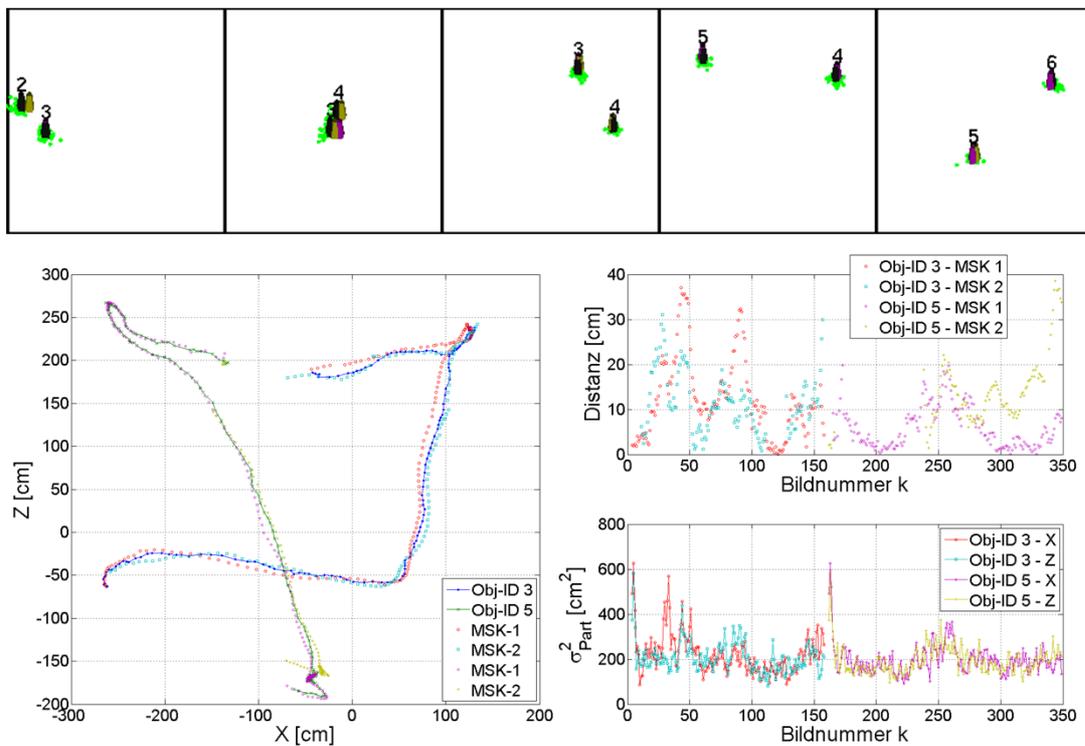


Abbildung 83: Objektverfolgung CONDENSATION-Algorithmus III.

Verwendeter Parametersatz:  $\sigma_{\Pi}=15$ ,  $\Phi=2$ .

## 5 Objekt- und Personenverfolgung

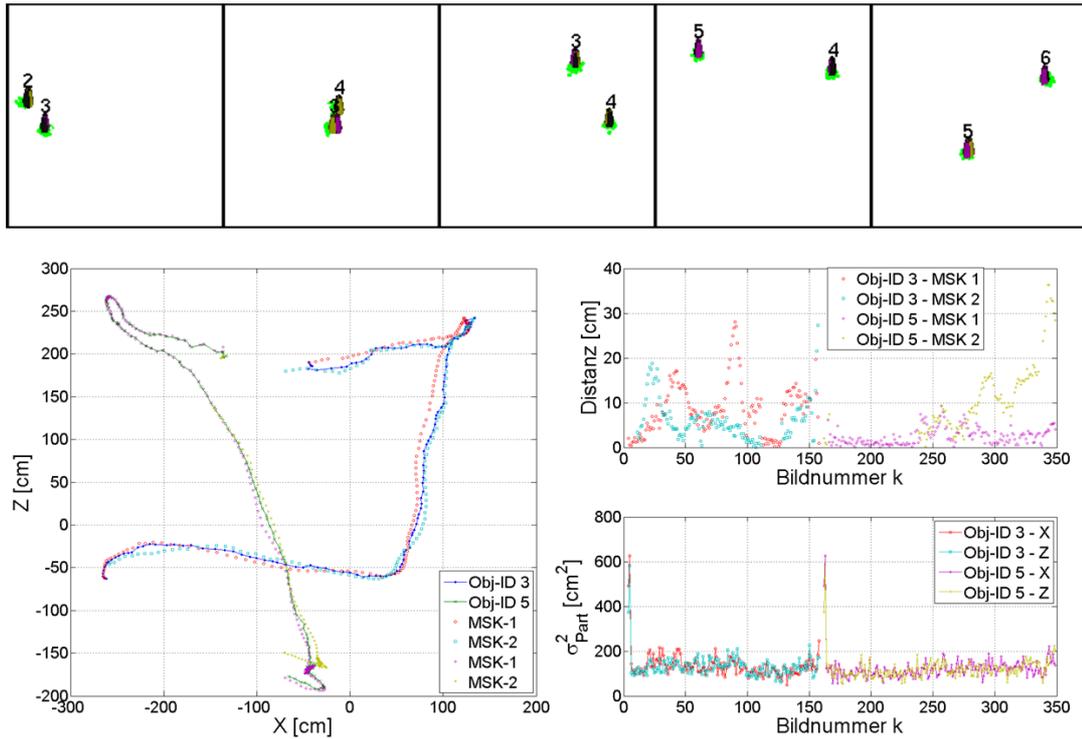


Abbildung 84: Objektverfolgung CONDENSATION-Algorithmus IV.

Verwendeter Parametersatz:  $\sigma_{\Pi}=15$ ,  $\Phi=1$ .

Die in den Bewegungstrajektorien der generierten Objekte (Abb. 81-84, linker unterer Graph) erkennbare Bevorzugung mancher Konstituenten, erkennbar an einer entsprechend geringeren Distanz, ist dadurch zu erklären, dass durch die Positionswichtung der einzelnen MSK-Objekte ein direkter Einfluss auf die Neuberechnung der Partikelgewichte durch (79) ausgeübt wird, sodass Objekte mit höheren Wichtungen bevorzugt werden.

### 5.4 STEIGERUNG DES INFORMATIONSGEHALTS

Der Abschluss dieses Kapitels widmet sich der Diskussion des mittels der Verwendung des im Rahmen dieser Arbeit implementierten Systems erreichten Steigerung des Informationsgehalts unter Verwendung mehrerer Perspektiven mit partiell überlappendem Sichtbereich.

Betrachtet wird zunächst die Erweiterung des Sicht- und Detektionsbereichs durch die Hinzunahme mehrerer MSK. Anzumerken ist, dass sich der im Kontext einer durchzuführenden Objektverfolgung ergebende Mehrwert jedoch erst durch eine zeitliche Synchronisation der vorhandenen Teilinformationen ausnutzen lässt.

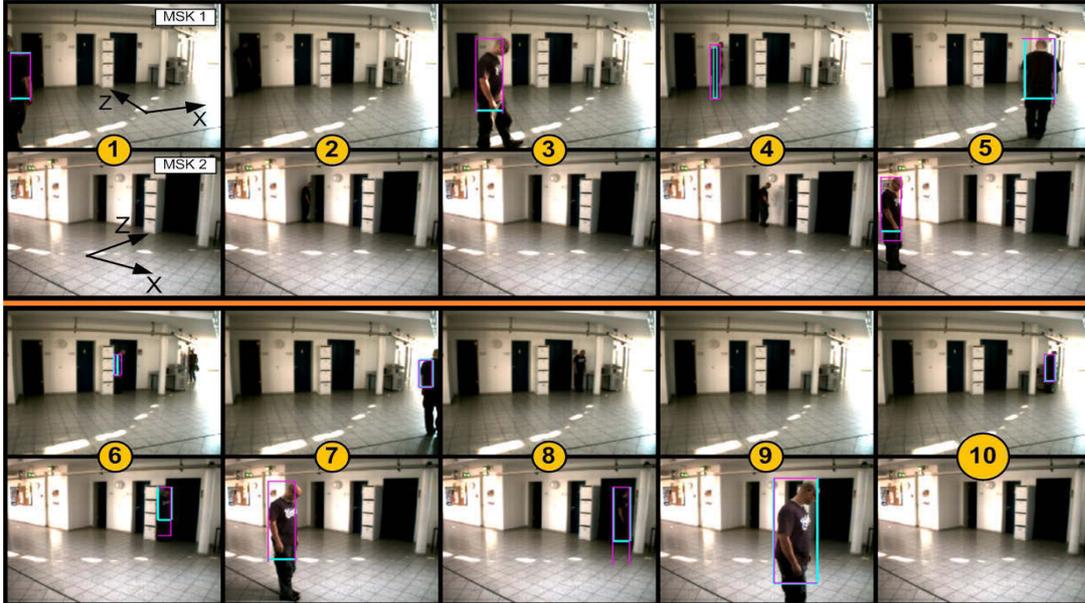
Im Anschluss dazu erfolgt eine Analyse der erreichten Reduktion gegenseitiger Objektverdeckungen. Eine Verringerung der Okklusionen ist besonders erstrebenswert, da sich dadurch Objekte naturgemäß über längere Zeiträume verfolgen lassen. Dieses führt letztendlich dazu, dass die entsprechenden Objekttrajektorien einen erhöhten Informationsgehalt vorweisen können, welcher sich in einem eventuell nachgeschalteten Analyseschritt ausnutzen lässt.

#### 5.4.1 ERWEITERUNG DES DETEKTIONSBEREICHS

Als Grundlage der Verifikation der Bereichserweiterung durch den Einsatz temporal synchronisierter MSK dient eine ausgewählte Szene mit definierter Objekttrajektorie (Abb. 86,

## 5 Objekt- und Personenverfolgung

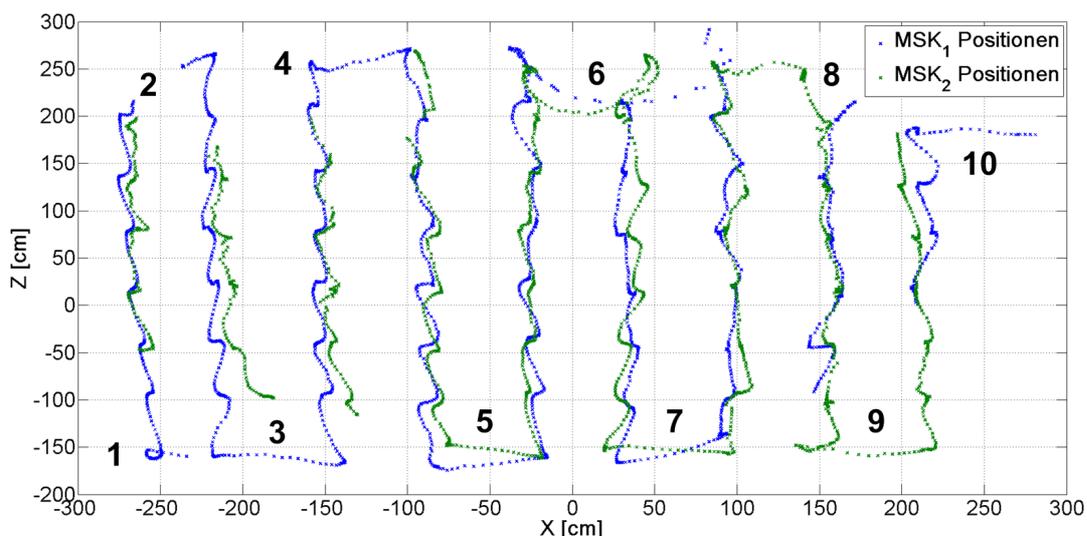
87). In Abb. 85 sind ausgewählte Momente der untersuchten Szene, welche die unterschiedlichen Aspekte des Systemverhaltens beleuchten, wiedergegeben worden. Jedes Bildpaar beinhaltet dabei die Perspektiven beider MSK. Die Nummerierung der Bildpaare findet sich wieder in den Abb. 86 und 88, wo sie den Positionen des realen Objekts sowie den entsprechenden Zeitpunkten entsprechen.



*Abbildung 85: Ausgewählte Aufnahmen im Rahmen der Überprüfung des erweiterten Detektionsbereichs.*

*Zwecks Verdeutlichung unterschiedlicher Aspekte der Bereichserweiterung sind ausgewählte Momente der untersuchten Szene abgebildet worden. Die Bewegung der Testperson entspricht dabei einem schrittweisen Ablauf eines vordefinierten Weges im Detektionsbereich des Komplettsystems (Abb. 86, 87).*

Aus den Bildpaaren 1, 3, 9 sowie 10 ist die direkte Erweiterung des Detektionsbereichs durch komplementäre Fusion der Daten ersichtlich (Abb. 86, 87).



*Abbildung 86: Positionen der durch die MSK erzeugten lokalen Objekte.*

*Die Position der Zahlen entspricht in etwa der Position des realen Objektes im globalen Koordinatensystem im zugehörigen Bildpaar von Abb. 85.*

## 5 Objekt- und Personenverfolgung

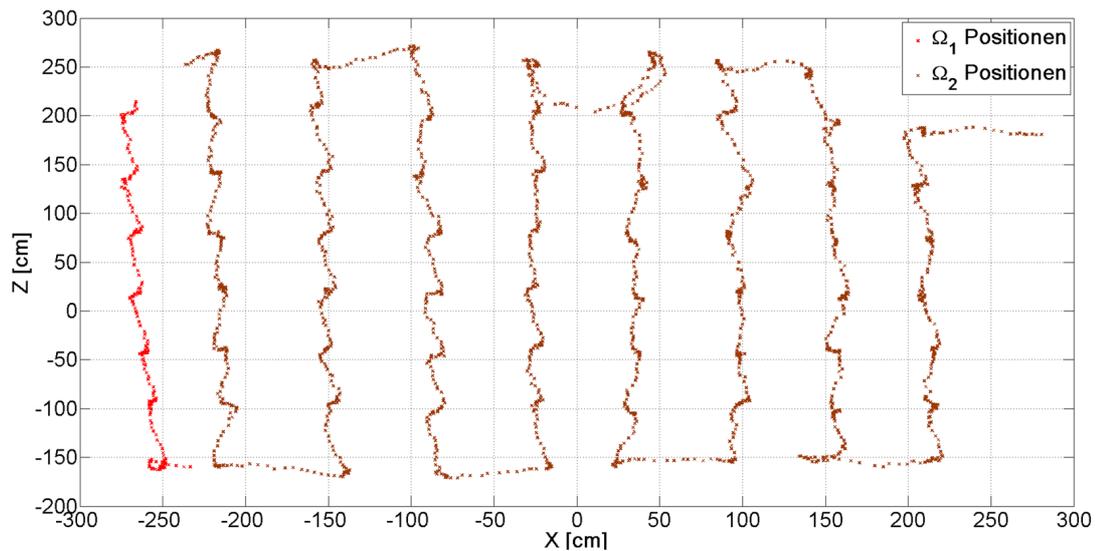


Abbildung 87: Durch Informationsfusion erzeugte Trajektorien globaler Objekte.

Erkennbar ist die durch Informationsfusion erreichte Verlängerung der Trajektorie von Objekt  $\Omega_2$ .

Durch die vorhandene zeitliche Synchronisation können die Teilinformationen der kooperierenden Multisensorknoten verwendet werden, um einen vergrößerten Lokalisationsbereich zu erzeugen.

Der Vorteil einer redundanten Informationserzeugung, bedingt durch überlappende Sichtbereiche, liefern die Bildpaare 4 und 8. Sporadische Aussetzer in der Objektlokalisierung einzelner Multisensorknoten (Abb. 86) können durch die zusätzlich vorhandenen Informationen unter Verwendung konkurrierender Fusion kompensiert werden (Abb. 87) und führt dazu, dass die Anzahl der erzeugten globalen Objekte verringert werden.

Einen Bereich geringer Lokalisationsleistung zeigt das Bildpaar 2. Aufgrund der Tatsache, dass beide Multisensorknoten in dem entsprechenden Gebiet mit geringer Beleuchtungsintensität keine Informationen liefern (Abb. 86), wird das anfänglich erzeugte Objekt ( $\Omega_1$ ) verloren (Abb. 87). Entsprechend führt das Wiedereintreten der Person in einen Bereich, welcher eine ausreichende Beleuchtung aufweist, zu einer erneuten Erzeugung eines globalen Objekts ( $\Omega_2$ ). Eine Reduktion von Gebieten mit geringer Beleuchtungsintensität lässt sich naturbedingt durch Hinzunahme zusätzlicher MSK erreichen.

Das vergrößerte Lokalisationspotenzial als Ergebnis der Fusion der Informationen beider MSK zeigt sich im direkten Vergleich zwischen Abb. 86 und Abb. 87. Während die Trajektorien der lokalen Objekte MSK1 und MSK2 Diskontinuitäten aufweisen, wird durch die Zusammenführung der Informationen eine permanente Verfolgung des Zielobjekts über einen großen Bereich ermöglicht, was letztendlich in der Erzeugung einer längeren Objekttrajektorie mit erhöhtem Informationsgehalt mündet.

Eine Aussage bzgl. der Lokalisierungsleistung der Multisensorknoten findet sich in Abb. 88. Abgebildet sind die Distanzen der durch die MSK lokal erzeugten Objektpositionen sowie die Entfernungen der globalen Objekte zu den jeweiligen Konstituenten.

Die Geschwindigkeiten der globalen Objekte  $\Omega_1$  und  $\Omega_2$  finden sich in Abb. 89. Bedingt durch die schrittweise Bewegung der Person im Detektionsbereich ergibt sich eine Oszillation der Geschwindigkeit. Dadurch sind auch die leichten, regelmäßigen Variationen in den Bewegungstrajektorien der globalen Objekte zu erklären. Vor Beginn eines durchzuführenden Schrittes wird jeweils eine Verlagerung des Körpers durchgeführt, welche sich in den Bewegungslinien als Muster erkennen lässt.

## 5 Objekt- und Personenverfolgung

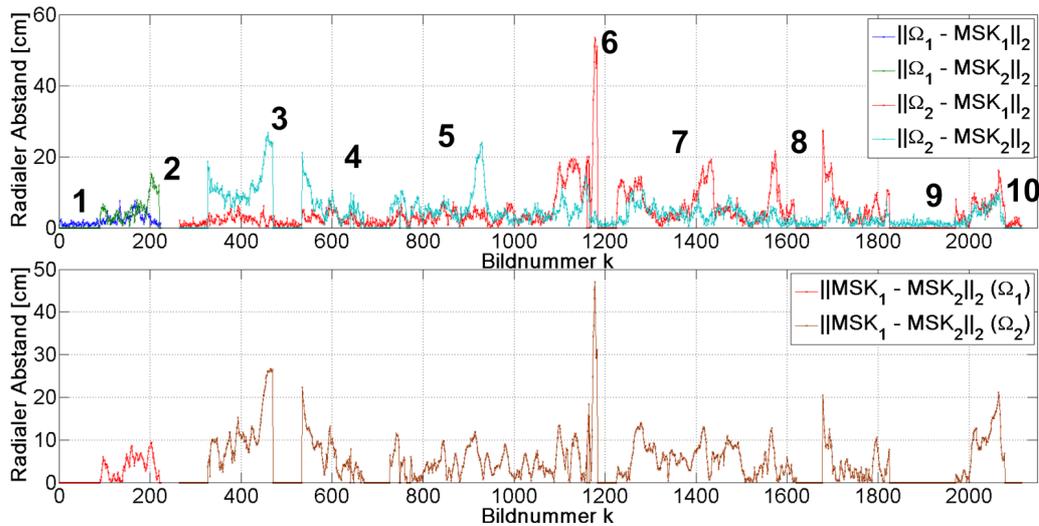


Abbildung 88: Radiale Abstände global und lokal erzeugter Objekte.

Oben: Distanzen der globalen Objekte  $\Omega_1$  und  $\Omega_2$  zu den jeweiligen lokal erzeugten Konstituenten. Unten: Abstände zwischen den beiden durch die unterschiedlichen MSK erzeugten Objekten. Die Nummerierungen im oberen Graphen entsprechen den Aufnahmezeitpunkten der Bildpaare von Abb. 85.

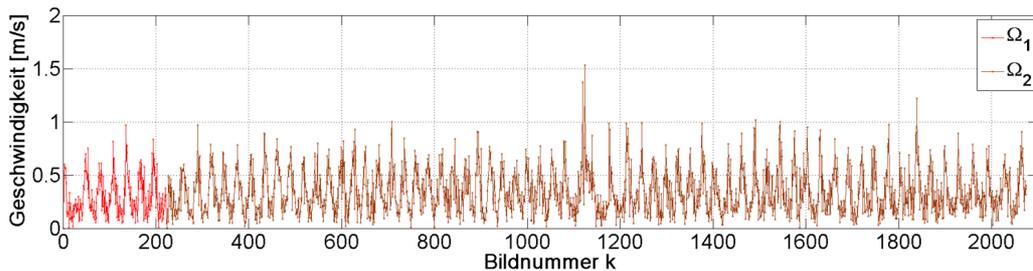


Abbildung 89: Geschwindigkeit der globalen Objekte.

### 5.4.2 REDUKTION GEGENSEITIGER OBJEKTVERDECKUNGEN

Unterschieden werden im Folgenden zwei Fälle:

1. Wiederholte gegenseitige Verdeckung durch Personen bzw. Objekte, welche sich eine gewisse Zeit im Detektionsbereich des Systems bewegen und eventuell miteinander interagieren.
2. Temporale gegenseitige Verdeckungen durch Objekte bzw. Personen, welche sich in einer ungünstigen Anordnung einmalig durch den Detektionsbereich des Systems bewegen.

#### WIEDERHOLTE VERDECKUNGEN DURCH UNKOORDINIERTER BEWEGUNG

Sich wiederholende Verdeckungen, als zwangsläufiges Ergebnis der nicht koordinierten Bewegung mehrerer Personen und/oder Objekte im Detektionsbereich des Sensorsystems führen dazu, dass diese nicht permanent verfolgt werden können. Ihr Wiederauftauchen in den unterschiedlichen Detektionsbereichen der MSK zu unterschiedlichen Zeitpunkten führt unter Umständen zu einer Erzeugung einer Vielzahl von globalen Objekten mit unterschiedlicher Identität<sup>92</sup>. Die gesteigerte Objektzahl führt jedoch zu einer Reduktion der Zeitperiode, in denen ein Objekt eindeutig verfolgt

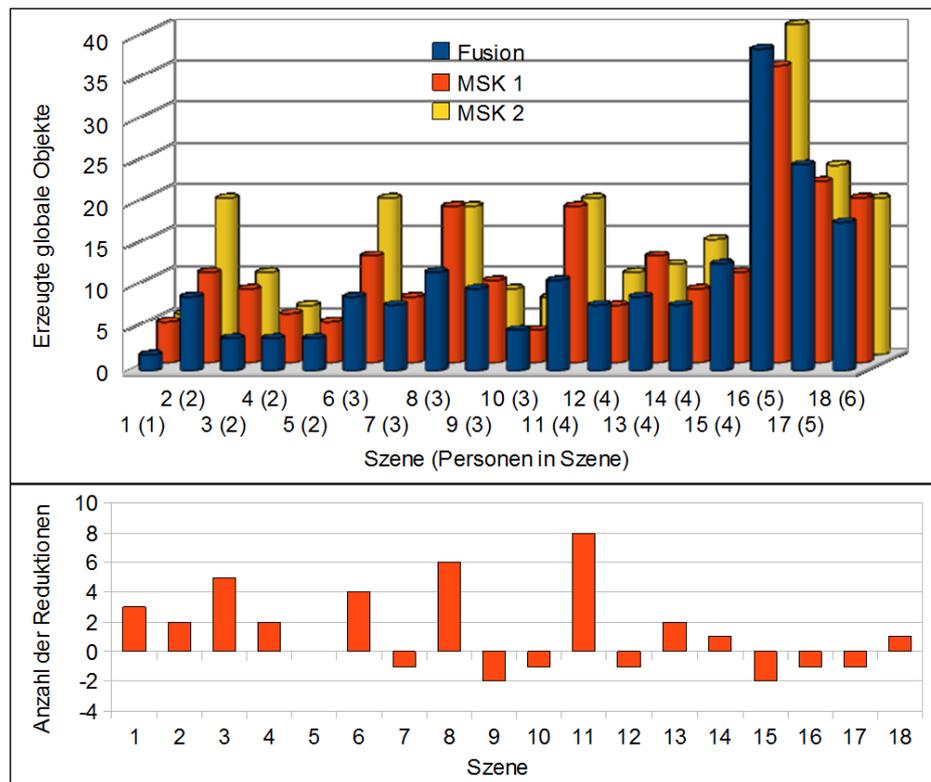
<sup>92</sup>Es sei daran erinnert, dass jedes Objekt während seiner Erzeugung eine eindeutige Kennung erhält.

## 5 Objekt- und Personenverfolgung

werden kann, was letztendlich einer Verringerung des Informationsgehalts entspricht. Eine Reduktion dieser unerwünschten Mehrfacherzeugungen kann naturgemäß durch die Verwendung unterschiedlicher Perspektiven auf die Szene erreicht werden. In Abb. 90 sind die diesbezüglichen Ergebnisse der Fusion in Szenen mit variierender Personenzahl angegeben.

Bei Betrachtung der Graphen zeigt sich, dass für den Fall einer geringen Anzahl von Personen, welche sich simultan im gemeinsamen Detektionsbereich der MSK bewegen, eine Reduktion der global erzeugten Objekte erreicht wird. Im Falle zusätzlicher Personen bzw. Objekten bietet der Einsatz von lediglich zwei Perspektiven nicht genügend Informationen, um die zwangsläufig größere Anzahl an gegenseitigen Verdeckungen zu behandeln.

Der Grund der teilweisen erhöhten Menge der durch Fusion generierten globalen Objekte in Szenen mit einer Personenzahl  $\geq 3$ , liegt in der zu unterschiedlichen Zeitpunkten durchgeführten Erzeugung lokaler Objekte mit gleichem realen Bezug. Die sich durch die Bewegung der Personen ergebenden Verdeckungen haben den Effekt, dass ein reales Objekt in verschiedenen MSK zu unterschiedlichen Zeitpunkten lokalisiert und für unterschiedlich lange Perioden erhalten werden kann. Eine fehlende temporale Überlappung der Existenzdauer der lokalen Objekte, welche sich auf das gleiche reale Objekt beziehen, führt dazu, dass der FSK diese als unterschiedliche Objekte interpretiert und zusätzliche globale Objekte erzeugt<sup>93</sup>.



*Abbildung 90: Reduktion der Erzeugung globaler Objekte.*

*Oben: Anzahl der pro Szene erzeugten globalen Objekte basierend auf lediglich den lokalen Informationen eines MSK sowie der fusionierten Daten. Unten: Menge der durch Fusion erreichten Objektreduktion. Die abgebildeten Werte ergeben sich aus der Differenz der durch Informationsfusion erzeugten Objekte sowie dem minimalen Wert der Erzeugungen durch die Verwendung lediglich eines MSK.*

<sup>93</sup>Wie in Abschnitt 5.3.2 diskutiert, besitzen globale Objekte kein Alter. Dieses impliziert, dass im Falle einer nicht erfolgten Erhaltung durch ein lokales Objekt, das globale Objekt direkt aus der Menge der Vorhandenen entfernt wird.

## 5 Objekt- und Personenverfolgung

### PARTIELLE VERDECKUNGEN DURCH UNGÜNSTIGE PERSPEKTIVEN

Ein häufig auftretender Fall ist die teilweise Verdeckung von Personen bzw. Objekten, welche, je nach Bewegungsablauf durch den Detektionsbereich des Systems, permanent oder wiederkehrend sein kann. Der problematische Aspekt in diesem Fall ist die fehlende Möglichkeit der Separation der Objekte auf Basis der PMD-Daten.

Aufgrund der geringen lateralen Auflösung führen bereits geringe Objekt- bzw. Personenabstände zu einer Verschmelzung in einen Aktivitätscluster. Zu erkennen ist dieser Effekt in den Bilderpaaren 2 und 3 in Abb. 91, welche eine Szene von vier in einer Reihe laufenden Personen durch den Detektionsbereich des Systems beschreibt.

Während eine permanente teilweise Verdeckung zu einer zu geringen Erzeugung lokaler MSK-Objekte führt, werden im Falle einer wiederkehrenden, teilweisen Okklusion zusätzliche lokale Objekte generiert, was unter Umständen<sup>94</sup> eine ungewollte Erzeugung globaler Objekte einleitet.

Die zeitliche Entwicklung der im Detektionsbereich der Multisensorknoten vorhandenen globalen Objekte beschreibt der Graph in der untersten Zeile von Abb. 91. Erkennbar ist die zu verschiedenen Zeitpunkten variierende Anzahl der MSK-Objekte, welche nicht immer mit der tatsächlichen Anzahl der real vorhandenen Personen in der Szene übereinstimmt. Erst durch kooperative Fusion der Informationen der verschiedenen Perspektiven bietet sich die Möglichkeit einer wahrheitsgemäßen Beschreibung der Szene.

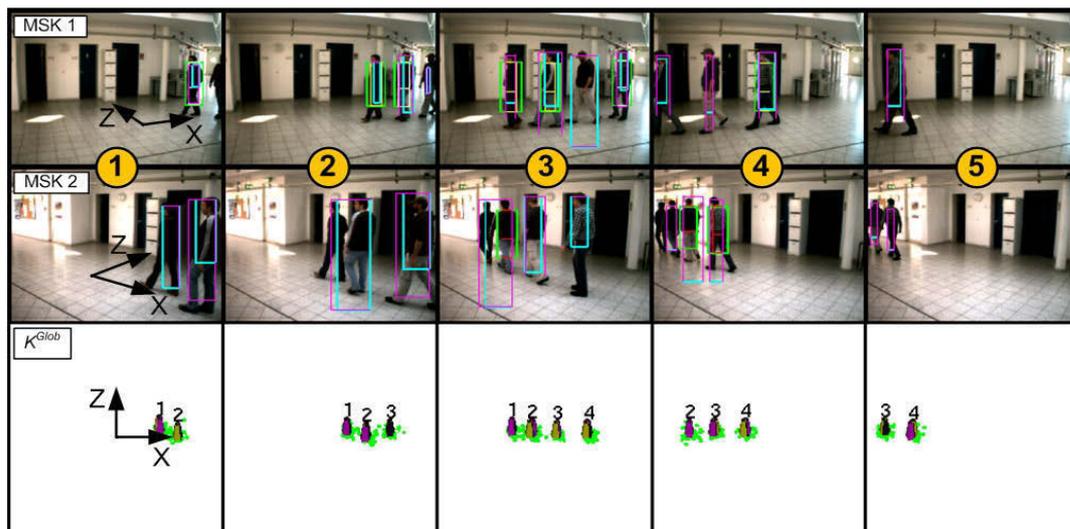


Abbildung 91: Temporale Verdeckungen durch ungünstige Perspektive.

Obere zwei Zeilen: Momentaufnahmen der untersuchten Szene aus den jeweiligen Perspektiven der verwendeten MSK. Zusätzlich zu den auf die jeweiligen Aktivitätscluster bezogenen Daten enthalten die Bilder die durch den HOG-Algorithmus im Falle einer erfolgreich durchgeführten Klassifikation erzeugten Informationen (Abb. 68). Dritte Zeile: Positionen und Identitäten der erzeugten Objekte im globalen Koordinatensystem  $K^{Glob}$ . Die Nummerierung bezieht sich hierbei auf die Identitäten der globalen Objekte.

<sup>94</sup>Siehe dazu die Erläuterungen bzgl. der wiederkehrenden Verdeckungen aufgrund unkoordinierter Bewegungen aus vorigem Abschnitt.

## 5 Objekt- und Personenverfolgung

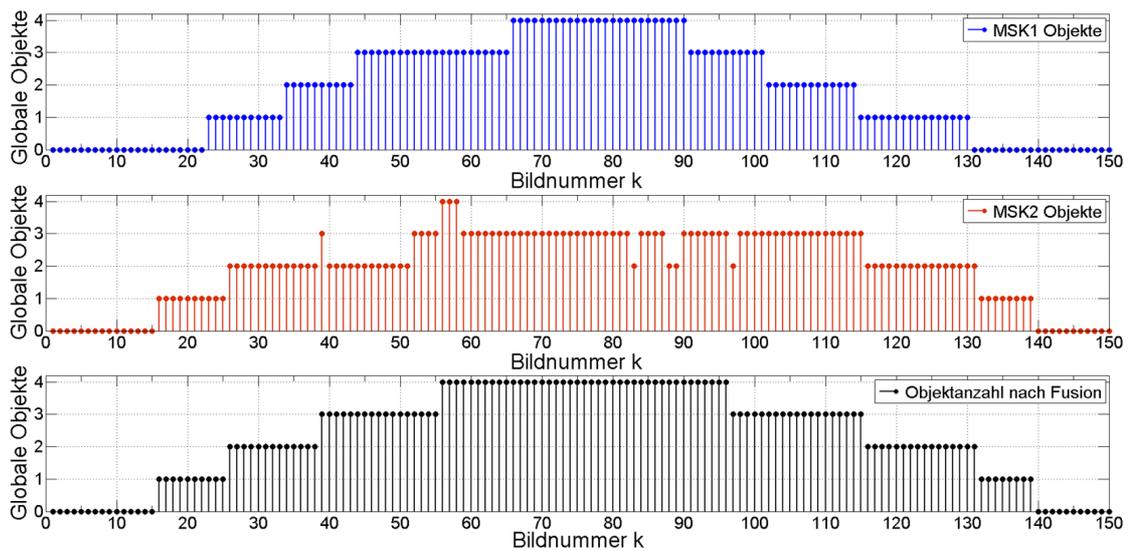


Abbildung 92: Objekterzeugungen auf Grundlage unterschiedlicher Informationen.

Analyse einer Sequenz bestehend aus 150 Bildern. Während die oberen beiden Graphen die Erzeugungen globaler Objekte auf Grundlage der Informationen lediglich eines MSK wiedergeben, spiegelt der untere Graph die Erzeugung auf Basis der fusionierten Daten wider. Durch die Fusion der unterschiedlichen Informationen ergibt sich die Möglichkeit einer adäquateren Beschreibung der in Abb. 91 teilweise wiedergegebenen Szene.

### 5.5 ANALYSE DES ZEITBEDARFS ZUR ERZEUGUNG GLOBALER INFORMATIONEN

Der letzte Abschnitt dieses Kapitels ist der Analyse der Performanz des Gesamtsystems im Hinblick auf die Erzeugung globaler Informationen gewidmet. Diskutiert wird zunächst die Fusionsgeschwindigkeit des verfolgten Ansatzes auf dem FSK.

Im Anschluss dazu erfolgt die Betrachtung der Operationsgeschwindigkeit des gesamten Systems. Die hierbei gezeigten Ergebnisse ermöglichen eine Aussage, inwieweit die Einsatzfähigkeit des gewählten Ansatzes unter den Forderungen einer echtzeitfähigen Operation sowie Skalierbarkeit gegeben ist.

#### 5.5.1 GESCHWINDIGKEIT DER DATENFUSION

Der verfolgte Ansatz zur Fusion der Informationen der MSK auf Basis ihrer lokalen Zustandsvektoren unter Verwendung eines Partikelfilters ist im Hinblick auf die aufzubringende Rechenleistung äußerst moderat (Abb. 93).

Während im oberen Graphen die zur Erzeugung der globalen Informationen unter Verwendung lediglich eines MSK notwendige Berechnungszeit angegeben ist und letztlich nur einen Hinweis auf die zu erwartende Größenordnung gibt, enthält der untere Graph die Berechnungszeiten unter Verwendung zweier MSK.

Die geringe Steigerung der Berechnungsdauer ist jedoch aufgrund der zugrunde liegenden Berechnung (Abschnitt 5.3.4.2) zu erwarten gewesen und lässt den Schluss zu, dass auch unter Hinzunahme weiterer MSK von einer drastischen Erhöhung der zur Berechnung der globalen Informationen notwendigen Zeitspanne nicht auszugehen ist. Somit bietet der Ansatz die Möglichkeit der Fusion mehrerer Dekaden von lokalen Zustandsvektoren, ohne einen wesentlichen Einfluss auf die Gesamtberechnungsdauer zu haben.

## 5 Objekt- und Personenverfolgung

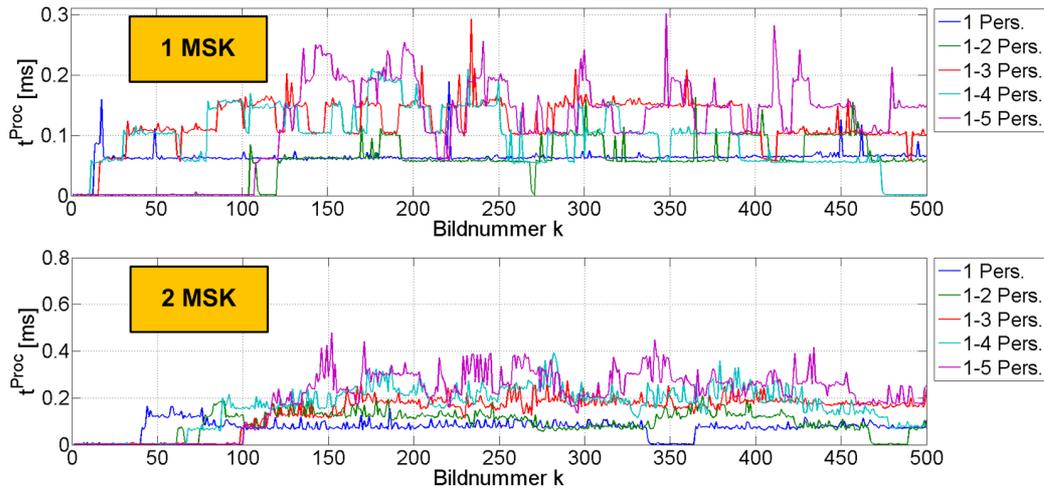


Abbildung 93: Fusionsgeschwindigkeit auf dem FSK.

Analyse einer Sequenz bestehend aus 500 Bildern. Oberer Graph: Die zur Fusion notwendige Berechnungsdauer unter Betrachtung von Szenen mit unterschiedlicher Personenzahl und nur einem MSK. Unterer Graph: Berechnungsdauer unter Hinzunahme der Informationen eines zusätzlichen MSK.

### 5.5.2 GESCHWINDIGKEIT DES GESAMTSYSTEMS

Die zur Bearbeitung eines Zeitpunkts notwendige Zeitspanne des Gesamtsystems, unter Variation der in einer Szene maximalen vorhandenen Personenzahl, ergibt sich aus Abb. 94.

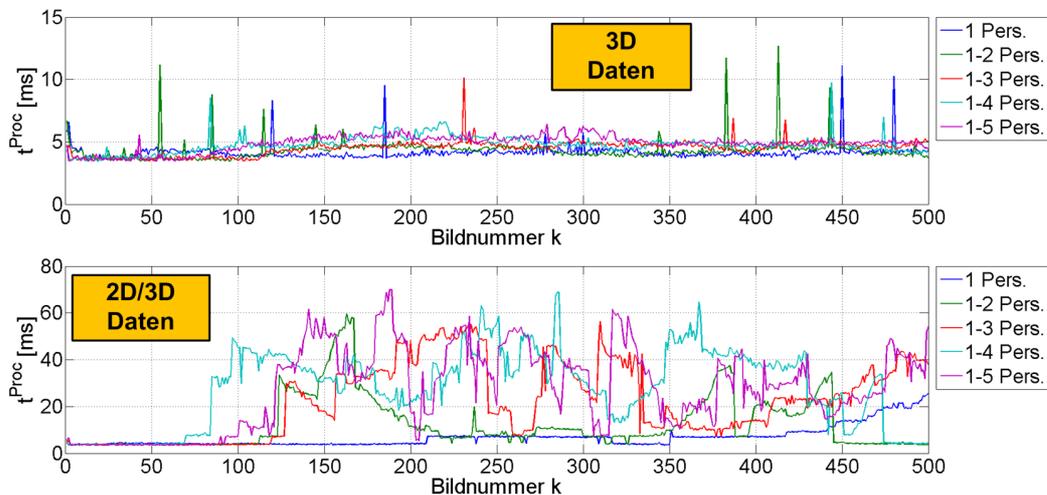


Abbildung 94: Totale Berechnungsdauer des Gesamtsystems.

Analyse einer Sequenz bestehend aus 500 Bildern (gleiche Sequenz wie auch in Abb. 93). Oberer Graph: Totale Berechnungsdauer unter lediglicher Verwendung der PMD-Daten (3D-Bild). Unterer Graph: Totale Berechnungsdauer unter Einsatz aller Modalitäten der MultiCam (2D/3D-Bild). In den angegebenen Berechnungszeiten sind die notwendigen Zeitspannen zur Notifikation der MSK durch den FSK, die Versendungen der lokalen Zustandsvektoren, die Datenangleichung sowie die Fusion der Informationen bereits enthalten.

## 5 Objekt- und Personenverfolgung

---

Während der obere Graph die kompletten Berechnungszeiten unter Verwendung von lediglich den PMD-basierten Informationen wiedergibt (3D-Daten), spiegelt der untere Graph die notwendige Prozessierungszeit eines Zeitpunkts unter Einsatz sämtlicher MultiCam-Daten (2D/3D-Daten) wider, und schließt somit die Klassifikation von Objekten auf Basis hoch aufgelöster Intensitätsdaten mit ein. In den in Abb. 94 angegebenen Berechnungszeiten sind die Benachrichtigung der MSK durch den FSK, die Erzeugung und Versendung der lokalen Zustandsvektoren durch die MSK an den FSK, sowie deren Fusion im FSK mit enthalten, und entsprechen damit der in Abb. 2 als  $t_k^{Proc}$  angegebenen Zeitspannen.

Für den Fall der monomodalen Operation auf dem 3D-Bild, was einer reinen Lokalisation und Verfolgung von unklassifizierten Objekten entspricht, ergibt sich anhand der in Abb. 94 abgebildeten Prozessierungszeiten eine Zeitauflösung des Systems, welche nur von der verwendeten Sensorik abhängig ist.

Ein Vergleich der zur Berechnung notwendigen Zeitspannen (Abb. 94), mit denen aus dem entsprechenden Graphen aus Abb. 79 zeigt, dass sich durch Hinzunahme der zur Kommunikation und Fusion notwendigen Zeiträumen lediglich eine geringe Erhöhung der zur Berechnung der globalen Information notwendigen Zeitspanne ergibt.

Gleiches gilt für den Fall der Verwendung sämtlicher MutiCam-Modalitäten. Durch direkten Vergleich der Daten des unteren Graphen von Abb. 94 mit den entsprechenden Angaben in Abb. 79 ist eine quantitative Veränderung in den Berechnungszeiten nicht erkennbar. Die Hinzunahme eines zusätzlichen MSK hat im Wesentlichen nur marginalen Einfluss auf die Geschwindigkeit des Gesamtsystems und ist eine direkte Folge des verwendeten Ansatzes.

Anzumerken ist, dass die in Abb. 94 enthaltenen Informationen auf Grundlage der Verwendung eines Standardrechners (CPU: E6600, 3GHz) sowie eines Laptops (CPU: Intel i5 M560, 2.67GHz) als Daten-prozessierende Einheiten erzeugt worden sind. Eine Optimierung des Quellcodes im Hinblick auf eine parallele Prozessierung der Daten auf der CPU wurde nicht durchgeführt.



# 6 SCHLUSSBETRACHTUNG

Im Folgenden werden zunächst die direkten Vorteile beleuchtet, welche sich durch den Einsatz eines Multisensorsystems wie der MultiCam bieten.

Anschließend folgt eine Betrachtung der im Rahmen dieser Arbeit erreichten Ziele. Diskutiert wird, ob und in welchem Masse das Erreichte mit den in Abschnitt 1.1 aufgestellten Zielsetzungen übereinstimmt.

Des Weiteren werden direkte Möglichkeiten des Einsatzes des beschriebenen Systems benannt, wobei zusätzlich auch auf die momentanen Beschränkungen eingegangen wird. Abgeschlossen wird dieses Kapitel mit einer Betrachtung möglicher Erweiterungen des Systems.

## 6.1 POTENZIAL DER SENSORIK

Das im Rahmen dieser Arbeit verwendete neuartige Multisensorkonzept, welches einer Kombination aus ToF- und CMOS-Sensor unter Verwendung eines monokularen Aufbaus entspricht, bietet folgende und teilweise einmalige Vorteile:

1. Durch die Verwendung von ToF-Kameras ist die Erzeugung von Tiefeninformationen auch unter fordernden Lichtbedingungen möglich. Aufgrund des aktiven Messprinzips kann die MultiCam auch in Umgebungen betrieben werden, die für konventionelle Kamerasysteme nicht zugänglich sind.
2. Durch den gewählten monokularen Ansatz erleichtert sich der Vorgang der Kalibration sowie der Registrierung der Bilder der unterschiedlichen Modalitäten erheblich. Im Vergleich zu binokularen Systemen, in denen die Kalibration sowie das Registrieren der Bildinformationen einen erheblichen Aufwand erfordert, bietet die MultiCam die Möglichkeit einer vereinfachten und rechen technisch effizienten Operation, welche sich im Rahmen eines echtzeitfähigen Systems mit hoher Zeitaufösung nutzen lässt.
3. Bedingt durch die interne Konzeption liefert das Sensorsystem zusätzlich zeitlich zueinander synchronisierte 2D/3D-Bilder. Dieses ermöglicht eine verlässliche simultane Akquisition von hoch aufgelösten Intensitätsdaten sowie Distanzdaten einer Szene. Synchronisationsmechanismen, welche im Falle der Verwendung separater Systeme notwendig werden, entfallen gänzlich.
4. Des Weiteren besitzen ToF-Systeme den Vorteil, dass sich eine Lokalisation von Objekten relativ einfach und schnell durchführen lässt. Aufgrund der im Vergleich zu normalen Kamerasystemen geringen Datendichte sowie der Güte der Informationen lässt sich die im Allgemeinen nichttriviale Aufgabe der räumlichen Positions- und Geschwindigkeitsbestimmung unter Einsatz relativ geringer Rechenleistung rasch und verlässlich durchführen.
5. Es bietet sich die Möglichkeit der Fusion von Informationen unterschiedlicher Modalitäten auf Pixel-, Merkmals- und Entscheidungsebene unter Verwendung lediglich eines Sensorsystems.

Unter Betrachtung obiger Punkte ist es evident, dass das im Rahmen dieser Arbeit eingesetzte Multisensorsystem in Form der MultiCam ein hohes Anwendungspotenzial bietet und prädestiniert für den Einsatz in Systemen scheint, welche unter wechselnden Umgebungssichtverhältnissen operieren müssen und in denen eine schnelle und zuverlässige Reaktion auf Objektbewegungen von essenzieller Bedeutung ist.

### 6.2 BETRACHTUNG DER ZIELSETZUNG UND ERRUNGENSCHAFTEN

An dieser Stelle sei erneut der Blick auf die in Abschnitt 1.1 aufgestellten Zielsetzungen gerichtet.

Als grundlegend zu erreichende Ziele wurde eine

1. echtzeitfähige Operation,
2. einfache Skalierbarkeit,
3. Operation auch unter fordernden Lichtverhältnissen bzw. Dunkelheit (*Graceful Degradation*),
4. sowie funktionale Erweiterbarkeit

des Systems zur Lokalisation und Verfolgung von Personen benannt.

#### ECHTZEITFÄHIGE OPERATION

Bezüglich der Zeitauflösung des realisierten Systems kann gesagt werden, dass für den Fall der lediglichen Verwendung von PMD-Daten die Bearbeitungsdauer des Gesamtsystems pro Bild bzw. Zeitpunkt (Abb. 94) mit höchstens 11 ms weit unter der zur Akquisition maximal angesetzten Zeitspanne  $t_{Max}^{Akq}$  (Abschnitt 1.2) liegt. Eine Beschränkung der zeitlichen Auflösung findet lediglich durch die MultiCam bzw. ihrer Hardwareeigenschaften selbst statt. Eine echtzeitfähige Operation ist in diesem Fall unabhängig von der Anzahl der im Sichtfeld der Kamera befindlichen Personen uneingeschränkt möglich.

Im Falle der Hinzunahme von 2D-Daten ergibt sich für den Fall geringer Personenzahlen, dass aufgrund der Beschleunigung der Personenklassifikatoren ebenfalls eine Berechnungszeit erreicht werden kann, welche unter der maximalen Periode zur Bildakquisition angenommenen Zeitspanne  $t_{Max}^{Akq}$  liegt. Im Falle des Vorhandenseins von vier oder mehr Personen in einer Szene ergibt sich aufgrund der großen Anzahl an notwendigen Berechnungen auf den Intensitätsdaten, dass  $t_{Max}^{Akq}$  nicht permanent unterschritten werden kann (Abb. 94). Bei Hinzunahme der Intensitätsdaten ergibt sich somit eine von der Personenzahl und Position der Personen im Sichtfeld der Kamera abhängige Geschwindigkeit, welche im Hinblick auf die Zielsetzung zu einer reduzierten Echtzeitfähigkeit führt. Eine simple Modifikation des Systems, welche zu einer permanenten Echtzeitfähigkeit des Systems führen würde, wird im letzten Abschnitt dieser Arbeit diskutiert (Abschnitt 6.5).

#### EINFACHE SKALIERBARKEIT

Aufgrund der Verwendung agentenbasierter MSK ist die Möglichkeit der einfachen Skalierung gegeben. Bedingt durch die inhärente Parallelität des Systems konzentriert sich die Steigerung der Rechenlast im Wesentlichen auf den Fusionsknoten (Abschnitt 5.5, Abb. 93) und erlaubt die simultane Operation einer Vielzahl von MSK. Damit bietet sich gleichzeitig die Möglichkeit der Anpassung der Systemtopologie an verschiedene Gegebenheiten und ermöglicht den Einsatz des Systems ohne fundamentale Anforderungen an die Geometrie der Räume zu stellen.

Durch Verwendung einer definierten Flussrichtung der Nutzinformation der Agenten hin zu einem einzelnen, definierten Fusionsagenten sowie der Übermittlung von lediglich per Bild erzeugten Zustandsvektoren (46) wird gewährleistet, dass die zur Übermittlung der Daten notwendige Bandbreite bei Weitem nicht überschritten wird<sup>95</sup> und nur moderat wächst. Dieses ist eine direkte Folge der Prozessierung der Sensorinformationen sowie der Erzeugung von abgeleitetem Wissen lokal auf den Agenten.

---

<sup>95</sup>Angenommen wird hier eine auf Fast-Ethernet (100Mbit/s) basierende Kommunikationsinfrastruktur.

## 6 Schlussbetrachtung

---

### OPERATION UNTER FORDERNDEN LICHTVERHÄLTNISSEN

Aufgrund der konzeptionellen Separation von 2D- und 3D-Daten bietet der verfolgte Ansatz die Grundfunktionalität der Lokalisation und Verfolgung von unklassifizierten Objekten auch in Umgebungen mit fordernden Lichtverhältnissen, was einen massiven Vorteil im Vergleich zur Verwendung von lediglich konventioneller Kamertechnik darstellt.

Wie in Kapitel 4 gezeigt, basiert die Erzeugung sowohl des Hintergrunds (Abschnitt 4.2.2.2) als auch der Aktivitätscluster (Abschnitt 4.2.4) lediglich auf der Verwendung der präprozessierten Informationen des PMD-Chips (Abschnitt 4.1.2). Die Beschränkung auf Distanzdaten gewährleistet die reduzierte Funktionalität der Lokalisation multipler Objekte auch im Falle der Abwesenheit vom Umgebungslicht und kann somit unabhängig von der Tageszeit durchgeführt werden.

### FUNKTIONALE ERWEITERBARKEIT

Bedingt durch die Erweiterbarkeit des zugrunde liegenden Agentensystems mittels zuladbaren Bibliotheken ist eine funktionale Erweiterung des Systems in einer einfachen Weise möglich (Abschnitt 3.1.7). Der Einsatz des implementierten Agentensystems ist somit nicht auf spezielle Sensoren beschränkt.

Aufgrund der relativ freien Wahl in Bezug auf die Topologie sowie der Funktionalität ist der Einsatz des agentenbasierten Grundsystems unter Verwendung einer auf die entsprechende Zielsetzung hin implementierten Funktionalität, erfolgreich demonstriert worden [218], [219]. Anzumerken ist an dieser Stelle, dass die im Rahmen des Projekts entwickelte Algorithmik zur Klassifikation mittels Thermalkameras durch Herrn Dr. Stefan Lammers implementiert worden ist. Die Algorithmik zur Detektion und Lokalisation mittels PIR-Sensoren sowie der Datenfusion (lokal und auf dem FSK) als auch die notwendigen Implementierungen zur Integration sämtlicher Sensorik (Sensor-APIs) in die entsprechenden Rechnersysteme wurden durch den Autor dieser Arbeit entwickelt. Zusätzlich wurde von Letzterem auch die Implementierung einer Applikation zur grafischen Darstellung der Ergebnisse übernommen.

Bei Betrachtung der Ergebnisse kann gesagt werden, dass die in Abschnitt 1.1 aufgestellten Forderungen einer echtzeitfähigen, skalierbaren Objekt- bzw. Personenlokalisierung und Verfolgung auch unter fordernden Umgebungslichtverhältnissen unter Verwendung der MultiCam als Datenquelle erreicht worden ist. Die erreichte Zeitauflösung des Systems ist für eine Verfolgung von sich normal bewegenden Personen in Innenräumen ausreichend. Durch die ausschließliche Verwendung der Informationen des PMD-Chips lässt sich die Zeitauflösung, unter Aussparung der Objektklassifikation, weiterhin erheblich verbessern.

Obwohl die im Rahmen dieser Arbeit betrachteten agentenbasierten Multisensorknoten synchron betrieben wurden, ist eine solche Operation nicht zwingend. Mittels der inhärenten Möglichkeit der Agentenkommunikation können Informationen prinzipiell zu beliebigen Zeitpunkten ausgetauscht werden. Denkbar ist z. B. eine Systemkonfiguration, indem ein Teil der Agenten ihre Informationen permanent und synchronisiert anhand einer externen Taktung erzeugen und dem Fusionsknoten zusenden. Die verbleibenden Agenten hingegen operieren unabhängig von der Taktung. Sie liefern ihre Daten lediglich im Falle bestimmter Ereignisse an den zur Fusion der Daten zuständigen Agenten ab.

### ZUSÄTZLICHE ERRUNGENSCHAFTEN

Durch die Wahl einer Optik mit einer Brennweite von 6mm sowie einer entsprechenden Beleuchtung konnte ein Sichtbereich pro MultiCam erzeugt werden, welcher eine ausreichende Größe besitzt, um auch mehrere, miteinander agierende Objekte bzw. Personen simultan lokalisieren und verfolgen zu können (z. B. Abb. 50, 51).

## 6 Schlussbetrachtung

---

Im Falle der Detektion und Lokalisation von Objekten (Abschnitte 4.2.1, 4.2.3) konnte gezeigt werden, dass durch die Beschränkung auf die Tiefeninformationen des PMD-Chips gewährleistet werden kann, dass eine Detektion und Lokalisation von Objekten auch unter Umständen möglich ist, in denen konventionelle Kamertechnik versagt.

Zusätzlich dazu führt die Beschränkung der Methoden auf 3D-Informationen dazu, dass die zur Lokalisation und Detektion angewandten Methoden eines rechentechnisch geringen Aufwands bedürfen. Eine maximale Bearbeitungsdauer von unter 6ms (Abb. 79) pro Aufnahme<sup>96</sup> spiegelt diesen Sachverhalt wider. Weiterhin ist ersichtlich, dass die zur Detektion und Lokalisation notwendige Berechnungsdauer nur unwesentlich von der Anzahl der Personen im Sichtbereich einer MultiCam rechentechnisch beeinflusst wird. Hinzuzufügen ist, dass die Berechnungsdauer der lokalen Objektverfolgung mittels des Kalman-Filters in dieser Dauer mit enthalten ist.

Mittels des Einsatzes einer koordinierten Datenakquisition aus unterschiedlichen Perspektiven ist eine erhöhte Lokalisationsleistung sowie eine Vergrößerung des effektiven Lokalisationsbereiches erreicht worden (Abschnitt 5.4.1, Abb. 87).

Aufgrund der einfach durchzuführenden Skalierung des Netzwerks durch simples Hinzufügen zusätzlicher Knoten bietet sich die Möglichkeit einer Erweiterung des Detektionsbereichs des Systems auch auf größere Räume. Anhand der Informationen aus Abb. 79, 93 sowie 94 ist zu entnehmen, dass sich das Hinzufügen zusätzlicher Knoten nur unwesentlich auf die Berechnungsdauer des Gesamtsystems auswirkt. Dieses ist nicht weiter überraschend, da zur Berechnung des Gesamtzustandes auf dem FSK lediglich die momentane Parametermenge  $c_k^j$  (46) der teilnehmenden MSK übertragen werden.

Ebenfalls gezeigt wurde die Möglichkeit der Verfolgung mehrerer Personen unter Steigerung des Informationsgehalts der erzeugten Trajektorien bedingt durch die Verringerung gegenseitiger Verdeckungen (Abschnitt 5.4.2). Anzumerken ist, dass die Steigerung des Informationsgehalts nur für den Fall der Anwesenheit einer geringen Anzahl von Personen ( $\leq 3$ ) im Detektionsbereich des Gesamtsystems gilt (Abb. 90). Als Hauptgrund hierfür wird die Verwendung von lediglich zwei MSK angenommen.

Weiterhin wurde demonstriert, dass durch eine relativ einfache Fusion der Informationen unter Verwendung von Merkmalen, erzeugt auf Grundlage der 3D-Daten, eine dramatische Beschleunigung der klassifizierenden Algorithmen AdaBoost und HOG erreicht werden konnte (Abschnitt 4.3, Abb. 56, 59, 63, 66). Die erreichte Reduktion der notwendigen Berechnungsschritte erlaubt damit den Einsatz dieser Algorithmen in Systemen zur echtzeitfähigen Überwachung dynamischer Objekte.

### 6.3 EINSATZMÖGLICHKEITEN

Aufgrund der genannten Eigenschaften des agentenbasierten Basissystems ergibt sich die Möglichkeit des vielfältigen Einsatzes, da weder die Topologie noch die explizite Funktionalität durch die gewählte Implementation a priori vorherbestimmt werden. Ein Einsatz des vorgestellten Systems ist vor allem in sicherheitstechnisch relevanten Räumen denkbar:

1. **Operationssaal:** Dokumentation der Anwesenheit von Personal an entsprechender Position im Raum und/oder Überwachung der Patientenposition.
2. **Intensivstation:** Detektion von ungewollten oder unzulässigen Patientenbewegungen. Überprüfung der Anwesenheit von Pflegepersonal.
3. **Altenheim:** Überwachung der Bewohner während des Schlafs. Überwachung von Räumen, z. B. dem Speisesaal, zur schnellen Detektion von gestürzten Menschen.
4. **Operationszentrale in Kraftwerken:** Dokumentation der Anwesenheit von Personal an entsprechender Position im Raum.

---

<sup>96</sup>Diese Zeitspanne entspricht der Berechnungsdauer der lokal auf einem Knoten prozessierten Informationen ohne zusätzliche Kommunikation mit dem Fusionsknoten.

---

## 6 Schlussbetrachtung

---

5. **Banken (Tresorräume):** Öffnen bzw. Schließen der Tür nur bei Anwesenheit einer bestimmten Anzahl von Personen an bestimmten Raumpositionen.
6. **Museum:** Definition von erlaubten Aufenthaltspositionen der Besucher zwecks Schutz der Ausstellungsobjekte. Aufgrund der Möglichkeit einer reduzierten Funktionalität auch im Dunklen könnte das System zusätzlich als Schutzmaßnahme gegen ungewolltes Eindringen verwendet werden.
7. **Geschäfte:** Einsatz des Systems zur Besucheranalyse. Die Funktionalität der auf diesem Gebiet existierenden Lösungen, z. B. [232], [233], könnten direkt durch den Einsatz des hier beschriebenen Systems, zumindest teilweise, übernommen werden. Zusätzlich böte sich durch die Eigenschaft der reduzierten Funktionalität in Dunkelheit, analog zum Fall des Museums, die Möglichkeit der Operation in einem Sicherheitsmodus.
8. Durch die bewiesene Geschwindigkeit der lokalen sowie globalen Verarbeitung (Abb. 79, 85), insbesondere im Falle der Beschränkung auf Tiefeninformationen, eignet sich der Ansatz auch für Umgebungen mit schnellen Objektbewegungen, z. B. in industriellen Anlagen mit Roboterfertigung. Durch Überwachung des Aktionsradius eines Roboters könnte verhindert werden, dass menschliches Personal beim Eintreten in die Aktivitäts- bzw. Gefahrenzone Schaden erleidet, indem der Roboter rechtzeitig gestoppt oder zumindest gebremst wird. Zusätzlich böte die Operationsstabilität des Systems in Bezug auf fordernde Lichtverhältnisse einen weiteren Vorteil.

### 6.4 MOMENTANE SYSTEMBESCHRÄNKUNGEN

Entscheidend für die Performanz eines auf einer MultiCam aufbauenden Multisensorknotens ist die verwendete Beleuchtung. Die zur Überwachung verwendbare maximale Entfernung, unter Beibehaltung einer zur Verfolgung von Objekten notwendigen Zeitauflösung, ist durch sie limitiert. Eine Ausweitung des Detektionsbereichs auf über den im Rahmen dieser Arbeit betrachteten Distanzbereich erfordert eine gesteigerte Intensität, was durch die Hinzunahme zusätzlicher Beleuchtungen prinzipiell möglich ist. Durch die begrenzte Verfügbarkeit von Beleuchtungen zum Zeitpunkt der Bildakquisitionen ist die Verwendung zusätzlicher Einheiten zwecks Steigerung der Emission der Beleuchtung jedoch nicht möglich gewesen.

Bedingt durch die notwendige Beleuchtungsintensität zur Abdeckung des betrachteten Distanzbereichs sind die Distanzinformationen von nahen Objekten aufgrund der auftretenden Multipropagation des reflektierten Lichts nur eingeschränkt verwendbar, sodass der Nahbereich eines MSK einem nicht zugänglichen Gebiet entspricht. Durch die Verwendung zusätzlicher MSK ließen sich solche Gebiete prinzipiell reduzieren, indem durch entsprechende Ausrichtung der MultiCam eine Überwachung des Nahbereichs anderer Einheiten übernommen werden kann.

Ein zusätzlich beschränkender Faktor stellt die geringe laterale Auflösung des PMD-Chips mit lediglich 64x48 Pixel dar. Durch die relativ grobe Aufteilung des Detektionsbereichs reduziert sich die laterale Separationsleistung in Bezug auf vorhandene Objekte mit zunehmender Entfernung. Eine räumliche Trennung von Objekten bzw. Personen wird dadurch erschwert und mindert die Performanz des in Abschnitt 4.2.4 vorgestellten Algorithmus zur Lokalisation.

Die in Kapitel 5 getroffenen Aussagen basieren auf der Verwendung von lediglich zwei Multisensorknoten. Aufgrund der prototypischen Natur der MultiCam und das damit verbundene Fehlen zusätzlicher Einheiten ist eine Untersuchung von Szenen unter Verwendung weiterer Perspektiven nicht möglich gewesen. Weiterhin konnte eine Untersuchung zwecks Beweis der Annahme, dass lediglich die geringe Anzahl der eingesetzten MultiCams für die Reduktion des Informationsgehalts der Trajektorien im Falle der Anwesenheit von mehreren Personen verantwortlich ist, im Rahmen dieser Arbeit nicht durchgeführt werden.

Zur sinngemäßen Operation bedarf es des Vorwissens bezüglich der Positionen der Multisensorknoten relativ zu einem a priori zu definierenden globalen Koordinatensystem. Diese Informationen müssen im Falle einer expliziten Anordnung manuell gesammelt werden.

### 6.5 ERWEITERUNGSMÖGLICHKEITEN UND AUSBLICK

Der wohl direkteste Ansatz zur Erweiterung des Systems bestünde in der Hinzunahme zusätzlicher Sensorknoten zur besseren Abdeckung des betrachteten Detektionsbereichs und der damit verbundenen Steigerung des Lokalisations-, Identifikations- und Verfolgungspotenzials. Des Weiteren ist davon auszugehen, dass der Einsatz lateral höher aufgelöster PMD-Chips zu einem erhöhten Leistungsvermögen des implementierten Algorithmus zur Lokalisation führt.

Es wurde bereits angemerkt (Abschnitt 6.2), dass im Falle der Verwendung von sowohl 2D- als auch 3D-Informationen die Forderung der Echtzeitfähigkeit vom System nur unter der Bedingung des Vorhandenseins einer geringer Anzahl von Personen im Sichtbereich erfüllt werden kann. Hier böte die zyklische Betrachtung der vorhandenen Objekte nach dem Round-Robin-Prinzip eine Möglichkeit der zuverlässigen Reduktion der pro Bild notwendigen Berechnungsdauer. Der Wahrscheinlichkeit, dass aufgrund der Verwendung einer geringen Anzahl von Multisensorknoten ein Objekt der Klassifikation entgeht, kann dabei durch die Verwendung zusätzlicher MSK entgegen gewirkt werden.

Bedingt durch den modularen Aufbau des agentenbasierten Sensorknotenkonzepts bietet sich weiterhin die Möglichkeit des Einsatzes zusätzlicher Sensorik wie PIR-Sensoren oder Thermalkameras zur Steigerung des Informationspotenzials an. Denkbar wäre der Einsatz heterogener MSK, um sich den Eigenarten eines zur Überwachung vorgesehenen Gebietes besser anzupassen. Ein diesbezüglicher Beweis der Einsetzbarkeit des Systems unter Verwendung inhomogener Sensorknoten wurde bereits im POMSe-Projekt geliefert (Abschnitt 3.3). Weiterhin denkbar wäre die Verwendung von zum Gesamtsystem asynchron operierenden MSK. Ausgestattet wären diese dann lediglich mit passiven, wenig Energie verbrauchenden Sensoren (z. B. PIR-Sensoren), und besäßen die ausschließliche Aufgabe, das Gesamtsystem im Falle der Detektion einer Objektpräsenz ganz oder teilweise zu aktivieren. Der sich daraus ergebende vorteilhafte Nebeneffekt wäre eine gesteigerte Energieeffizienz.

Eine der momentanen Systembeschränkungen ist die Notwendigkeit der manuellen Messung der Position und Orientierung der MSK in einem globalen Koordinatensystem. Durch Hinzunahme von Lokalisations- [230] und Orientierungssensorik [231] könnte dieser Prozess automatisiert werden. Ein sich direkt ergebender Vorteil wäre die Möglichkeit der Detektion von Veränderungen der Positionen und/oder Orientierungen der MSK.

Ein interessanter, im Sinne einer beschleunigten Personenklassifikation zusätzlich gangbarer Weg liegt in der Verwendung neuartiger Berechnungsmodelle. Durch die heutzutage einfache Verfügbarkeit von Mehrkern-CPU's sowie Grafikkarten und unter Verwendung von z. B. OpenCL [200], CUDA [201] oder DirectCompute [202] öffnet sich der Weg einer massiv parallelen Bearbeitung der Daten und ermöglicht eine teilweise erhebliche Reduktion der notwendigen Berechnungszeiträume.

In Bezug auf die vom System gelieferten Informationen wäre der nächste logische Schritt die Verwendung dieser zur Situationsanalyse. Speziell für den Nahbereich der MultiCam sind mittlerweile vielversprechende Softwarepakete erhältlich, welche direkt eingesetzt werden könnten. Z. B. bietet sich durch die Verwendung der *faceAPI* von *seeingmachines* [203] die Möglichkeit, Gesichtsausdrücke von Personen zu erkennen und diese als Informationen im Rahmen einer Situationsabschätzung weiterzuverarbeiten. Im Sinne einer Verhaltens- oder Bewegungsanalyse bieten die Produkte von *Softkinetic* (iisu-SDK) [204] oder *Omek-Interaktive* (Omek-SDK) [205] Werkzeuge an, um die Positionen von Personen in der Nähe einer Kamera aufzunehmen und bis zu einem gewissen Grad auch zu interpretieren. Zusätzlich ergibt sich dadurch auch eine interessante Möglichkeit zur Interaktion mit dem System selbst.

Eine lohnenswerte Weiterentwicklung wäre die simultane Verwendung mehrerer überwachender Systeme. Betrachtet man die zum Einsatz kommenden kooperierenden Knoten zur Überwachung eines Gebietes als einen virtuellen Sensor, dann bietet sich durch die Vernetzung dieser prinzipiell die Möglichkeit einer gleichzeitigen Überwachung mehrerer, möglicherweise weit verteilter Räume oder sogar ganzer Gebäude an.

## A LOKALES OBJEKTMANAGEMENT

### VERWENDETE GRÖSSEN

$\Omega_{a,k}$	Lokales Objekt $a \in \mathbb{N}$ zum Zeitpunkt $k$ .
$\Omega_{a,k}^{age} \in \mathbb{N}$	Alter des lokalen Objekts zum Zeitpunkt $k$ .
$N_k \in \mathbb{N}$	Anzahl der vorhandenen Aktivitätscluster zum Zeitpunkt $k$ .
${}^L N_k \in \mathbb{N}$	Anzahl der vorhandenen lokalen Objekte zum Zeitpunkt $k$ . Der Index $L$ dient dabei zur Unterscheidung der lokalen Objektanzahl von der Anzahl vorhandenen Aktivitätscluster $N_k$ .
$C_k^i$	Aktivitätscluster des Zeitpunkts $k$ .
$T_k^i \in \mathbb{N}$	Anzahl zusammenhängender Zeitpunkte von $C_k^i$ .
$O^{in} \in \mathbb{N}$	Anzahl zusammenhängender Zeitpunkte, welche zur Erzeugung eines lokalen Objekts notwendig sind.
$O^{out} \in \mathbb{N}$	Maximales Alter lokaler Objekte.
$\hat{P}_k^j \in \mathbb{R}^2$	Position der primären Partition des Aktivitätsclusters $C_k^i$ in der $x-z$ -Ebene des Koordinatensystems $K^{MCam}$ .
$\Delta_{k,l}^{i,j} \in \mathbb{R}^3$	Auf Positions-, Modulationsamplituden- und Intensitätsinformationen basierender Entfernungsvektor (81) zwischen $C_k^i$ und $C_l^j$ , welcher zur Berechnung der Ähnlichkeit in (82) verwendet wird.
$D_{k,l}^i = \{\Delta_{k,l}^{i,1}, \dots, \Delta_{k,l}^{i,N_l}\}$	Entfernungsmenge, welche die Entfernungen zwischen $C_k^i$ und den $N_l$ Aktivitätscluster des Zeitpunkts $l$ enthalten.
$S_{k,l}^{i,j} \in \mathbb{R}$	Similarität zwischen $C_k^i$ und $C_l^j$ .
$\Delta^{max} \in \mathbb{R}$	Schwellendistanz in der $x-z$ -Ebene des Koordinatensystems $K^{MCam}$ .

### BESTIMMUNG ZEITLICH ZUSAMMENHÄNGENDER AKTIVITÄTSCLUSTER

Der erste Schritt im Prozess des lokalen Objektmanagements, welcher für jeden Zeitpunkt durchzuführen ist, besteht in der Berechnung der Ähnlichkeit zwischen Aktivitätsclustern des momentanen sowie des vorangegangenen Zeitpunkts. Im Folgenden wird angenommen, dass der zu betrachtenden Zeitpunkt durch  $k$  bestimmt ist.

Zunächst werden diesbezüglich für jeden der  $N_k$  Aktivitätscluster die Entfernungen zu den Aktivitätsclustern des vorangegangenen Zeitpunkts, welche durch

$$\Delta_{k,k-1}^{i,j} = [\Delta_{k,k-1}^{i,j,1}, \Delta_{k,k-1}^{i,j,2}, \Delta_{k,k-1}^{i,j,3}]^T = [\|\hat{P}_k^i - \hat{P}_{k-1}^j\|_2, \|\mathbf{n}_k^i - \mathbf{n}_{k-1}^j\|_2, \|\mathbf{t}_k^i, \mathbf{t}_{k-1}^j\|_2]^T \in \mathbb{R}^3 \quad (81)$$

definiert werden, berechnet und in die zugehörige Entfernungsmenge  $D_{k,k-1}^i$ ,  $i=1, \dots, N_k$  eingetragen.

## A Lokales Objektmanagement

---

Die Komponenten von (81) sind im einzelnen:

1.  $\Delta_{k,k-1,1}^{i,j} = \|\hat{P}_k^i - \hat{P}_{k-1}^j\|_2$ : Abstände der primären Partitionen der Aktivitätscluster  $C_k^i$  und  $C_{k-1}^i$  in der  $x-z$ -Ebene von  $K^{MCam}$ . Durch Vernachlässigung der Höhenposition der Aktivitätscluster in der Berechnung von (81) wird die Fluktuation der Abstände  $\Delta_{k,k-1,1}^{i,j}$  reduziert, da im Allgemeinen die  $v$ -Komponenten der Positionsvektoren  ${}^{2D3D}P_k^i, {}^{2D3D}P_{k-1}^j$  der primären Partitionen, besonders im Falle entfernter Objekte, relativ große Schwankungen<sup>97</sup> aufweisen.
2.  $\Delta_{k,k-1,2}^{i,j} = \|\mathbf{n}_k^i - \mathbf{n}_{k-1}^j\|_2$ : Abstand der in der Parametermenge (45) der Aktivitätscluster  $C_k^i$  und  $C_{k-1}^i$  enthaltenen Modulationsamplitudenvektoren<sup>98</sup>.
3.  $\Delta_{k,k-1,3}^{i,j} = \|\mathbf{I}_k^i, \mathbf{I}_{k-1}^j\|_2$ : Abstand der in den Parametermengen (45) der entsprechenden Aktivitätscluster enthaltenen Intensitätsvektoren.

Eine fundamentale Annahme, welche getroffen wird, ist, dass eine Ähnlichkeit zwischen zwei Aktivitätscluster nur dann existiert, falls die Bedingung  $\Delta_{k,k-1,1}^{i,j} \leq \Delta^{max}$  erfüllt wird. Das Aufstellen dieser Forderung hat zur Folge, dass zur Berechnung einer Similarität zwischen zwei Aktivitätscluster eine festgelegte Nähe beider Aktivitätscluster notwendig ist, und führt dazu, dass nach Erstellung der Distanzmengen alle Elemente mit  $\Delta_{k,k-1,1}^{i,j} > \Delta^{max}$  entfernt werden.

Das Entfernen von Distanzvektoren aus der Distanzmenge der Aktivitätscluster führt dabei zu drei unterschiedlichen Fällen:

1.  $D_{k,k-1}^i = \emptyset$ : Es existieren keine Aktivitätscluster des vorausgegangenen Zeitpunkts mit einer zur Berechnung von Ähnlichkeiten hinreichenden Nähe zu  $C_k^i$ . Als Folge davon kann dem Aktivitätscluster  $C_k^i$  kein Vorgänger zugeordnet werden. Die Anzahl der zeitlich zusammenhängenden Zeitpunkte  $T_k^i$  von  $C_k^i$  wird auf 0 gesetzt.
2.  $|D_{k,k-1}^i| = 1$ : Aufgrund der Eindeutigkeit wird der entsprechende Aktivitätscluster des vorigen Zeitpunkts als Vorgänger angenommen<sup>99</sup>.
3.  $|D_{k,k-1}^i| > 1$ : Bedingt durch das Vorhandensein mehrerer potenzieller Vorgänger bedarf es einer zusätzlichen Untersuchung, welche im Folgenden diskutiert wird.

Eine zusätzlich getroffene Annahme ist, dass sich die Signatur der primären Partitionen<sup>100</sup> eines Objektes in zwei aufeinanderfolgenden Bildern nicht sprunghaft verändert. Eingebaut wird diese Annahme in die Ähnlichkeitsbildung (82) durch die Verwendung der entsprechenden Abstände  $\Delta_{k,k-1,2}^{i,j}$  und  $\Delta_{k,k-1,3}^{i,j}$ .

Für jeden der  $|D_{k,k-1}^i|$  Aktivitätscluster des vorausgegangenen Zeitpunkts mit einer zu  $C_k^i$  hinreichenden Nähe wird zunächst die Ähnlichkeit mittels

$$S_{k,k-1}^{i,j} = \frac{\alpha_1}{pos(\Delta_{k,k-1,1}^{i,j})} + \frac{\alpha_2}{pos(\Delta_{k,k-1,2}^{i,j})} + \frac{\alpha_3}{pos(\Delta_{k,k-1,3}^{i,j})}, \alpha_m \in [0,1], \sum_m \alpha_m = 1 \quad (82)$$

---

<sup>97</sup>Die Schwankungen dieser Partitionen rühren von der nicht deterministischen Natur der Erzeugung primärer Partitionen her. Besonders in entfernten Bereichen ergeben sich Variationen ihrer Höhen, welche durch die geringe Auflösung des PMD-Chips, besonders in  $v$ -Richtung, noch verstärkt werden.

<sup>98</sup>Die Definitionen der Modulationsamplituden und Intensitätsvektoren finden sich in Abschnitt 4.2.4.2.

<sup>99</sup>Der Term  $|D_{k,k-1}^i|$  entspricht der Kardinalität der Menge.

<sup>100</sup>Als Signatur werden in diesem Kontext die in der primären Partition enthaltenen Modulationsamplituden und Intensitäten interpretiert.

## A Lokales Objektmanagement

berechnet, wobei der Aktivitätscluster des vorausgegangenen Zeitpunkts mit der höchsten Similarität als Vorgänger von  $C_k^i$  ausgewählt wird. Die Hilfsfunktion in den Nennern von (82) dient der Ermittlung der Position eines beliebigen Elementes einer Menge, deren Elemente nach Größe absteigend geordnet sind. Übertragen auf die Berechnung der Ähnlichkeit zweier Aktivitätscluster liefert  $pos(\Delta_{k,k-1}^{i,j})$  die Position von  $\Delta_{k,k-1}^{i,j}$  aus der Menge, welche alle ersten Komponenten von  $\Delta_{k,l}^{i,j}$  in ihrer nach Größe absteigenden Reihenfolge enthält. Die Anzahl der Komponenten ist dabei gleich  $|D_{k,k-1}^i|$ . Gleiches gilt für  $pos(\Delta_{k,k-1}^{i,j,2})$  und  $pos(\Delta_{k,k-1}^{i,j,3})$ , welche auf den jeweils durch die entsprechenden Komponenten von  $\Delta_{k,l}^{i,j}$  erzeugten Mengen operieren.

Im letzten Schritt des Prozesses der Bestimmung zeitlich zusammenhängender Aktivitätscluster wird untersucht, ob ein Aktivitätscluster des vorangegangenen Zeitpunktes mehrfach als ein Vorgänger bestimmt worden ist, was im Falle neu auftauchender oder mehrerer naher Aktivitätscluster möglich sein kann<sup>101</sup>. Sind keine Mehrfachzuweisungen vorhanden, wird die Anzahl der zusammenhängenden Zeitpunkte  $T_{k-1}^j$  des Vorgängers  $C_{k-1}^j$  übernommen und nachträglich inkrementiert.

Im Falle von vorhandenen Mehrfachzuweisungen wird für jeden der mehrfach zugewiesenen Vorgänger eine Nachfolgerbestimmung durchgeführt, was dem obigen Prozess der Vorgängerfindung mit vertauschten Rollen entspricht<sup>102</sup>. Nach Auswahl des Nachfolgers mit der größten Similarität (82) wird dessen Zähler zusammenhängender Zeitpunkte entsprechend modifiziert. Den Ablauf der Bestimmung räumlich zusammenhängender Aktivitätscluster liefert Abb. 95.

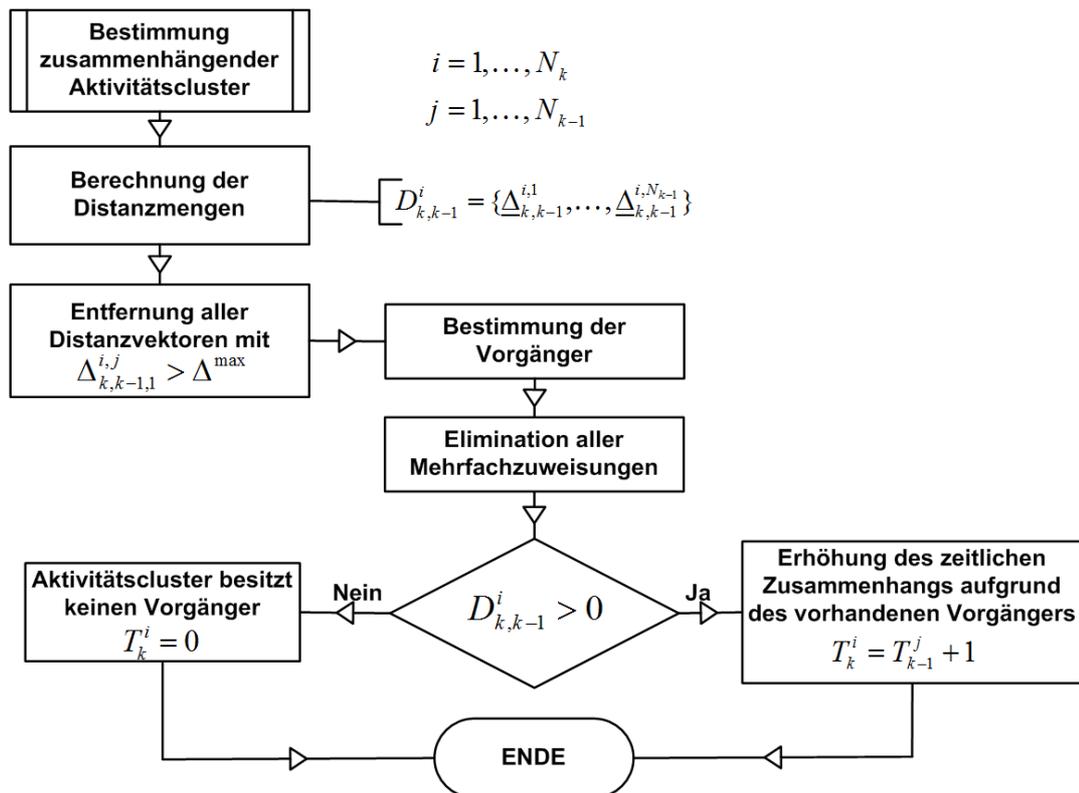


Abbildung 95: Prozess zur Bestimmung zeitlich zusammenhängender Aktivitätscluster.

<sup>101</sup>Dieses gilt insbesondere für die Wahl eines großen Wertes für  $\Delta^{max}$ .

<sup>102</sup>Für einen mehrfach zugewiesenen Vorgänger  $C_{k-1}^j$  setzt sich seine Distanzliste  $D_{k-1,k}^j$  aus seinen potenziellen Nachfolgern zusammen.

## A Lokales Objektmanagement

Im Anschluss an die Bestimmung zeitlich zusammenhängender Aktivitätscluster findet eine Unterscheidung zweier Fälle statt:

1.  ${}^L N_{k-1} = 0$ : Es existieren keine zu vorausgegangenen Zeitpunkten erzeugte lokale Objekte.
2.  ${}^L N_{k-1} > 0$ : Existenz lokaler Objekte, welche zu früheren Zeitpunkten erzeugt worden sind.

Während es im ersten Fall direkt zum Prozess der Erzeugung lokaler Objekte kommt, wird im zweiten Fall zunächst der Vorgang der Erhaltung sowie Entfernung vorgeschaltet.

### LOKALE OBJEKTERZEUGUNG

Als Basisinformationen der lokalen Objekterzeugung dienen die in  $T_k^i$  festgehaltenen Längen der zeitlichen Zusammenhänge der  $N_k$  Aktivitätscluster. Erfüllt ein Aktivitätscluster  $C_k^i$  die Forderung  $T_k^i \geq O^{in}$ , dann führt das zur Erzeugung eines eindeutigen lokalen Objekts  $\Omega_{a,k}$  mit  $a \in \mathbb{N}$  und  $\Omega_{a,k}^{age} = 0$  auf Grundlage von  $C_k^i$  (Abb. 96). Für den das Objekt erzeugende Cluster  $C_k^i$  gilt, dass dieser eine Kennung erhält, welche seine Zuweisung zu dem Objekt  $\Omega_{a,k}$  angibt.

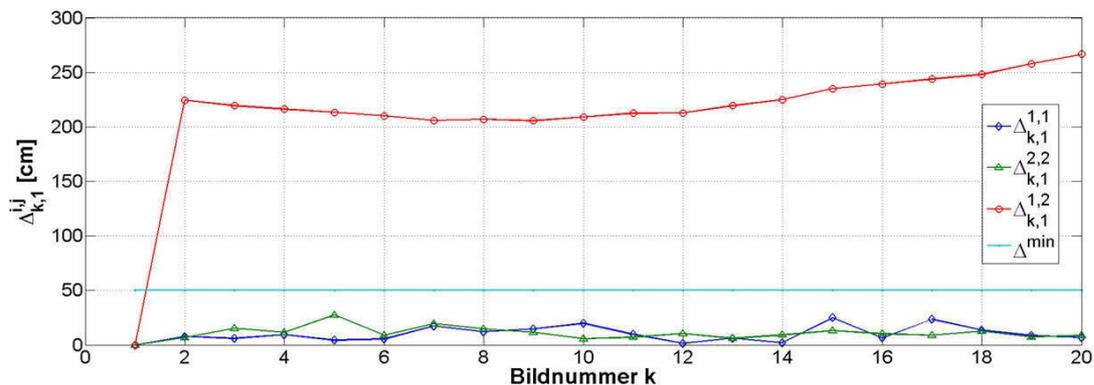
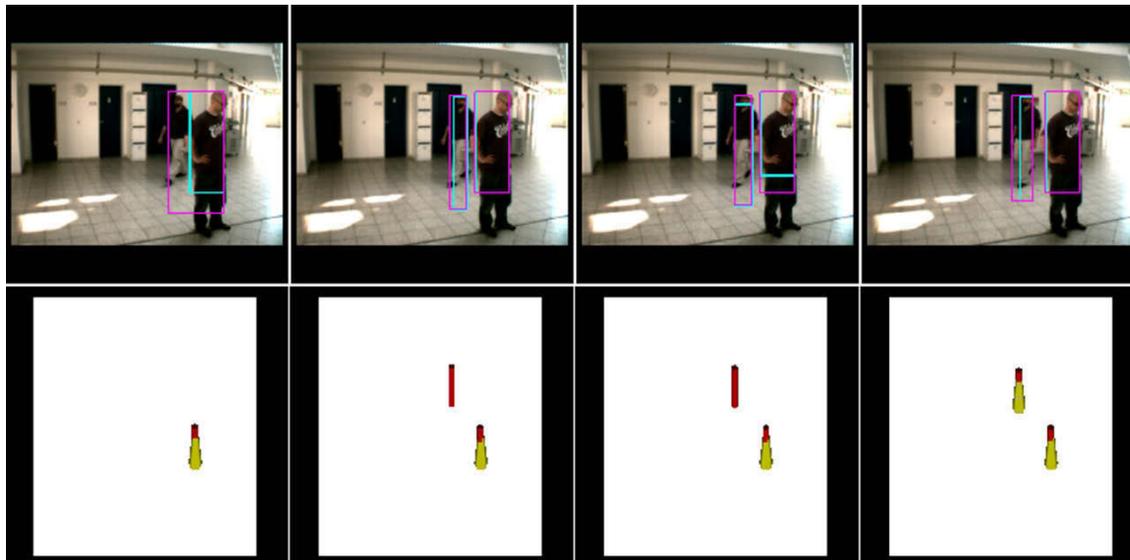


Abbildung 96: Lokale Objektgeneration II.

Beginn einer Assoziationssequenz durch Wiedereintritt eines Objekts (nach Verdeckung) in den Detektionsbereich der MultiCam. Im unteren Graphen sind die Entwicklungen der Abstände der einzelnen Aktivitätscluster der Sequenz in der  $x$ - $z$ -Ebene von  $K^{MCam}$  abgebildet. Parameter:  $O^{in} = 2$ ,  $\Delta^{max} = 50$ ,  $\alpha_1 = 0,5$ ,  $\alpha_2 = 0,3$ ,  $\alpha_3 = 0,2$ .

## A Lokales Objektmanagement

---

### OBJEKTERHALTUNG UND OBJEKTENTFERNUNG

Im ersten Schritt werden zunächst alle Aktivitätscluster des Zeitpunkts  $k$  betrachtet, für die ein Vorgänger gefunden werden konnte. Ist der entsprechende Vorgänger im Rahmen des lokalen Objektmanagements des vorausgegangenen Zeitpunkts bereits einem lokalen Objekt zugewiesen worden, wird der Nachfolger als erhaltender Aktivitätscluster interpretiert und ebenfalls demselben lokalen Objekt zugewiesen. Die Informationen (Position, Geschwindigkeit) werden dem Kalman-Filter des lokalen Objekts übergeben, welcher die Statusinformationen gemäß der neu erhaltenen Daten (Messungen) zum Zeitpunkt  $k$  auffrischt ( $\Omega_{a,k-1} \rightarrow \Omega_{a,k}$ , Abschnitt 5.2.3). Gleichzeitig wird das Alter des Objekts auf null zurückgesetzt.

Nach Betrachtung aller Aktivitätscluster mit Vorgänger und der sich eventuell daraus ergebenden direkten Erhaltung lokaler Objekte erfolgt eine zusätzliche Unterscheidung zweier Fälle:

1. Alle Objekte des vorausgegangenen Zeitpunkts sind durch Aktivitätscluster des Zeitpunkts  $k$  direkt erhalten worden und entsprechen damit lokalen Objekten des Zeitpunkts  $k$ . Somit kann Fall  ${}^L N_{k-1} = 0$  angenommen werden. Aufgrund der Erhaltung aller lokalen Objekte wird der Prozess der Entfernung übersprungen und direkt zur Objekterzeugung übergegangen. Da einige Aktivitätscluster des Zeitpunkts  $k$  bereits zur direkten Erhaltung von lokalen Objekten verwendet wurden, operiert die Objekterhaltung lediglich auf der Menge der Aktivitätscluster, deren Informationen nicht verwendet wurden.
2. Es existieren lokale Objekte des vorausgegangenen Zeitpunkts  $k-1$ , welche keine direkte Erhaltung durch Aktivitätscluster des Zeitpunkts  $k$  erfahren haben. Gründe hierfür sind z. B. Fluktuationen in der Erzeugung der Aktivitätscluster, Verlassen des Sichtbereiches der Kamera sowie gegenseitige Verdeckung der realen Objekte. Der Diskussion dieses Falles dienen die folgenden Abschnitte.

Angenommen wird, dass die direkte Objekterhaltung bereits stattgefunden hat. Für die weitere Betrachtung sind damit lediglich die Aktivitätscluster des Zeitpunkts  $k$ , welche nicht zur direkten Objekterhaltung beigetragen haben sowie lokale Objekte des Zeitpunkts  $k-1$  ohne erhaltenden Aktivitätscluster, von Interesse.

Durchgeführt wird zunächst eine Bestimmung zusammenhängender Objekte und Aktivitätscluster. Im Wesentlichen handelt es sich dabei um eine leichte Variation des Prozesses, welcher bereits zur Bestimmung zeitlich zusammenhängender Aktivitätscluster zu Beginn des lokalen Objektmanagements verwendet wurde. Die Änderung ist, dass als Menge der Aktivitätscluster des Zeitpunkts  $k$  jetzt lediglich die Aktivitätscluster verwendet werden, welche nicht zu einer direkten Objekterhaltung beigetragen haben. Als Menge der Aktivitätscluster des Zeitpunkts  $k-1$  dienen die lokalen Objekte des vorausgegangenen Zeitpunkts ohne einen sie erhaltenden Aktivitätscluster. Bedingt durch das Fehlen der Parametermenge (46) in lokalen Objekten reduziert sich der Distanzvektor (81) zu einem Skalar, der gleich der ersten Komponente des Vektors ist:

$$\underline{\Delta}_{k,k-1}^{i,j} = \Delta_{k,k-1,1}^{i,j}. \quad (83)$$

Die in (82) angegebene Ähnlichkeitsberechnung ändert sich damit zu:

$$S_{k,k-1}^{i,j} = \text{pos}(\Delta_{k,k-1,1}^{i,j})^{-1}. \quad (84)$$

Als Positionsinformation des lokalen Objekts werden die entsprechenden Werte des zugehörigen Kalman-Filters verwendet.

## A Lokales Objektmanagement

Da die gefundenen Vorgänger lokalen Objekten entsprechen, erhalten diese durch den entsprechenden Aktivitätscluster eine sogenannte indirekte Erhaltung. Eine Zuweisung des erhaltenden Aktivitätsclusters zum erhaltenden Objekt sowie das Rücksetzen des Objektalters auf null erfolgt zusätzlich.

Für lokale Objekte, denen kein Nachfolger bzw. einen sie indirekt erhaltenden Aktivitätscluster zugewiesen werden konnte, gilt, dass ihr Alter inkrementiert wird. Erreicht das Alter einen durch  $O^{out}$  bestimmten Wert, führt das zur Entfernung des lokalen Objekts.

Da Objekte mit einem Alterungszähler  $0 < \Omega_a^{age} < O^{out}$  keine auf Messungen basierenden Informationen zur Auffrischung ihres Zustands zur Verfügung haben, wird in diesem Falle die Information aus dem Prädiktionsschritt (73) des Filters des zugehörigen Kalman-Filters als Messung benutzt, um eine Auffrischung des Objektzustandes zu erhalten (Abschnitt 5.2.3).

Existieren zu diesem Zeitpunkt noch Aktivitätscluster, welche keinem Objekt zugeordnet worden sind, werden diese zu einer Menge zusammengefasst, welche die Informationsgrundlage des nachgeschalteten Objekterzeugungsprozesses bildet. Den kompletten Prozess des lokalen Objektmanagements liefert Abb. 97.

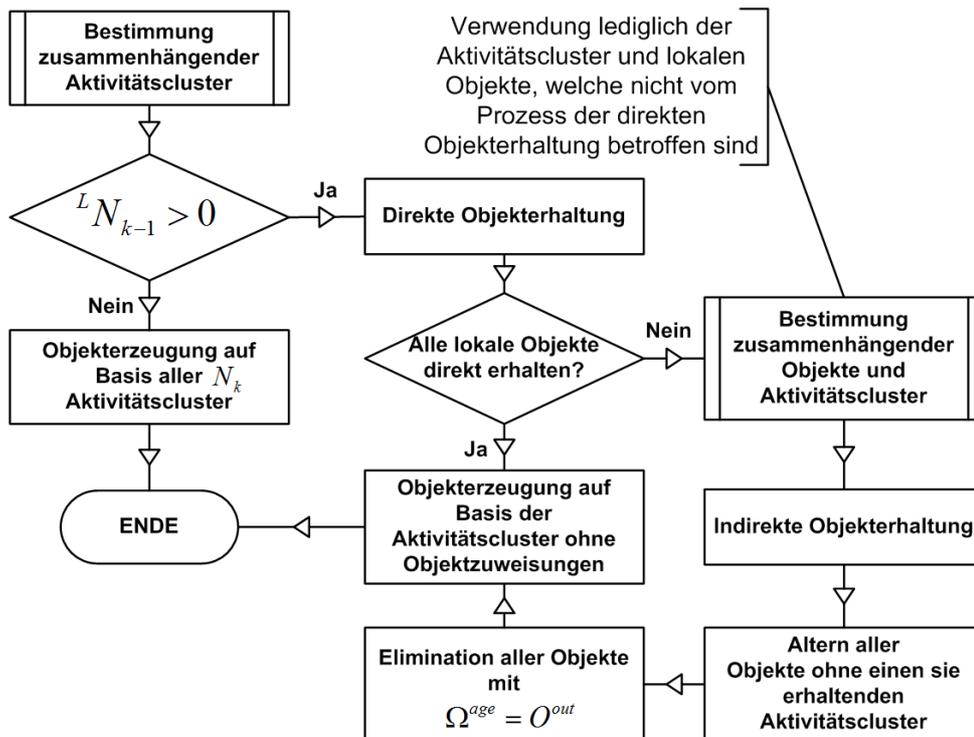


Abbildung 97: Prozess des lokalen Objektmanagements.

## B GLOBALES OBJEKTMANAGEMENT

### VERWENDETE GRÖSSEN

$N_L \in \mathbb{N}$	Anzahl der vorhandenen MSK.
$N_{n,k}, n=1, \dots, N_L$	Anzahl der lokalen Objekte des MSK $n$ zum Zeitpunkt $k$ .
${}^G N_k \in \mathbb{N}$	Anzahl globaler Objekte zum Zeitpunkt $k$ .
$P_k^n = \{ \rho_{1,k}^n, \dots, \rho_{N_{n,k}}^n \}$	Menge der $N_{n,k}$ Positionsvektoren der lokalen Objekte des MSK $n$ zum Zeitpunkt $k$ in der $x-z$ -Ebene des Koordinatensystems $K^{Glob}$ .
${}^G \Omega_{a,k} = \{ {}^G \Xi_{a,k}, {}^G O_{a,k}, {}^G \beta_a, \Pi_a \}$	Globales Objekt $a \in \mathbb{N}$ zum Zeitpunkt $k$ . Zusammengesetzt wird das globale Objekt aus dem Tupel, bestehend aus der Position ${}^G \Xi_{a,k}$ in der $x-z$ -Ebene von $K^{Glob}$ , der Konstituentenmenge ${}^G O_{a,k}$ , der Zeitindexvariable ${}^G \beta_a \in \mathbb{N}$ sowie dem zur Positionsbestimmung zuständigen Partikelfilter $\Pi_a$ .
$\Omega_{i,k}^n$	Lokales Objekt $i$ des MSK $n$ zum Zeitpunkt $k$ .
$\lambda_{i,j,k}^{n,m} \in \mathbb{R}$	Distanz zwischen $\Omega_{i,k}^n$ , dem Quellobjekt und $\Omega_{j,k}^m$ , dem Zielobjekt in der $x-z$ -Ebene von $K^{Glob}$ zum Zeitpunkt $k$ . Aufgrund der Symmetrie der Distanzberechnung gilt: $\lambda_{i,j,k}^{n,m} = \lambda_{j,i,k}^{m,n}$ .
$\Lambda_k^{n,m}$	Distanzmenge der MSK $n$ und $m$ . Enthält alle Distanzen zwischen den lokalen Objekten zweier MSK zum Zeitpunkt $k$ .
$\hat{\lambda}_{i,k}^{n,a} \in \mathbb{R}$	Distanz zwischen dem lokalen Objekt $\Omega_{i,k}^n$ und dem globalen Objekt ${}^G \Omega_{a,k}$ zum Zeitpunkt $k$ in der $x-z$ -Ebene des Koordinatensystems $K^{Glob}$ .
$\Lambda_k$	Menge aller gültigen Distanzen zwischen lokalen und globalen Objekten zum Zeitpunkt $k$ .
${}^G \Delta^{max} \in \mathbb{R}$	Maximale Entfernung zwischen Konstituenten sowie Radius des globalen Objekts.

### ANNAHMEN

1. Jedes reale Objekt wird in einem MSK lediglich durch ein lokales Objekt repräsentiert. Daraus folgt, dass bei  $N_L$  MSK maximal  $N_L$  Repräsentationen eines realen Objekts auf dem FSK vorhanden sind.
2. Informationsgrundlage des Objektmanagements sind die Positionen der durch die MSK lokal erzeugten Objekte in der  $x-z$ -Ebene des globalen Koordinatensystems  $K^{Glob}$ . Definiert wird jedes globale Objekt durch räumlich nahe oder sich sogar teilweise überlappende lokale Objekte unterschiedlicher MSK, welche als seine Konstituenten interpretiert werden. Die notwendige räumliche Nähe unterschiedlicher Konstituenten ist dabei durch die Größe  ${}^G \Delta^{max}$  definiert. Aufgrund der ersten Annahme ist die maximale Konstituentenzahl durch  $N_L$  begrenzt.

## B Globales Objektmanagement

---

3. Modelliert wird ein globales Objekt durch ein sich in der  $x-z$ -Ebene von  $K^{Glob}$  bewegender Zylinder mit Radius  ${}^G\Delta^{max}$ , wobei der Mittelpunkt des Zylinders in der  $x-z$ -Ebene der Position des globalen Objekts entspricht.

### BERECHNUNG DER DISTANZINFORMATIONEN

Der initiale Schritt im Rahmen des globalen Objektmanagements zum Zeitpunkt  $k$  besteht in der Berechnung sämtlicher Distanzmengen. Beispielfhaft ergibt sich diese für die MSK  $n$  und  $m$  zu

$$\Lambda_k^{n,m} = \{\lambda_{1,1,k}^{n,m}, \dots, \lambda_{1,N_{m,k},k}^{n,m}, \lambda_{2,1,k}^{n,m}, \dots, \lambda_{N_{n,k},N_{m,k},k}^{n,m}\} \text{ mit } \lambda_{i,j,k}^{n,m} = \left\| \rho_{i,k}^n - \rho_{j,k}^m \right\|_2, \quad n \neq m. \quad (85)$$

Naturgemäß ist die Bestimmung der Abstände für Objekte desselben MSK nicht notwendig, da davon ausgegangen wird, dass es sich hierbei bereits um unterschiedliche reale Objekte handelt, welche nicht zu einem globalen Objekt zusammengefasst werden können.

Der zweite Schritt dient der Bestimmung der potenziellen Zusammengehörigkeit von lokalen Objekten unterschiedlicher MSK. Dazu werden die Elemente aller erzeugten Distanzmengen (85) der MSK mit der Schwellendistanz  ${}^G\Delta^{max}$  verglichen, wobei Abstände, welche die Bedingung

$$\lambda_{i,j,k}^{n,m} < {}^G\Delta^{max} \quad (86)$$

nicht erfüllen, aus der entsprechenden Menge entfernt werden. Für jeden der  $N_L$  MSK gilt im Anschluss an die Distanzfilterung gemäß (86), dass seine  $N_L - 1$  Distanzmengen zu diesem Zeitpunkt nur noch die Elemente beinhalten, die den lokalen Objekten des MSK entsprechen, welche eine durch  ${}^G\Delta^{max}$  bestimmte Nähe zu den lokalen Objekten anderer MSK besitzen. Zu beachten ist, dass durch die Distanzfilterung die Möglichkeit besteht, dass sämtliche Distanzelemente, welche einem lokalen Objekt zugeordnet werden können, aus den Distanzmengen entfernt werden. Lokale Objekte ohne zugehörige Distanzelemente werden dabei als isolierte lokale Objekte betrachtet. Ihre Behandlung wird im Rahmen der globalen Objekterhaltung diskutiert. Im Anschluss zur Distanzfilterung werden zwei grundlegende Fälle unterschieden:

1.  ${}^GN_{k-1} = 0$ : Es existieren keine globalen Objekte und somit auch keine Vorinformationen, welche zum Zeitpunkt  $k$  verwendet werden können. Ein Umstand, der direkt zum Prozess der globalen Objekterzeugung führt.
2.  ${}^GN_{k-1} > 0$ : Die Informationen bezüglich bereits vorhandener globaler Objekte müssen mit einbezogen werden, was zunächst die Prozesse der Erhaltung und Entfernung globaler Objekte einleitet und welche dem Erzeugungsprozess vorgeschaltet sind.

### GLOBALE OBJEKTERZEUGUNG

Der erste Schritt diesbezüglich besteht in der Untersuchung aller Distanzmengen (85) auf Mehrfachzuweisungen. Existieren in einer beliebigen Distanzliste, z. B.  $\Lambda_k^{n,m}$ , mehrere Elemente mit gleichen Quellobjekten, bedeutet dieses, dass mehrere Objekte des MSK  $m$  eine der Bedingung (86) erfüllenden Nähe zu einem Objekt des MSK  $n$  besitzen. Aufgrund der Annahme, dass ein globales Objekt von jedem MSK maximal einen Konstituenten besitzen darf, wird lediglich das Element aus  $\Lambda_k^{n,m}$  mit dem geringsten Wert behalten. Wichtig ist, dass aus Symmetriegründen neben den aus  $\Lambda_k^{n,m}$  entfernten Elementen auch die entsprechenden Elemente aus  $\Lambda_k^{m,n}$  gelöscht

---

## B Globales Objektmanagement

werden müssen. Durch die Entfernung von Mehrfachzuweisungen besteht die Möglichkeit, dass es zur Erzeugung zusätzlicher isolierter lokaler Objekte kommt.

Der nächste durchzuführende Schritt beinhaltet die Erzeugung globaler Objekte auf Basis der von Mehrfachzuweisungen bereinigten Distanzmengen. Aufgrund der redundant vorliegenden Informationen<sup>103</sup> ist dabei lediglich die Behandlung der Distanzmengen  $\Lambda_k^{n,m}$ ,  $n < m$ , notwendig. Begonnen wird mit der ersten zu betrachtenden Distanzmenge  $\Lambda_k^{1,2}$ . Für jedes Element aus  $\Lambda_k^{1,2}$  wird zunächst ein globales Objekt erzeugt und als Konstituenten werden die entsprechenden Quell- und Zielobjekte eingetragen<sup>104</sup>.

Für jedes Element der restlichen Distanzmengen wird danach überprüft, ob das entsprechende Quell- oder Zielobjekt bereits in der Konstituentenmenge eines erzeugten Objektes vorhanden ist. Trifft dieses zu, wird jeweils das MSK-Objekt des Distanzelements, welches nicht in der Konstituentenmenge des Objekts enthalten ist, als ein Konstituent hinzugefügt. Für den Fall, dass weder Quell- noch Zielobjekt eines Distanzelements in den Konstituentenmengen der erzeugten Objekte gefunden werden, wird ein weiteres globales Objekt erzeugt, wobei als dessen Konstituenten die Quell- und Zielobjekte des jeweiligen Distanzelements eingetragen werden.

Der letzte Schritt im Rahmen der globalen Objekterzeugung überprüft, ob im Prozess der Entfernungsfiterung sowie der Elimination von Mehrfachzuweisungen isolierte lokale Objekte erzeugt worden sind. Ist dieses der Fall, wird für jedes dieser Objekte ein globales erzeugt, wobei die Konstituentenmenge des globalen Objekts lediglich das isolierte lokale Objekt enthält.

### GLOBALE ERHALTUNG UND ENTFERNUNG VON OBJEKTEN

Begonnen wird mit der Überprüfung, ob MSK-Objekte des gegenwärtigen Zeitpunkts  $k$  bereits in den Konstituentenmengen der vorhandenen globalen Objekte zu finden sind. Alle erneut gefundenen lokalen Objekte werden dabei in die den globalen Objekten zugehörigen Konstituentenmengen übertragen und spielen im weiteren Verlauf des Prozesses der Erhaltung keine Rolle mehr. Die Reduktion der zu betrachtenden MSK-Objekte ist möglich, da bereits im Vorfeld eine Erzeugung bzw. Erhaltung eines globalen Objekts auf Basis der entsprechenden MSK-Objekte durchgeführt worden ist, sodass diese erneut vorhandenen Konstituenten direkt übernommen werden können und somit der Erhaltung eines globalen Objekts dienen. Nach Beendigung dieses direkten Erhaltungsschrittes werden aus allen Konstituentenmengen der vorhandenen globalen Objekte die Konstituenten des vorausgegangenen Zeitpunktes entfernt.

Im Anschluss daran werden für alle verbliebenen MSK-Objekte die Distanzen

$$\hat{\lambda}_{p,k}^{n,a} = \left\| \rho_{p,k}^n - \Xi_{a,k-1} \right\|_2 \quad (87)$$

zu allen globalen Objekten berechnet, wobei sämtliche Distanzen, welche die Entfernung  ${}^G\Delta^{max}$  unterschreiten, in  $\Lambda_k$  gespeichert werden.  $\Lambda_k$  enthält somit Informationen über MSK-Objekte des Zeitpunkts  $k$ , welche zu den vorhandenen globalen Objekten eine hinreichende Nähe vorweisen können und eventuell als Konstituenten infrage kommen.

Der Berechnung der Distanzen folgt eine Untersuchung auf Mehrfachzuweisungen. Falls ein lokales Objekt mehreren globalen Objekten zuweisbar ist, wird das globale Objekt mit der geringsten Entfernung ausgesucht und das lokale Objekt in die entsprechende Konstituentenmenge eingetragen. Zusätzlich wird untersucht, ob mehrere MSK-Objekte eines MSK dem gleichen globalen Objekt zugeordnet worden sind. Ist dieses der Fall, wird das lokale Objekt mit der geringsten Entfernung ausgewählt und in die Konstituentenmenge eingetragen.

<sup>103</sup>Aufgrund ihrer Symmetrie sind die Informationen aus  $\Lambda_k^{n,m}$  ebenfalls in  $\Lambda_k^{m,n}$  enthalten.

<sup>104</sup>Es sei daran erinnert, dass durch die Indizierung der Elemente einer Distanzmenge die jeweiligen Quell- und Zielobjekte eindeutig bestimmt sind.

## B Globales Objektmanagement

Für alle globalen Objekte, deren Konstituentenmenge eine Auffrischung durch MSK-Objekte des Zeitpunkts  $k$  erfahren haben, wird im Anschluss eine Neubestimmung ihrer Objektposition durch den dem globalen Objekt zugewiesenen Partikelfilter durchgeführt (Abschnitt 5.3.4.2).

In dem sich anschließenden Objektentfernungsprozess werden alle globalen Objekte, deren Konstituentenmenge keine sie erhaltenden MSK-Objekte des gegenwärtigen Zeitpunkts vorweisen können, entfernt. Globale Objekte besitzen somit im Vergleich zu den lokal auf einem MSK erzeugten keinen Alterungsprozess.

Da der globale Erhaltungs- und Eliminationsprozess alle notwendigen Informationen bzgl. des vorangegangenen Zeitpunkts bereits verwendet hat, kann an dieser Stelle von der Annahme  ${}^G N_{k-1} = 0$  ausgegangen werden. Auf Basis der Menge der MSK-Objekte, welche noch nicht globalen Objekten zugeordnet werden konnten, findet im Anschluss der Prozess der globalen Objekterzeugung statt. Einen kompakten Ablauf des globalen Objektmanagements liefert Abb. 98.

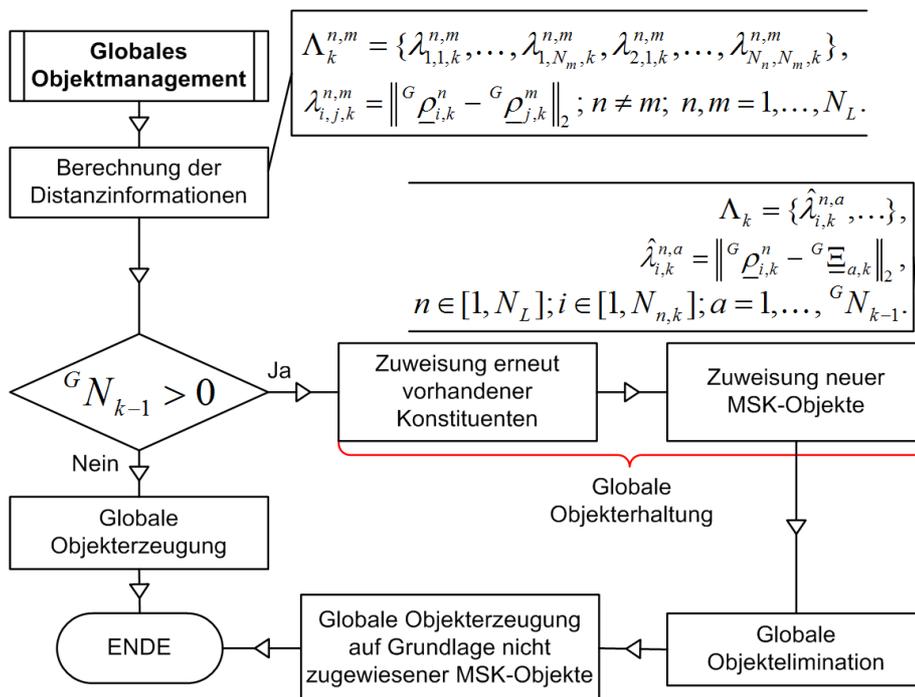


Abbildung 98: Globales Objektmanagement.

## C BAYES'SCHE ESTIMATION

Das in diesem Abschnitt gewählte Vorgehen basiert in seiner Form im Wesentlichen auf den Büchern von Gordon [183] und Doucet [184]. Für eine detailliertere Beschreibung der bayes'schen Estimation sei an dieser Stelle auf diese Werke verwiesen.

### VERWENDETE GRÖSSEN

$\underline{x}_k, \xi_k$	Zustandsvektor $\underline{x}_k$ des Zeitpunkts $k$ mit zugehöriger Dummyvariable $\xi_k$ .
$\underline{y}_k, \zeta_k$	Messvektor $\underline{y}_k$ des Zeitpunkts $k$ mit zugehöriger Dummyvariable $\zeta_k$ .
$Y_k = \{\underline{y}_1, \dots, \underline{y}_k\},$ $Z_k = \{\zeta_1, \dots, \zeta_k\}$	Menge aller bis zum Zeitpunkt $k$ vorhandener Messvektoren sowie Menge aller bis zum Zeitpunkt $k$ vorhandener Dummymessvektoren.
$X_k = \{\underline{x}_0, \dots, \underline{x}_k\},$ $\Xi_k = \{\xi_0, \dots, \xi_k\}$	Menge aller bis zum Zeitpunkt $k$ vorhandener Zustandsvektoren sowie Menge aller bis zum Zeitpunkt $k$ vorhandener Dummyzustandsvektoren.
$f_{\underline{x}_0 Y_0}(\xi_0 Z_0) = f_{\underline{x}_0}(\xi_0)$	Wahrscheinlichkeitsdichte des initialen Zustands.
$f_{\underline{x}_k \underline{x}_{k-1}, Y_{k-1}}(\xi_k \xi_{k-1}, Z_{k-1})$	Übergangswahrscheinlichkeitsdichte.
$f_{\underline{x}_k \underline{x}_{k-1}}(\xi_k \xi_{k-1})$	Übergangswahrscheinlichkeitsdichte (Markov-Prozess).
$f_{\underline{y}_k \underline{x}_k}(\zeta_k \xi_k)$	Messwahrscheinlichkeitsdichte, likelihood-Dichte.
$f_{\underline{x}_{k-1} Y_{k-1}}(\xi_{k-1} Z_{k-1})$	Wahrscheinlichkeitsdichte des Zustands zum Zeitpunkt $k-1$ ( <i>prior</i> ).
$f_{\underline{x}_k Y_{k-1}}(\xi_k Z_{k-1})$	Dynamischer Prior ( <i>dynamic prior</i> [183]). Prädiktion des Systemzustands.
$f_{\underline{x}_k Y_k}(\xi_k Z_k)$	Wahrscheinlichkeitsdichte des Zustands zum Zeitpunkt $k$ ( <i>posterior</i> ).

### ANNAHMEN

1. Vorhandensein eines initialen Systemzustands  $f_{\underline{x}_0|Y_0}(\xi_0|Z_0) = f_{\underline{x}_0}(\xi_0)$ .
2. Übergangswahrscheinlichkeitsdichte entspricht einem Markov-Prozess erster Ordnung:

$$f_{\underline{x}_k|\underline{x}_{k-1}, Y_{k-1}}(\xi_k|\xi_{k-1}, Z_{k-1}) = f_{\underline{x}_k|\underline{x}_{k-1}}(\xi_k|\xi_{k-1}). \quad (88)$$

3. Gegenseitig unabhängige Messungen:

$$f_{Y_k|X_k}(Z_k|\Xi_k) = \prod_{i=1}^k f_{\underline{y}_i|\underline{x}_i}(\zeta_i|\xi_i). \quad (89)$$

## C Bayes'sche Estimation

---

### ESTIMATIONSPROZESS

Ziel der bayes'schen Estimation ist die Berechnung der konditionierten Wahrscheinlichkeitsdichte  $f_{\underline{x}_k|Y_k}(\xi_k|Z_k)$  des Systemzustands  $\underline{x}_k$  unter Verwendung sämtlicher Informationen bzw. Messungen bis zum Zeitpunkt  $k$  [183].

Gegeben sei zunächst die Wahrscheinlichkeitsdichte des Zustands  $f_{\underline{x}_{k-1}|Y_{k-1}}(\xi_{k-1}|Z_{k-1})$  zum Zeitpunkt  $k-1$ . Unter Verwendung der Chapman-Kolmogorov-Gleichung [183], [199], wird durch

$$\begin{aligned} f_{\underline{x}_k|Y_{k-1}}(\xi_k|Z_{k-1}) &= \int f_{\underline{x}_k|\underline{x}_{k-1}, Y_{k-1}}(\xi_k|\xi_{k-1}, Z_{k-1}) f_{\underline{x}_{k-1}|Y_{k-1}}(\xi_{k-1}|Z_{k-1}) d\xi_{k-1} \\ &= \int f_{\underline{x}_k|\underline{x}_{k-1}}(\xi_k|\xi_{k-1}) f_{\underline{x}_{k-1}|Y_{k-1}}(\xi_{k-1}|Z_{k-1}) d\xi_{k-1} \end{aligned} \quad (90)$$

zunächst eine Prädiktion des Systemzustands berechnet.

Im Falle der Verfügbarkeit von  $y_k$  schließt sich ein Auffrischungsschritt an, welcher die gemessenen Informationen bezüglich des Systemzustands einbindet und die Estimation

$$f_{\underline{x}_k|Y_k}(\xi_k|Z_k) = \frac{f_{y_k|\underline{x}_k}(\xi_k|\xi_k) f_{\underline{x}_k|Y_{k-1}}(\xi_k|Z_{k-1})}{f_{y_k|Y_{k-1}}(\xi_k|Z_{k-1})} \quad (91)$$

abschließt. Der normalisierende Nenner in (91) ergibt sich umgeformt zu

$$f_{y_k|Y_{k-1}}(\xi_k|Z_{k-1}) = \int f_{y_k|\underline{x}_k}(\xi_k|\xi_k) f_{\underline{x}_k|Y_{k-1}}(\xi_k|Z_{k-1}) d\xi_k. \quad (92)$$

Zum Beweis von (91) sei zunächst  $f_{\underline{x}_k|Y_k}(\xi_k|Z_k)$  unter Anwendung der Bayes-Regel umgeformt zu

$$f_{\underline{x}_k|Y_k}(\xi_k|Z_k) = \frac{f_{Y_k|\underline{x}_k}(Z_k|\xi_k) f_{\underline{x}_k}(\xi_k)}{f_{Y_k}(Z_k)}. \quad (93)$$

Durch Separation

$$\frac{f_{Y_k|\underline{x}_k}(Z_k|\xi_k) f_{\underline{x}_k}(\xi_k)}{f_{Y_k}(Z_k)} = \frac{f_{y_k, Y_{k-1}|\underline{x}_k}(\xi_k, Z_{k-1}|\xi_k) f_{\underline{x}_k}(\xi_k)}{f_{y_k, Y_{k-1}}(\xi_k, Z_{k-1})} \quad (94)$$

und anschließender Faktorisierung der zusammengesetzten Wahrscheinlichkeitsdichten

$$\frac{f_{y_k, Y_{k-1}|\underline{x}_k}(\xi_k, Z_{k-1}|\xi_k) f_{\underline{x}_k}(\xi_k)}{f_{y_k, Y_{k-1}}(\xi_k, Z_{k-1})} = \frac{f_{y_k|\underline{x}_k, Y_{k-1}, \underline{x}_k}(\xi_k|Z_{k-1}, \xi_k) f_{Y_{k-1}|\underline{x}_k}(Z_{k-1}|\xi_k) f_{\underline{x}_k}(\xi_k)}{f_{y_k|Y_{k-1}}(\xi_k|Z_{k-1}) f_{Y_{k-1}}(Z_{k-1})} \quad (95)$$

## C Bayes'sche Estimation

---

und unter erneuter Anwendung der Bayes-Regel ergibt sich (93) zunächst zu

$$f_{x_k|Y_k}(\xi_k | Z_k) = \frac{f_{y_k|Y_{k-1}, x_k}(\xi_k | Z_{k-1}, \xi_k) f_{x_k|Y_{k-1}}(\xi_k | Z_{k-1}) f_{Y_{k-1}}(Z_{k-1}) f_{x_k}(\xi_k)}{f_{y_k|Y_{k-1}}(\xi_k | Z_{k-1}) f_{Y_{k-1}}(Z_{k-1}) f_{x_k}(\xi_k)}. \quad (96)$$

Unter Verwendung der Annahme von gegenseitig unabhängigen Messungen und der Elimination redundanter Terme ergibt sich (96) zu

$$f_{x_k|Y_k}(\xi_k | Z_k) = \frac{f_{y_k|x_k}(\xi_k | \xi_k) f_{x_k|Y_{k-1}}(\xi_k | Z_{k-1})}{f_{y_k|Y_{k-1}}(\xi_k | Z_{k-1})}. \quad (97)$$



## D CONDENSATION-ALGORITHMUS

### VERWENDETE GRÖSSEN

$\mathbf{x}_k, \xi_k$	Zustandsvektor $\mathbf{x}_k$ des Zeitpunkts $k$ mit zugehöriger Dummyvariable $\xi_k$ .
$\mathbf{y}_k, \zeta_k$	Messvektor $\mathbf{y}_k$ des Zeitpunkts $k$ mit zugehöriger Dummyvariable $\zeta_k$ .
$\{s_k, \pi_k\}$	Repräsentation eines Systemzustands des Zeitpunkts $k$ durch $N$ Partikel (Hypothesen) $s_k = \{s_k^{(1)}, \dots, s_k^{(N)}\}$ mit zugehörigen Gewichten $\pi_k = \{\pi_k^{(1)}, \dots, \pi_k^{(N)}\}$ . Für die Menge der Partikelgewichte eines Zeitpunkts gilt: $\sum_{i=1}^N \pi_k^{(i)} = 1$ .
$Y_k = \{\mathbf{y}_1, \dots, \mathbf{y}_k\},$ $Z_k = \{\zeta_1, \dots, \zeta_k\}$	Menge aller bis zum Zeitpunkt $k$ vorhandener Messvektoren sowie Menge aller bis zum Zeitpunkt $k$ vorhandener Dummymessvektoren.
$X_k = \{\mathbf{x}_0, \dots, \mathbf{x}_k\}$	Menge aller bis zum Zeitpunkt $k$ vorhandener Zustandsvektoren.
$f_{\mathbf{y}_k \mathbf{x}_k}(\zeta_k \xi_k)$	Observationsdichte.
$f_{\mathbf{x}_{k-1} \mathbf{y}_{k-1}}(\xi_{k-1} Z_{k-1})$	Wahrscheinlichkeitsdichte des Zustands zum Zeitpunkt $k-1$ ( <i>prior</i> ).
$f_{\mathbf{x}_k \mathbf{y}_{k-1}}(\xi_k Z_{k-1})$	Prädiktion des Systemzustands ( <i>dynamic prior</i> ).
$f_{\mathbf{x}_k \mathbf{y}_k}(\xi_k Z_k)$	Wahrscheinlichkeitsdichte des Zustands zum Zeitpunkt $k$ ( <i>posterior</i> ).

### ANNAHMEN

1. Propagation des Systemzustands gemäß

$$f_{\mathbf{x}_k|\mathbf{y}_k}(\xi_k|Z_k) = \frac{f_{\mathbf{y}_k|\mathbf{x}_k}(\zeta_k|\xi_k) f_{\mathbf{x}_k|\mathbf{y}_{k-1}}(\xi_k|Z_{k-1})}{\int f_{\mathbf{y}_k|\mathbf{x}_k}(\zeta_k|\xi_k) f_{\mathbf{x}_k|\mathbf{y}_{k-1}}(\xi_k|Z_{k-1}) d\xi_k},$$

wobei die Annahmen aus Anhang C übernommen werden.

### ESTIMATIONSPROZESS

Zu Beginn der Operation wird auf Grundlage der initialen Dichte<sup>105</sup>  $f_{\mathbf{x}_0}(\xi_0)$  zunächst mittels faktorisierten Entnehmens eine anfängliche Repräsentation des Systemzustands  $\{s_0, \pi_0\}$  erzeugt, welche als Ausgangsbasis für die iterative Bestimmung der posterioren Wahrscheinlichkeitsdichte

<sup>105</sup>Im Allgemeinen lässt sich ein Anfangszustand  $\{s_0, \pi_0\}$  immer durch zufällig initialisierte Partikel definieren. Alle Partikelgewichte werden dabei auf den reziproken Wert der Partikelanzahl  $N^{-1}$  gesetzt.

## D CONDENSATION-Algorithmus

nachfolgender Zeitpunkte dient. Im Folgenden wird davon ausgegangen, dass zwecks Bestimmung der posterioren Dichte  $f_{\mathbf{x}_k|Y_k}(\xi_k|Z_k)$  eines Zeitpunkts  $k$  die entsprechende Darstellung der prior-Dichte des vorausgegangenen Zeitpunkts  $f_{\mathbf{x}_{k-1}|Y_{k-1}}(\xi_{k-1}|Z_{k-1})$ , in Form des Tupels  $\{s_{k-1}, \pi_{k-1}\}$ , gegeben ist. Die Estimation eines Folgezeitpunkts gemäß des CONDENSATION-Algorithmus geschieht in drei Stufen [192].

### 1. STUFE: ERZEUGUNG DER PARTIKELMENGE DES NACHFOLGEZEITPUNKTS

Aus der Partikelmenge  $s_{k-1}$  wird durch zufälliges Entnehmen mit Rücksetzen die Partikelmenge  $s'_k$  gebildet, wobei die Anzahl  $N$  der Partikel erhalten bleibt. Die Wahrscheinlichkeit der Entnahme eines Partikels aus  $s_{k-1}^{(i)}$  wird dabei durch sein Partikelgewicht  $\pi_{k-1}^{(i)}$  bestimmt. Aufgrund der unterschiedlichen Partikelgewichte werden Partikel mit hoher Wichtung häufiger ausgewählt, was dazu führt, dass in der Menge  $s'_k$  identische Elemente enthalten sind und entspricht einer Degeneration der Partikelmenge [182]. Die Gewichte aller neu erzeugten Partikel aus der Menge  $\pi'_k$  werden dabei auf die reziproke Partikelanzahl  $N^{-1}$  gleichgesetzt.

### 2. STUFE: ANWENDUNG DER SYSTEMDYNAMIK (PRÄDIKTION)

Unter Anwendung der Systemübergangsgleichung erfahren alle Partikel der Menge  $s'_k$  zunächst einen deterministischen Drift. Durch anschließende stochastische Diffusion der einzelnen Partikel erfolgt eine Reduktion der Partikeldegeneration. Die diesbezüglich beeinflussten Partikel bilden die neue Partikelmenge  $s_k$ . Im Kontext einer probabilistischen Estimation (Anhang C) entspricht dieses dem Prädiktionsschritt, sodass das Tupel  $\{s_k, \pi_k\}$  als der dynamische Prior  $f_{\mathbf{x}_k|Y_{k-1}}(\xi_k|Z_{k-1})$  des Zeitpunkts  $k$  interpretiert werden kann.

### 3. STUFE: WICHTUNG DER PARTIKELMENGE DES NACHFOLGEZEITPUNKTS (AKTUALISIERUNG)

Mittels der Observationsdichte sowie des Observationsschritts des partitionierten Entnehmens (*partitioned sampling*)

$$\pi_k^{(i)} = f_{\mathbf{y}_k|\mathbf{x}_k}(\zeta_k | \xi_k = s_k^{(i)}) \quad (98)$$

erfolgt eine Neubestimmung der Gewichte der Hypothesen aus  $s_k$ . Die erzeugte Menge der Partikelgewichte  $\pi_k$  ergibt nach Normalisierung  $\sum_i \pi_k^{(i)} = 1$  zusammen mit der Partikelmenge  $s_k$  eine Repräsentation der posterioren Dichte  $\{s_k, \pi_k\} \triangleq f_{\mathbf{x}_k|Y_k}(\xi_k|Z_k)$  des Nachfolgezustands. Der komplette Ablauf eines Iterationsschrittes des CONDENSATION-Algorithmus findet sich in Abb. 99.

Mittels der Informationen aus  $\{s_k, \pi_k\}$  können im Anschluss durch

$$\mathbb{E} \| f(\mathbf{x}_k) \| = \sum_{i=1}^N \pi_k^{(i)} f(s_k^{(i)}) \quad (99)$$

statistische Momente geschätzt werden.

## D CONDENSATION-Algorithmus

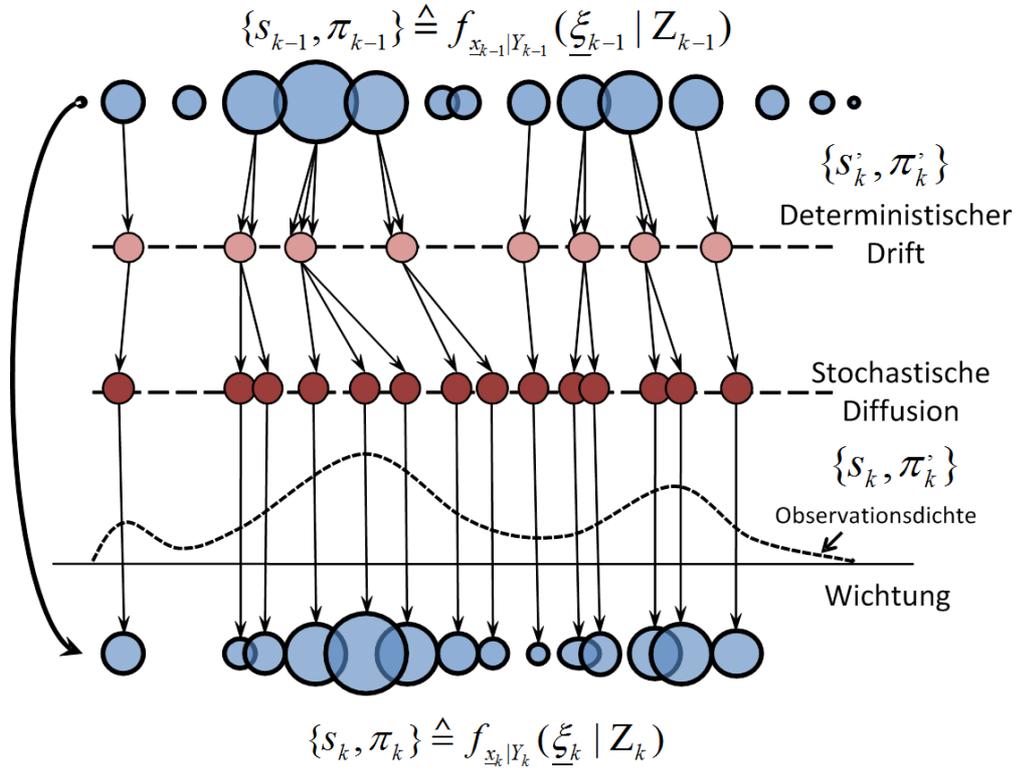


Abbildung 99: Prozessablauf einer Iteration des CONDENSATION-Algorithmus. Ausgehend von der durch den Tupel  $\{s_{k-1}^{(n)}, \pi_{k-1}^{(n)}\}$  repräsentierten Dichte  $f_{x_{k-1}|Y_{k-1}}(\underline{\xi}_{k-1} | Z_{k-1})$  findet nach der Erzeugung einer neuen Partikelmenge  $s'_k$  zunächst ein durch die Zustandsübergangsfunktion (50) definierter, deterministischer Drift sowie eine zufällige Diffusion aller Elemente aus  $s'_k$  statt:  $s'_k \rightarrow s_k$ . Angeschlossen dazu ist eine Neuberechnung der Partikelgewichte (49) unter Zuhilfenahme einer zum Zeitpunkt  $k$  verfügbaren Observationsdichte  $f_{y_k|x_k}(z_k | \underline{\xi}_k)$ , was zu einer der posterioren Dichte partikelbasierten Beschreibung des Nachfolgezustands  $\{s_k, \pi_k\} \hat{=} f_{x_k|Y_k}(\underline{\xi}_k | Z_k)$  führt.



## E QUATERNIONEN

### ROTATIONEN DURCH DREHMATRIZEN

Rotationen im euklidischen Raum können durch die Elemente der speziellen orthogonalen Gruppe

$$SO(3) = \{ M \in \mathbb{R}^{3 \times 3} \mid M^T M = I_3, |M| = 1 \}, \quad (100)$$

repräsentiert durch Drehmatrizen, beschrieben werden. Rotationsinversion sowie Kombinationen von Rotationen werden durch Inversionen bzw. Multiplikationssequenzen der Drehmatrizen erreicht. Eine Drehung  $R = (\hat{x}, \phi)$  um eine Achse  $\hat{x}$  auf der Einheitskugel mit Winkel  $\phi$  wird dabei durch

$$R(\hat{x}, \phi) = D(\hat{x}) + \cos \phi (I_3 - D(\hat{x})) + \sin \phi \Omega(\hat{x}) \quad (101)$$

berechnet, mit

$$D(\hat{x}) = \begin{pmatrix} x_1^2 & x_1 x_2 & x_1 x_3 \\ x_2 x_1 & x_2^2 & x_2 x_3 \\ x_3 x_1 & x_3 x_2 & x_3^2 \end{pmatrix}, \quad \Omega(\hat{x}) = \begin{pmatrix} 0 & -x_3 & x_2 \\ x_3 & 0 & -x_1 \\ -x_2 & x_1 & 0 \end{pmatrix}. \quad (102)$$

Im Falle einer Partitionierung in drei Einzeldrehungen kann die komplette Drehung als das Produkt

$$R(\hat{x}, \phi) = R(\hat{x}_1, \phi_1) R(\hat{x}_2, \phi_2) R(\hat{x}_3, \phi_3) \quad (103)$$

geschrieben werden. Legt man die Drehachsen  $\hat{x}_1, \hat{x}_2, \hat{x}_3$  fest, z. B. durch die Einheitsvektoren  $\hat{e}_1, \hat{e}_2, \hat{e}_3$  eines euklidischen Koordinatensystems, kann eine Rotation in diesem Raum lediglich durch Angabe der Winkel (Euler-Winkel)  $\phi_1, \phi_2, \phi_3$  durchgeführt werden. Es gilt zu beachten, dass Rotationsoperationen im euklidischen Raum nicht kommutativ sind, eine Rotationssequenz muss vorher festgelegt werden. Eine häufige in der Navigation verwendete Konvention ist die Drehung in einem globalen ortsfesten Koordinatensystem, wobei als Drehachsen die Standardbasis verwendet wird. Die entsprechenden Winkel werden dann mit *roll*, *pitch* sowie *yaw* benannt (Häng-, Neigungs- und Gierwinkel).

### ROTATIONEN DURCH QUATERNIONEN

Eine alternative Formulierung der Drehungen im dreidimensionalen Raum ist durch die Einführung von Quaternionen möglich. Mit

$$H = \{ q = x_0 + x_1 \mathbf{i} + x_2 \mathbf{j} + x_3 \mathbf{k} \mid x_i \in \mathbb{R}, i = 0, 1, 2, 3; \underbrace{\mathbf{i}^2 = \mathbf{j}^2 = \mathbf{k}^2 = \mathbf{i} \mathbf{j} \mathbf{k} = -1}_{\text{Hamilton-Regeln}} \} \quad (104)$$

## E Quaternionen

ist eine Erweiterung der reellen Zahlen definiert.  $H$  wird zuweilen auch nach ihrem Entdecker, Sir William Rowan Hamilton als die Menge der Hamilton-Zahlen oder schlicht Hamiltonsche Quaternionen genannt. Die Elemente lassen sie sinnvoll in einen reellen und vektoriellen Anteil separieren:  $q = [x_0, \underline{x}]^T$  mit  $\underline{x} = [x_1, x_2, x_3]^T$ . Für die Elemente von  $H$  gilt [206]:

$$\begin{aligned}
 q + q' &= [x_0 + x_0', \underline{x} + \underline{x}']^T && \text{Addition,} \\
 q \cdot q' &= [x_0 x_0' - \langle \underline{x}, \underline{x}' \rangle, x_0 \underline{x}' + x_0' \underline{x} - \underline{x} \times \underline{x}']^T && \text{Multiplikation,} \\
 q^* &= [x_0, -\underline{x}]^T && \text{Konjugation,} \\
 \|q\| &= \sqrt{q q^*} = \sqrt{x_0^2 + \|\underline{x}\|^2} && \text{Norm,} \\
 q^{-1} &= \frac{q^*}{\|q\|^2} && \text{Inversion.}
 \end{aligned} \tag{105}$$

Definiert sei weiterhin

$$H_1 = \{ \hat{q} \in H \mid x_0^2 + x_1^2 + x_2^2 + x_3^2 = \|\hat{q}\| = 1 \}, \quad H_i \subset H \tag{106}$$

als eine Untermenge von  $H$ , welche die Menge der Einheitsquaternionen (Versoren) enthält. Durch Transformation der Elemente von  $H_1$  in Polarform [206], [208] ergibt sich ein Versor zu

$$\hat{q} = x_0 + x_1 \mathbf{i} + x_2 \mathbf{j} + x_3 \mathbf{k} = [x_0, \underline{x}]^T \rightarrow [\cos \varphi, \hat{x} \sin \varphi]^T = \cos \varphi + \hat{x} \sin \varphi, \tag{107}$$

mit  $\hat{x} = \underline{x} / \|\underline{x}\|$  und  $\varphi = 2 \cos^{-1} w = 2 \sin^{-1} \sqrt{x^2 + y^2 + z^3}$ .

Der interessante Aspekt dieser Darstellung ist, dass ein Versor in Polardarstellung (107) als eine Rotation  $R = (\hat{x}, 2\varphi)$  (101) interpretiert werden kann [207]. Eine Drehung  $R(\hat{x}, \varphi)$  eines Vektors  $r \rightarrow r'$  mit  $r, r' \in \mathbb{R}^3$  lässt sich dann im Formalismus der Quaternionen durch

$$p' = q p q^{-1}, \quad q = [\cos \varphi / 2, \hat{x} \sin \varphi / 2]^T, \quad p = [0, r]^T, \quad p' = [0, r']^T \tag{108}$$

beschreiben [206], [207]. Vorteile, welche sich durch die Verwendung von Quaternionen ergeben, sind:

1. Einfacher und intuitiver in der Handhabung im Vergleich zu der Darstellung durch Matrizen. Kompaktere Darstellung (Quaternion 4 Werte, Matrix 9 Werte).
2. Schnellere und energiesparendere Berechnung der Drehungen: 16 Multiplikationen und 12 Additionen im Vergleich zu 27 Multiplikationen und 18 Additionen im Falle der Verwendung von Matrizen [209].
3. Vermeidung von Singularitäten (*Gimbal Locks*) [210].

## F TABELLEN

Table 4: Durch Datenfusion erzielbarer Nutzen (in Anlehnung an [87], [88], [90], [92]).

Nutzkategorie	Genereller Nutzen	Betriebsbedingte Vorteile
Robuste Operation	Einige Sensoren können Informationen liefern, während andere durch Störung, Ausfall oder Verdeckung in ihrer Operation behindert sind.	<ul style="list-style-type: none"> <li>• Erhöhte Detektionswahrscheinlichkeit</li> <li>• Möglichkeit der fortgesetzten Operation trotz Störung</li> <li>• <i>Graceful Degradation</i></li> </ul>
Erweiterte räumliche Abdeckung	Verschiedene Abdeckungsbereiche erlauben eine Erweiterung des Sichtfeldes.	<ul style="list-style-type: none"> <li>• Erhöhte Detektionswahrscheinlichkeit</li> <li>• Erhöhte Überlebenswahrscheinlichkeit</li> </ul>
Erweiterte temporale Abdeckung	Einige Systeme erlauben die Detektion / Messung in der Totzeit der anderen Sensoren.	<ul style="list-style-type: none"> <li>• Erhöhte Detektionswahrscheinlichkeit</li> <li>• Schutz im Falle etwaiger modaler Unterdrückung (Stealth, IR-Suppression etc.)</li> </ul>
Erhöhtes Vertrauen	Mehrere Sensoren können das Ereignis oder Objekt bestätigen.	<ul style="list-style-type: none"> <li>• Regelungen bzgl. des Eingriffs benötigen eine (verlässliche) positive Zielidentifikation</li> <li>• Sicherheit im Falle eingeleiteter situationsbedingter Gegenmaßnahmen</li> </ul>
Verringerte Mehrdeutigkeit	Gemeinsame Information von idealerweise orthogonalen Sensoren reduziert die Menge der entstehenden Hypothesen bzgl. eines Zieles resp. Objekts.	<ul style="list-style-type: none"> <li>• Erhöhte Zielidentifikation</li> <li>• Möglichkeit der Priorisierung</li> <li>• Erhöhtes Potenzial im Rahmen einer Assistenz</li> </ul>
Verbesserte Detektion	Effektive Integration mehrerer (idealerweise multimodaler) Messungen des Objekts / Zieles / Ereignisses führt zu einer erhöhten Verlässlichkeit in Bezug auf eine Detektion.	<ul style="list-style-type: none"> <li>• Verlängerung der zur Verfügung stehenden Reaktionszeit</li> <li>• Erhöhte Überlebensfähigkeit</li> </ul>
Erhöhte räumliche Auflösung	Eine Anordnung aus multiplen Sensoren kann (geometrisch) im Vergleich zu einem einzelnen Sensor eine Apertur mit erhöhter Raumaufklärung formen.	<ul style="list-style-type: none"> <li>• Erhöhte Genauigkeit in Bezug auf Positionsdaten unterstützt sowohl aktive als auch passive Maßnahmen</li> </ul>
Erhöhte Zuverlässigkeit	Eine Anordnung aus multiplen Sensoren besitzt inhärent eine erhöhte Redundanz.	<ul style="list-style-type: none"> <li>• Redundante Systeme</li> <li>• <i>Graceful Degradation</i></li> </ul>
Gesteigerte Dimensionalität	Ein System mit verschiedenen (idealerweise orthogonalen) Sensoren ist generell weniger anfällig für Störungen.	<ul style="list-style-type: none"> <li>• Erhöht die Überlebensfähigkeit des Systems</li> <li>• Erlaubt eine kontinuierliche Operation</li> </ul>

## F Tabellen

Tabelle 5: Applikationen und Projekte zum Thema Multisensor-Datenfusion.

Projekt	Zielsetzung Auswertung	Sensorik	Sensor-Knotenanzahl	Überwachungsgebiet
DARPA-NEST [126]	Detektion, Klassifikation und Verfolgung von Personen, Soldaten und Fahrzeugen	Radar, Magnetfeld	90	140m <sup>2</sup>
SensIT [127]	Militärische Fahrzeugklassifikation in Kampfgebieten	Akustik, Seismik, pol. IR-Licht	75	2700m <sup>2</sup>
RDECOM [128]	Fahrzeugeterkennung	IR-Kamera, Farbkamera, Akustik	3	ca. 150m <sup>2</sup>
Surveillance [129]	Personentracking	PIR, Akustik, Magnetfeld	144	5000m <sup>2</sup>
VIGILNET [130]	Detektion schneller Ziele	PIR, Akustik, Magnetfeld, Temperatur	70	500m <sup>2</sup>
Line in the Sand [131]	Detektion, und Klassifikation von Personen, Soldaten und Fahrzeugen	PIR, Akustik, Magnetfeld	10000	10000m <sup>2</sup>
ScatterWeb [132]	Zaunüberwachung	Beschleunigung	Abhängig von Gebäudegröße / Zaunlänge	Abhängig von Gebäudegröße / Zaunlänge
SAFE-Project [133]	Situationsanalyse	Optische Kameras	Abhängig von Gebäudegröße	Abhängig von Gebäudegröße
Battlefield-Surveillance [134]	Detektion und Verfolgung von Fahrzeugen in einem Kampfgebiet	Magnetfeld, Akustik	Abhängig von Flächengröße	Abhängig von Flächengröße

Tabelle 6: Zusammensetzung des Headers eines 2D/3D-Bildes bzw. Videobildes.

Wort (16 Bit)	Bedeutung
1	<b>Zeitstempel:</b> Jahr
2	<b>Zeitstempel:</b> Monat (1 ... 12)
3	<b>Zeitstempel:</b> Tag (1 ... 31)
4	<b>Zeitstempel:</b> Wochentag (1 ... 7)
5	<b>Zeitstempel:</b> Stunde (0 ... 23)
6	<b>Zeitstempel:</b> Minute (0 ... 59)
7	<b>Zeitstempel:</b> Sekunde (0 ... 59)
8	<b>Zeitstempel:</b> Millisekunde (0 ... 999)
9 - 14	Nicht belegt
15	<b>Datenquellen:</b> 10 -> 3D-Daten 11-> 2D- + 3D-Daten
16	0

## F Tabellen

*Tabelle 7: Zusammensetzung eines MultiCam-Videobildes.*

Byte	Bedeutung
0...31	Bildheader (Tab. 6).
32...307231	Bayer-Patch des 2D-Bildes (1 Byte pro Pixel $\rightarrow$ 640x480 Pixel $\hat{=}$ 300 kB).
307232...319519	Entfernungsdaten im Fließkommaformat (64x48 Pixel x 4Byte $\hat{=}$ 12 kB).
319520...331807	Modulationsamplitude im Fließkommaformat (64x48 Pixel x 4Byte $\hat{=}$ 12 kB).
331808...344096	Phasenbild $\psi_1=0^\circ$ (Abschnitt 2.1.1) im vorzeichenbehafteten Ganzzahlformat (16Bit pro Pixel $\rightarrow$ 64x48 Pixel $\hat{=}$ 6kB). Mit jeweils zwei Werten pro Pixel ( $N_{+, -}^{\psi_1}$ , Abschnitt 2.1.1) ergibt sich die Größe eines Phasenbildes damit zu 12 kB.
344096...344111	Footer Phasenbild $\psi_1$ (16 Byte).
344112...356399	Phasenbild $\psi_2=90^\circ$ (12 kB).
356400...356415	Footer Phasenbild $\psi_2$ (16 Byte).
356416...368703	Phasenbild $\psi_3=180^\circ$ (12 kB).
368704...368719	Footer Phasenbild $\psi_3$ (16 Byte).
368720...381007	Phasenbild $\psi_4=270^\circ$ (12 kB).
381008...381023	Footer Phasenbild $\psi_4$ (16 Byte).



## G LITERATURVERZEICHNIS

- [1] A. Weibel, R. Stiefelhagen: *CHIL - Computers in the Human Interaction Loop*. Springer Verlag, 2009. ISBN 978-1-84882-053-1.
- [2] *AMI - Augmented Multiparty Interaction*, <http://www.amiproject.org>
- [3] *SUBITO - Surveillance of Unattended Baggage and the Identifikation and Tracking of the Owner*, <http://www.subito-project.eu>
- [4] *PROMETHEUS - Prediction and Interpretation of Human Behaviour Based on Probabilistic Structures and Heterogeneous Sensors*, <https://www.informationsystems.foi.se/~prometheus-fp7>
- [5] *SAMURAI - Suspicious and Abnormal Behaviour Monitoring Using a Network of Cameras & Sensors for Situation Awareness Enhancement*, <http://www.samurai-eu.org>
- [6] *PASCAL2 - Pattern Analysis, Statistical Modelling and Computational Learning* <http://www.pascal-network.org>
- [7] *SAFROS - Patient Safety in Robotic Surgery*, <http://www.safros.eu>
- [8] M. Valera, S. A. Velastin: *Intelligent Distributed Surveillance Systems: A Review*. IEE Proc-Vis. Image Signal Process. Vol 152, No. 2, April 2005.
- [9] J. Zhu, L. Wang, R. Yang, J Davis: *Fusion of Time-of-Flight Depth and Stereo for High Accuracy Maps*. Computer Vision and Pattern Recognition (CVPR), 2008.
- [10] U. Hahne, M. Alexa: *Combining Time-of-Flight Depth and Stereo Images without Accurate Extrinsic Calibration*. Dynamic 3D Workshop in Conjunction with DAGM, 2007.
- [11] K. D. Kuhnert, M. Stommel: *Fusion of stereo-camera and pmd-camera data for real-time suited precise 3D environment reconstruction*. International Conference on Intelligent Robots and Systems, 2006.
- [12] L. Romero, A. Nunez, S. Bravo, L. E. Gamboa: *Fusing a Laser Range Finder and a Stereo Vision System to Detect Obstacles in 3D*. LNAI, IBERAMIA 2004.
- [13] Q. Mühlbauer, K. Kühnlenz, M. Buss: *Fusing Laser and Vision Data with a Genetic ICP Algorithm*. 10th Int. Conf. on Control, Automation, Robotics and Vision (ICARCV), 2008.
- [14] M. Perollaz, R. Labayrade, C. Royere, N. Hautiere, D. Aubert: *Long Range Obstacle Detection Using Laser Scanner and Stereovision*. Intelligent Vehicles Symposium, Tokio, Japan, 2006.
- [15] D. C. Ghiglia, M. D. Pritt: *Two-Dimensional Phase Unwrapping: Theory, Algorithms, and Software*. Wiley & Sons, ISBN 978-0471249351.
- [16] J. Krumm, S. Harris, B. Meyers, B. Brumitt, M. Hale, S. Shafer: *Multi-Camera Multi-Person Tracking for EasyLiving*. Third IEEE International Workshop on Visual Surveillance, Dublin, Ireland, 2000.
- [17] J. P. Batista: *Tracking Pedestrians under Occlusion using Multiple Cameras*. International Conference on Image Analysis and Recognition, Lecture Notes in Computer Science 3212, 2004.
- [18] A. Hampapur, L. Brown, J. Connell, A. Ekin, N. Haas, M. Lu, H. Merkl, S. Pankanti, A. Senior, C. Shu, Y. L. Tan: *Smart Video Surveillance*. IEEE Signal Processing Magazine, March 2005.
- [19] M. Harville, D. Li: *Fast, Integrated Person Tracking and Activity Recognition with Plan-View Templates from a Single Stereo Camera*. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2004.

## G Literaturverzeichnis

---

- [20] R. Eshel, Y. Moses: *Homography based multiple camera detection and tracking of people in a dense crowd*. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2008.
- [21] A. Tyagi, G. Potamianos, J. W. Davis, S. M. Chu: *Fusion of Multiple Camera Views for Kernel-Based 3D Tracking*. IEEE Workshop on Motion and Video Computing, Austin TX, 2007.
- [22] A. Mittal, L. S. Davis: *M2Tracker: A Multi-View Approach to Segmenting and Tracking People in a Cluttered Scene*. International Journal of Computer Vision 51, 2003.
- [23] K. Bernardin, T. Gehrig, R. Stiefelhagen: *Multi- and Singleview Multiperson Tracking for Smart Room Environments*. Lecture Notes in Computer Science, Vol. 4122, 2007.
- [24] T. Mori, T. Matsumoto, M. Shimosaka, H. Noguchi, T. Sato: *Multiple Persons Tracking with Data Fusion of Multiple Cameras and Floor Sensors Using Particle Filter*. Workshop on Multi-camera and Multi-modal Sensor Fusion Algorithms and Applications M2SFA2, Marseille, 2008.
- [25] M. K. Bhuyan, B. C. Lovell, A. Bigdeli: *Tracking with Multiple Cameras for Video Surveillance*. 9th Biennial Conference of the Australian Pattern Recognition Society, Glenelg, Australia, 2007.
- [26] T. Zhao, M. Aggarwal, R. Kumar, H. Sawhney: *Real-time Wide Area Multi-Camera Stereo Tracking*. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), 2005.
- [27] J. F. Martinez, S. Bravo, A. B. Garcia, I. Corredor, M. S. Familiar, L. Lopez, V. Hernandez, A. Da Silva: *Pervasive surveillance-agent system based on wireless sensor networks: design and deployment*. IOP Publishing in Measurement Science and Technology 21, 2010.
- [28] M. A. Patricio, J. Carbo, O. Perez, J. Garcia, J. M. Molina: *Multi-agent Framework in Visual Sensor Networks*. EURASIP Journal on Advances in Signal Processing, Vol. 2007.
- [29] J. Orwell, S. Massey, P. Remagnino, D. Greenhill, G. A. Jones: *A Multi-agent Framework for Visual Surveillance*. Proceedings of the 10th International Conference on Image Analysis and Processing (ICIAP), 1999.
- [30] T. Matsuyama, N. Ukita: *Real-Time Multitarget Tracking by a Cooperative Distributed Vision System*. Proceedings of the IEEE, Vol. 90, No. 7, July 2002.
- [31] F. Castaneda, M. A. Patricio, J. Garcia, J. M. Molina: *Extending Surveillance Systems Capabilities Using BDI Cooperative Sensor Agents*. Proceedings of the 4th ACM International Workshop on Video Surveillance and Sensor Networks (VSSN), 2006.
- [32] M. H. Cohen: *Building Brains for Rooms: Designing Distributed Software Agents*. Proceedings of the 9th Annual Conference on Innovative Applications of Artificial Intelligence (IAAI), 1997.
- [33] D. R. Karrupiah, Z. Zhu, P. Shenoy, E. Riseman: *A Fault-Tolerant Distributed Vision System Architecture for Object Tracking in a Smart Room*. International Workshop on Computer Vision Systems (ICCV), 2001.
- [34] F. Castaneda, J. Garcia, M. A. Patricio, J. M. Molina: *A Multi-Agent Architecture to Support Active Fusion in a Visual Sensor Network*. AMC/IEEE International Conference on Distributed Smart Cameras (ICDSC), 2008.
- [35] L. Marchesotti, L. Piva, C. Regazzoni: *An agent-based approach for tracking people in indoor complex environments*. Proceedings of the 12th International Conference on Image Analysis and Processing (ICIAP), 2003.
- [36] B. Schneider: *Der Photomischdetektor zur schnellen 3D-Vermessung für Sicherheitssysteme und zur Informationsübertragung im Automobil*. Dissertation, Universität Siegen, 2003.
-

## G Literaturverzeichnis

---

- [37] S. Hußmann: *Schnelle 3-D Objektvermessung mittels PMD/CMOS-Kombizeilensensor und Signalkompressions-Hardware*. Dissertation, Universität Siegen, 2000.
- [38] T. P. U. Krieger: *Innovatives Sensorkonzept und Signalverarbeitungsstrategien zur Bewegungserkennung und Präsenzkontrolle von Personen*. Dissertation, Universität Siegen, 2002.
- [39] M. Zacher: *Integration eines optischen 3D-Sensors in ein Koordinatenmessgerät für die Digitalisierung komplexer Oberflächen*. Dissertation, RWTH Aachen, 2003.
- [40] R. Lange: *3D Time-of-Flight Distance Measurement with Custom Solid-State Image Sensors in CMOS / CCD-Technology*. Dissertation, Universität Siegen, 2000.
- [41] R. Langenbach: *Methoden der Farbbildverarbeitung für die automatische Oberflächeninspektion*. Dissertation, Universität Siegen, 2002.
- [42] X. Luan: *Experimental Investigation of Photonic Mixer Device and Development of ToF 3D Ranging Systems Based on PMD Technology*. Dissertation, Universität Siegen, 2001.
- [43] M. Albrecht: *Untersuchung von Photogate-PMD-Sensoren hinsichtlich qualifizierender Charakterisierungsparameter und -methoden*. Dissertation, Universität Siegen, 2007.
- [44] D. Justen: *Untersuchung eines neuartigen 2D-gestützten 3D-PMD-Bildverarbeitungssystems*. Dissertation, Universität Siegen, 2000.
- [45] K. Engelhardt: *Methoden und Systeme der optischen 3-D Messtechnik. XI. Internationaler Kurs für Ingenieurvermessung Zürich*. Dümmler Verlag, Bonn, 1992.
- [46] M. Bernfeld, C. E. Cook: *Radar signals: An introduction to theory and application*. Artech House, Boston, 1993. ISBN 0-89006-733-3.
- [47] A. Ludloff: *Praxiswissen Radar und Radarsignalverarbeitung*. Vieweg, Braunschweig, 2002. ISBN 3-528-26568-X.
- [48] M. Zollner, E. Zwicker: *Elektroakustik*. Springer, 1993.
- [49] E. Stadler: *Modulationsverfahren*. Vogel, 2000. ISBN 3802318404.
- [50] H. Weidenfeller, A. Vlcek: *Digitale Modulationsverfahren mit Sinusträger*. Springer, 1996. ISBN 354060622X.
- [51] J. Zheng: *Optical Frequency-modulated Continuous-wave Interferometry*. Springer, 2005. ISBN 038723002.
- [52] M. Meyer: *Kommunikationstechnik*. Vieweg und Teubner, 2002. ISBN 3528138653.
- [53] B. Jähne, H. Haußecker: *Computer Vision and Applications*. Academic Press. London, San Diego 2000 ISBN 0-12-379777-2.
- [54] F. Pedrotti : *Optik für Ingenieure*. Springer, ISBN 978-3-540-22813-4.
- [55] P. Hariharan: *Optical Interferometry*. Springer, ISBN 978-3-540-22813-4.
- [56] K. Hartmann et al.: *Klassifizierungsaspekte bei der 3D-Szenenexploration mit einer neuen 2D/3D Multichip-Kamera*. ASM 2005, pp. 74-80, Springer Verlag.
- [57] E. Hecht: *Optik*. Oldenbourg Wissenschaftsverlag 2005 ISBN 3486273590.
- [58] O. Schreer: *Grundlagen der projektiven Geometrie*. Springer, Berlin, Heidelberg 2005.
- [59] T. Spirig, P. Seitz: *The Lock-In CCD. Two-Dimensional Synchronous Detection of Light*. IEEE Journal of Quantum Electronics, Vol. 31, No. 9, S. 1705-1708, September 1995.
- [60] J. Olk: *Untersuchung von Laufzeitentfernungsmesssystemen unter besonderer Berücksichtigung des Referenzproblems*. Dissertation, Universität Siegen, 1997.
- [61] B. Möller: *Full frame 3D snapshot. Possibilities and limitations of 3D image acquisition without scanning*. Examensarbeit, Universität Linköping, 2005.
-

## G Literaturverzeichnis

---

- [62] O. Lottner: *Investigations of Optical 2D/3D-Imaging with Different Sensors and Illumination Configurations*. Dissertation, Universität Siegen, 2011.
- [63] N. Santrac, G. Friedland, R. Rojas: *High Resolution Segmentation with a ToF 3D-Camera using the Example of a Lecture Scene*. Fakultät für Informatik, Universität Berlin, 2006.
- [64] M. J. Valls, G. Dissanayake: *Robotic 3D visual mapping for augmented situational awareness in unstructured environments*. International Workshop on Robotics for Risky Interventions and Surveillance of the Environment (RISE), 2008.
- [65] B. Huhle, S. Fleck, A. Schilling: *Integrating 3D ToF Camera Data and High Resolution Images for 3DTV Applications*. 3DTC CON – The True Vision, 2007.
- [66] R. Reulke: *Combination of Distance Data with High Resolution Images*. ISPRS, Commission V Symposium, Image Engineering and Vision Metrology, 2006.
- [67] PMD Technologies GmbH: „PMD PhotonICs: 3k-S, 41k-S“ <http://www.pmdtec.com>
- [68] T. Kahlmann: *Range imaging metrology – Investigation, Calibration and Development*. Dissertation, ETH Zürich, 2007.
- [69] M. Lindner, A. Kolb: *Lateral and Depth Calibration of PMD-Distance Sensors*. Advances in Visual Computing, Springer, 2/2006.
- [70] B. Jähne: *Digitale Bildverarbeitung*. 6. Ausgabe, Springer Verlag, 2005. ISBN: 978-3-540-27384-4.
- [71] B. Büttgen, T. Oggier, M. Lehmann: *CCD/CMOS Lock-In Pixel for Range Imaging: Challenges, Limitations and State-of-the-Art*, Mesa Imaging, 1<sup>st</sup> Range Imaging Research Day, 2005.
- [72] T. Ringbeck, T. Möller, B. Hagebecker: *Multidimensional measurement by using 3-D PMD sensors*, Advances in Radio Science 5, 2007.
- [73] Micron Technology, Inc, Boise, USA: Data sheet for the 1/2-inch VGA Digital Image Sensor MT9V403.
- [74] S. E. Ghobadi, O. E. Loepprich, K. Hartmann, O. Loffeld: *Hand Segmentation Using 2D/3D Images*. IVCNZ 2007, Hamilton, New Zealand, 2007.
- [75] O. Lottner, A. Sluiter, K. Hartmann, W. Weihs: *Movement Artefacts in Range Images of Time-of-Flight Cameras*. EOS DOI: 10.1109/ISSCS.2007.4292665, Romania, 2007.
- [76] J. A. Parian, A. Gruen: *Integrated Laser Scanner and Intensity Image Calibration and Accuracy Assessment*. ISPRS Workshop "Laser Scanning", Enschede, September 2005.
- [77] H. Baltsakis, A. Argyros, P. Trahanias: *Fusion of range and visual data for the extraction of scene structure information*. Intl. Conference on Pattern Recognition (ICPR), 2002.
- [78] MESA Imaging: „Swissranger SR4000“ <http://www.mesa-imaging.ch>
- [79] Micron MT9V403C12: <http://pdf1.alldatasheet.com/datasheet-pdf/view/97421/MICRON/MT9V403.html>
- [80] S. Gould, P. Baumstarck, M. Quigley, A. N. Ng, D. Koller: *Integrating Visual and Range Data for Robotic Object Detection*. Workshop on Multi-camera and Multi-modal Sensor Fusion Algorithms and Applications (M2SFA2), France, 2008.
- [81] G. Monteiro, C. Premebida, P. Peixoto, U. Nunes: *Tracking and Classification of Dynamic Obstacles Using Laser Range Finder and Vision*. Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2006.
- [82] S. E. Ghobadi, O. E. Loepprich, O. Lottner, K. Hartmann, O. Loffeld: *2D/3D Image Data Analysis for Object Tracking and Classification*. Advances in Machine Learning and Data Analysis, Lecture Notes in Electrical Engineering, Vol. 48, 2010. ISBN 978-90-481-3176-1.
-

## G Literaturverzeichnis

---

- [83] S. E. Ghobadi, O. E. Loepprich, F. Ahmadov, J. Bernshausen, K. Hartmann, O. Loffeld: *Real Time Hand Based Robot Control Using 2D/3D Images*. Advances in Visual Computing, Lecture Notes in Computer Science, Vol. 5359, 2008. ISBN 978-3-540-89646-3.
- [84] M. Wooldridge: *An Introduction to Multi Agent Systems*. John Wiley & Sons, Chichester, H. F. Duran-Whyte: *Integration, Coordination and Control of Multi-Sensor Robot Systems*. England 2002. ISBN 0-471-49691-X.
- [85] L. Padgham, M. Winikoff: *Developing Intelligent Agent Systems*. John Wiley & Sons, Chichester, England 2004. ISBN 0-470-46120-7.
- [86] R. H. Bordini, J. F. Hübner, M. Wooldridge: *Programming Multi-Agent Systems in AgentSpeak using Jason*. J. Wiley & Sons, Chichester England, 2007. ISBN 978-0-470-02900-8.
- [87] D. L. Hall, S. A. H. McMullen: *Mathematical Techniques in Multisensor Data Fusion*. Second Edition. Artech House, Boston 2004. ISBN 978-1580533355.
- [88] E. L. Waltz: *Data Fusion for C<sup>3</sup>I in Command, Control, Communications Intelligence Handbook*. Palo Alto, CA: EW Communications Inc., 1986.
- [89] D. L. Hall, A. Steinberg: *Dirty Secrets in Multisensor Data Fusion*. National Symposium on Sensor Data Fusion, San Antonio, TX 2000.
- [90] D. L. Hall, J. Llinas, M. E. Liggins: *Handbook of Multisensor Data Fusion Theory and Practice, 2<sup>nd</sup> Edition*. Boca Raton, CRC Press 2009. ISBN 978-1-4200-5308-1.
- [91] H. F. Durant-Whyte: *Integration, Coordination and Control of Multi-Sensor Robot Systems*. Kluwert Academic Publishers, Norwell, Massachusetts 1988.
- [92] E. L. Waltz, J. Llinas: *Multisensor Data Fusion*. Artech House, Boston MA 1990.
- [93] P. K. Varshney: *Distributed Detection and Data Fusion*. Springer, New York 1996.
- [94] M. D. McNeese, E. Salas, M. Endsley: *New Trends in Cooperative Activities: System Dynamics in Complex Environments*. Santa Monica, CA: Human Factors and Ergonomics Society 2001.
- [95] D. L. Hall, R. J. Linn, J. Llinas: *Survey of Data Fusion Systems in SPIE Conference on Data Structure and Target Classification*. Orlando FL, 1991.
- [96] A. N. Steinberg: *Threat Management System for Combat Aircraft in Tri-Service Data Fusion Symposium*. Johns Hopkins University, Baltimore, MD, 1987.
- [97] R. Sessler: *Eine modulare Architektur für dienstbasierte Interaktionen zwischen Agenten*. Dissertation, Technische Universität Berlin, 2002.
- [98] M. Fathi-Torbaghan, A. Höffmann: *Fuzzy Logik und Blackboard-Modelle in der technischen Anwendung*. Oldenbourg Verlag 1994. ISBN 3-486-22650-9.
- [99] M. J. Wooldridge, N. R. Jennings: *Intelligent Agents: Theory and Practice*. Knowledge Engineering Review, 10, 1995.
- [100] M. d'Inverno, D. Kinny, M. Luck, M. J. Wooldridge: *A Formal Specification of dMARS*. Proceedings of ATAL'97, 1997.
- [101] SOAP Spezifikation: <http://www.w3.org/TR/soap>
- [102] Object Management Group: <http://www.omg.org>
- [103] CORBA Spezifikation: [http://www.omg.org/technology/documents/spec\\_catalog.htm](http://www.omg.org/technology/documents/spec_catalog.htm)
- [104] M. d'Inverno, M. Luck: *Understanding Agent Systems*. Springer Verlag 2001. ISBN 3-540-41975-6.
-

## G Literaturverzeichnis

---

- [105] N. R. Jennings, P. Faratin, M. J. Johnson, P. O'Brien, M. E. Wiegand: *Agent-based business process management*. International Journal of Cooperative Information Systems, 5(2&3), 1996.
- [106] D. Kinny, M. Georgeff, A. Rao: *A methodology and modelling technique for systems of BDI agents*. Proceedings of the Seventh European Workshop on Modelling Autonomous Agents in a Multi-Agent World, Lecture Notes in Artificial Intelligence, Vol. 1038, Springer-Verlag, 1996.
- [107] C. Toomey, W. Mark: *Satellite image dissemination via software agents*. IEEE Expert, 10(5), 1995.
- [108] R. H. Guttman, A. G. Moukas, P. Maes: *Agent-mediated electronic commerce: a survey*. Knowledge Engineering Review, 13(2), 1998.
- [109] S. Grand, D. Cliff: *Entertainment software agents with artificial life*. Autonomous Agents and Multi-Agent Systems, 1(1), 1998.
- [110] S. Franklin, A. Graesser: *Is it an Agent, or just a Program? A Taxonomy for Autonomous Agents*. Proceedings of the Third International Workshop on Agent Theories, Architectures and Languages, Springer-Verlag, 1996.
- [111] FIPA Agent Management, 2003. <http://www.fipa.org/specs/fipa00023>
- [112] M. E. Bratman: *Intention, Plans and Practical Reason*. Harvard University Press, Cambridge MA, 1987.
- [113] M. E. Bratman, D. J. Israel, M. E. Pollack: *Plans and resource bounded practical reasoning*. Computational Intelligence 4, 1998.
- [114] V. Mascardi, D. Demergasso, D. Ancona: *Languages for Programming BDI-style Agents*. WOA 2005 - Workshop: From Objects to Agents, 2005.
- [115] M. E. Bratman: *Intentions in Communication*. Harvard University Press, Cambridge MA, 1990.
- [116] D. Nau, M. Ghallab, P. Traverso: *Automated Planning: Theory & Practice*. Morgan Kaufman Publ. Inc., 2004. ISBN 1558608567.
- [117] M. P. Georgeff, A. L. Lansky: *Reactive reasoning and planning*. In Proceedings of Advancement of Artificial Intelligence (AAAI), 1987.
- [118] J. R. Searle: *Speech Acts*. Cambridge University Press, 1969.
- [119] J. L. Austin: *How to Do Things With Words*. Oxford University Press, 1962.
- [120] T. Finin, J. Weber, G. Wiederhold, M. Genesereth, R. Fritzson, D. McKay, J. McGuire, R. Pelavin, S. Shapiro, C. Beck: *Specification of the KQML agent communication language*. DARPA Knowledge Sharing Initiative, External Interfaces Working Group, 1993.
- [121] P. R. Cohen, C. R. Perrault: *Elements of a plan based theory of speech acts*. Cognitive Science, Vol. 3, 1979.
- [122] University of Michigan Procedural Reasoning System UM-PRS: [http://www.marcush.net/IRS/irs\\_downloads.html](http://www.marcush.net/IRS/irs_downloads.html)
- [123] L. Lamontagne, M. Belanger, J. Berger, P. Pelletier: *EVALUATION D'UN SYSTEME A RAISONNEMENT PROCEDURAL (UM-PRS)*. Research and Development Branch, Department of National Defence, Canada, 1998.
- [124] J. Lee, M. J. Huber, E. H. Durfee, P. G. Kenny: *UM-PRS: An Implementation Of The Procedural Reasoning System For Multirobot Applications System*. In Conference on Intelligent Robotics in Field, Factory, Service, and Space (CIRFFSS), 1994.
- [125] University of Michigan Procedural Reasoning System UM-PRS: *Programmer and User Guide*. <ftp://ftp.eecs.umich.edu/people/durfee/umprs3.0-man.ps>
-

## G Literaturverzeichnis

---

- [126] A. Arora, P. Dutta, S. Bapat, V. Kulathumani, H. Zhang: *A Line in the Sand: A Wireless Sensor Network for Target Detection, Classification, and Tracking*. Computer Networks (Elsevier) Vol. 46, 2004.
- [127] C. Meesookho, S. Narayanan, C. S. Raghavendra: *Collaborative Classification Applications In Sensor Networks*. IEEE Sensor Array and Multichannel Signal Processing Workshop, 2002.
- [128] P. Perconti, M. Loew, J. Hilger: *Overview of Sensor Fusion Research at RDECOM NVESD & Recent Results on Vehicle Detection Using Multiple Sensor Nodes*. Proceedings of the Sixth International Conference of Information Fusion, 2003.
- [129] P. Chen, S. Oh, M. Manzo: *Experiments in Instrumenting Wireless Sensor Networks for Real-Time Surveillance*. In International Conference on Robotics and Automation, 2005.
- [130] T. He, P. Vicaire, A. Yang, T. Yan: *Achieving Real-Time Target Tracking Using WSN*. In 12th IEEE Real-Time and Embedded Technology and Applications Symposium (RTAS), 2006.
- [131] P. Dutta, M. Grimmer, A. Arora, S. Bibyk, D. Culler: *Design of a Wireless Sensor Network Platform for Detecting Rare, Random, and Ephemeral Events*. In Proc. of the 4th Intl. Conf. on Information Processing in Sensor Networks (IPSN'05), Los Angeles, CA, 2005.
- [132] G. Wittenburg, K. Terfloth, F. Villafuerte: *Fence Monitoring – Experimental Evaluation of a Use Case for Wireless Sensor Networks*. In Proceedings of the 4th European Conference on Wireless Sensor Networks (EWSN'07), Delft, The Netherlands, January 2007.
- [133] N. L. Carter, J. M. Ferryman: *The SAFEE On-Board Threat Detection System*. ICVS 2008.
- [134] T. Bokareva, W. Hu, S. Kanhere: *Wireless Sensor Networks for Battlefield Surveillance*. Land Warfare Conference, Brisbane Australia. October 2006.
- [135] Meinberg Funkuhren GmbH & Co. KG: <http://www.meinberg.de/german/sw/ntp.htm>
- [136] SPTimeSync: <http://www.softpedia.com/get/Desktop-Enhancements/Clocks-Time-Management/SP-TimeSync.shtml>
- [137] D. D. Corkill: *Collaborating Software. Blackboard and Multi-Agent Systems & the Future*. In Proceedings of the International Lisp Conference, New York 2003.
- [138] Y. Benezeth, P. M. Jodoin, B. Emile, H. Laurent, C. Rosenberger: *Review and Evaluation of Commonly Implemented Background Subtraction Algorithms*. IEEE International Conference on Pattern Recognition, 2008.
- [139] S. Herrero, J. Bescos: *Background Subtraction Techniques: Systematic Evaluation and Comparative Analysis*. Advanced Concepts for Intelligent Vision Systems, Springer Berlin/Heidelberg, 2009. ISBN 978-3-642-04696-4.
- [140] R. Ewerth, B. Freisleben: *Frame difference normalization: an approach to reduce error rates of cut detection algorithms for MPEG videos*. International Conference on Image Processing ICIP, 2003.
- [141] S. C. Cheung, C. Kamath: *Robust Techniques for background subtraction in urban traffic video*. Visual Communications and Image Processing, Proceedings of the SPIE, Vol. 5308, 2004.
- [142] Q. Zhou, J. Aggarwal: *Tracking and classifying moving objects from video*. IEEE Proc. PETS Workshop, 2001.
- [143] A. Wren, A. Azarbayejani, P. Darrell, A. Pentland: *Real-time tracking of human body*. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 19, 1997.
- [144] C. Stauffer, W. E. L. Grimson: *Adaptive background mixture models for real-time tracking*. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1999.
-

## G Literaturverzeichnis

---

- [145] C. Stauffer, W. E. L. Grimson: *Learning patterns of activity using real-time tracking*. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2000.
- [146] P. KaewTraKulPong, R. Bowden: *An Improved Adaptive Background Mixture Model for Realtime Tracking with Shadow Detection*. Proceedings of the 2nd European Workshop on Advanced Video Based Surveillance Systems (AVBS), September 2001.
- [147] A. M. Elgammal, L. Harwood, L. S. Davis: *Non-parametric model for background subtraction*. Computer Vision-ECCV 2000, Springer Berlin/Heidelberg. ISBN 978-3-540-67686-7.
- [148] J. Han, M. Kamber, A. K. H. Tung: *Spatial Clustering Methods in Data Mining: A Survey*. Geographic Data Mining and Knowledge Discovery, Taylor and Francis, 2001.
- [149] P. Berkhin: *Survey of Clustering Data Mining Techniques*. Technical report, Accrue Software, San Jose, CA, 2002.
- [150] G. McLachlan, K. Basford: *Mixture Models: Inference and Applications to Clustering*. Marcel Dekker, New York, 1988.
- [151] G. McLachlan, T. Krishnan: *The EM Algorithm and Extensions*. Wiley & Sons, New York, 1997.
- [152] D. W. Scott: *Multivariate Density Estimation*. Wiley & Sons, New York, 1992.
- [153] <http://www.inf.ethz.ch/personal/befische/ml1/demos/index.html>
- [154] J. B. Macqueen: *Some Methods for Classification and Analysis of Multivariate Observations*. Proc. 5th Berkeley Symp. on Mathematical Statistics and Probability, 1967.
- [155] E. Forgy: *Cluster analysis of multivariate data: Efficiency versus interpretability of classification*. Biometrics, Vol. 21, 1965.
- [156] P. Viola, M. Jones: *Robust real-time Object Detection*. Second International Workshop on Statistical and Computational Theories of Vision-Modeling, Learning, Computing and Sampling, Vancouver Canada, 2001.
- [157] P. Viola, M. Jones: *Rapid object detection using a boosted cascade of simple features*. Proceedings of the Conference on Computer Vision and Pattern Recognition Vol. 1, Hawaii, USA, 2001.
- [158] P. Viola, M. Jones: *Robust real-time face detection*. International Journal of Computer Vision, 2004. ISSN 0920-5691.
- [159] R. Lienhart, J. Maydt: *An Extended Set of Haar-like Features for Rapid Object Detection*. IEEE ICIP, Vol. I, 2002.
- [160] C. Papageorgiou, M. Oren, T. Poggio: *A general framework for object detection*. International Conference on Computer Vision, 1998.
- [161] R. E. Schapire, Y. Freund, P. Bartlett, W. S. Lee: *Boosting the margin: a new explanation for the effectiveness of voting methods*. Proceedings of the 14th International Conference on Machine Learning, 1997.
- [162] R. E. Schapire, Y. Freund, P. Bartlett, W. S. Lee: *Boosting the margin: a new explanation for the effectiveness of voting methods*. The Annals of Statistics, Vol. 26, Number 5, 1998.
- [163] R. E. Schapire: *A Brief Introduction to Boosting*. Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence, 1999.
- [164] M. Kearns: *Thoughts on Hypothesis Boosting*. Unpublished manuscript, 1988. Zu finden unter: <http://www.cis.upenn.edu/~mkearns/papers/boostnote.pdf>
- [165] M. Kearns, L. G. Valiant: *Learning Boolean formulae or finite automata is as hard as factoring*. Technical Report TR-14-88, Harvard University, Aiken Computation Laboratory, August 1988.
-

## G Literaturverzeichnis

---

- [166] M. Kearns, L. G. Valiant L: *Cryptographic limitations on learning Boolean formulae and finite automata*. Journal of the Association for Computer Machinery, January 1994.
- [167] R. E. Schapire, Y. Freund: *Experiments With a New Boosting Algorithm*. Proceedings of the Thirteenth International Conference on Machine Learning, San Francisco, 1996.
- [168] J. Schochman, J. Matas: *AdaBoost Talk*. [http://www.comp.leeds.ac.uk/scsjsso/adaboost\\_talk.pdf](http://www.comp.leeds.ac.uk/scsjsso/adaboost_talk.pdf)
- [169] M. Warmuth, S. V. N. Vishwanathan: *Survey of Boosting from an Optimization Perspective*. Tutorial at the International Conference of Machine Learning, ICML, 2009.
- [170] J. Friedman, T. Hastie, R. Tibshirani: *A Statistical View of Boosting*. The Annals of Statistics, 2000.
- [171] O. H. Jensen: *Implementing the Viola-Jones Face Detection Algorithm*. Project at the Technical University of Denmark, 2008. ISBN 87-643-0008-0.
- [172] <http://www.meinberg.de/german/sw/ntp.htm>
- [173] <http://www.spdialer.com/timesync>
- [174] D. L. Mills: *Network Time Protocol Version 4 Reference and Implementation Guide*. NTP Working Group, Technical Report 06-6-1, University of Delaware, 2006.
- [175] N. Dalal: *Finding People in Images and Videos*. Dissertation, Institut National Polytechnique de Grenoble, 2006.
- [176] N. Dalal, B. Triggs, C. Schmid: *Human Detection using Oriented Histograms of Flow and Appearance*. European Conference on Computer Vision, 2006.
- [177] D. G. Lowe: *Distinctive image features from scale-invariant keypoints*. International Journal of Computer Vision, 2004.
- [178] The OpenCV Machine-Vision-Library: <http://opencv.willowgarage.com/wiki>
- [179] S. E. Ghobadi, O. E. Loepprich, K. Hartmann, O. Loffeld: *Real Time Person Tracking and Localization Using Multimodal Visual Data*. Workshop on Pattern Recognition and Artificial Intelligence for Human Behaviour Analysis (PRAI\*HBA), Reggio Emilia, Italien, 2009.
- [180] O. Loffeld: *Estimationstheorie I, Grundlagen und stochastische Konzepte*. Oldenbourg Verlag GmbH München. ISBN 3-486-21616-3.
- [181] O. Loffeld: *Estimationstheorie II, Anwendungen: Kalman-Filter*. Oldenbourg Verlag GmbH München. ISBN 3-486-21616-3.
- [182] P. Zarchan, H. Musoff: *Fundamentals of Kalman Filtering: A Practical Approach*. American Institute of Aeronautics & Ast (AIAA); 2<sup>nd</sup> dition, 2005. ISBN 978-1563476945.
- [183] B. Ristic, S. Arulampalam, N. Gordon: *Beyond the Kalman Filter*. Artech House, London 2004. ISBN 1-58053-631-x.
- [184] A. Doucet, N. deFreitas, N. Gordon: *Sequential Monte Carlo Methods in Practice*. Springer Verlag, 2001. ISBN 0-387-95146-6.
- [185] Y. Bar-Shalom: *Multitarget-Multisensor Tracking: Advanced Applications*. Artech House, Norwood 1990. ISBN 0-89006-377-x.
- [186] Y. Bar-Shalom, X. Li: *Estimation and Tracking: Principles, Techniques and Software*. Artech House, Norwood MA, 1993. ISBN 0-89006-643-4.
- [187] A. Yilmaz, O. Javed, M. Shah: *Object Tracking: A Survey*. ACM Computing Surveys, Vol 38, 2006.
- [188] S. E. Ghobadi: *Real-Time Object Recognition and Tracking Using 2D/3D Images*, Dissertation, Universität Siegen, 2010.
-

## G Literaturverzeichnis

---

- [189] F. Aherne, N. Thacker, P. Rockett: *The Bhattacharyya Metric as an Absolute Similarity Measure for Frequency Coded Data*. Kybernetika 32(4), 1997.
- [190] A. Bhattacharyya: *On a measure of divergence between two statistical populations defined by their probability distribution*. Bulletin of the Calcutta Mathematical Society, Vol. 35, 1943.
- [191] J. L. Cassidis, A. Junkins: *Optimal Estimation of Dynamic Systems*. Chapman & Hall, 2004. ISBN 1-58488-391-X.
- [192] M. Isard, A. Blake: *Contour Tracking by Stochastic Propagation of Conditional Density*. 4th European Conference on Computer Vision, Vol. 1, 1996.
- [193] M. Isard, A. Blake: *A Mixed-State Condensation Tracker with Automatic Model-Switching*. In 6th International Conference on Computer Vision 29, 1998.
- [194] M. Isard, A. Blake: *CONDENSATION - Conditional Density Propagation for Visual Tracking*. International Journal on Computer Vision 29(1), 1998.
- [195] M. Isard: *Visual Motion Analysis by Probabilistic Propagation of Conditional Density*. Dissertation, Universität Oxford, 1998.
- [196] U. Grenader, Y. Chow, D. M. Keenan: *HANDS, A Pattern Theoretic Study of Biological Shapes*. Springer-Verlag, 1991.
- [197] N. Gordon, D. Salmond, A. Smith: *Novel approach to nonlinear/non-gaussian bayesian state estimation*. IEE Proc. F, 140, 2.
- [198] G. Kitagawa: *Monte-carlo filter and smoother for non-Gaussian nonlinear state space models*. Journal of Computational and Graphical Statistics, Vol. 5, 1996.
- [199] A. Papoulis: *Probability, Random-Variables and Stochastic Processes*. McGraw-Hill 2. ed., New York, 1984. ISBN 0070484686.
- [200] <http://www.khronos.org/OpenGL>
- [201] [http://www.nvidia.com/object/cuda\\_home\\_new.html](http://www.nvidia.com/object/cuda_home_new.html)
- [202] <http://channel9.msdn.com/tags/DirectCompute-Lecture-Series>
- [203] <http://www.seeingmachines.com/product/faceapi>
- [204] <http://www.softkinetic.net/index.php>
- [205] <http://www.omekinteraktive.com>
- [206] [http://en.wikipedia.org/wiki/Quaternions\\_and\\_spatial\\_rotation](http://en.wikipedia.org/wiki/Quaternions_and_spatial_rotation)
- [207] L. Byung-Uk: *Stereo Matching of Skull Landmarks*. Dissertation, Universität Stanford, CA, 1991.
- [208] L. Vicci: *Quaternions and Rotations in 3-Space: The Algebra and its Geometric Interpretation*, Technical Report TR01-014, Universität North Carolina, Chapel Hill, 2001.
- [209] <http://code.google.com/p/coloradocollegegame/wiki/Quaternions>
- [210] E. B. Dam, M. Koch, M. Lillholm: *Quaternions, Interpolations and Animation*. Technical Report DIKU-TR-98/5, Fakultät für Informatik, Universität Kopenhagen, Dänemark, 1998.
- [211] E. Bradski, A. Kaehler: *Learning OpenCV*. O'Reilly Media Inc., 2008. ISBN: 978-0-596-51613-0.
- [212] T. Kahlmann, F. Remondino, H. Ingensand: *Calibration for increased accuracy of the range imaging camera SwissRanger<sup>TM</sup>*. Image Engineering and Vision Metrology (IEVM), 2006.
- [213] M. Lindner, A. Kolb: *Calibration of intensity-related distance error of the PMD ToF-Camera*. Intelligent Robots and Computer Vision XXV, 6764, 2007.
-

## G Literaturverzeichnis

---

- [214] R. Schwarte: *Dynamic3D-Vision*. IEEE Int. Symp. on Electron Devices for Microwave and Opto-electronic Applications. EDMO 2001, Wien.
- [215] H. Heinol: *Untersuchung und Entwicklung von modulationslaufzeitbasierten 3D-Sichtsystemen*. Dissertation, Universität Siegen, 2001.
- [216] B. Buxbaum, R. Schwarte, Z. Xu, T. Ringbeck, H. Heinol, J. Schulte, W. Tai, Z. Zhang, X. Luan, J. Fricke: *Charge Transfer Simulation in PMD-Structures*. Sensor 99, 1999.
- [217] ZF-Projekt Nr. 666c – Cluster 7: P0mSe (Personen- und Objektdetektion mit mobilen Sensoren). IT-Sicherheitsforschungsprogramm des BSI (Bundesamt für Sicherheit in der Informationstechnik). Unterstützt mit Mitteln aus dem Zukunftsfonds der Bundesregierung (2006-2010). <http://www.bsi.bund.de/>
- [218] Abschlussbericht Projekt P0mSe, Arbeitspaket 1. August 2009. Anfragen diesbezüglich an: [info@zess.uni-siegen.de](mailto:info@zess.uni-siegen.de)
- [219] Abschlussbericht Projekt P0mSe, Arbeitspaket 2, Teil A. Juli 2010. Anfragen diesbezüglich an: [info@zess.uni-siegen.de](mailto:info@zess.uni-siegen.de)
- [220] M. Zubair, O. Loepprich, S. Ghobadi, K. Hartmann, M. Koch: *Object Detection and Classification in MultiChannel Seismic Network*. SPPRA, 11th IASTED International Conference on Signal Processing, Pattern Recognition and Applications. Innsbruck, Austria, 17-19.02.2010.
- [221] H. Ruser, K. Wenzl, Ch. Kargel: *Detection and tracking of humans with a sparse network of LIDAR sensors*. Defense, Security, and Sensing - SPIE 2009, 13.-17.April 2009, Orlando, Florida, USA.
- [222] T. Ringbeck, T. Möller, B. Hagebeucker: *Multidimensional measurement by using 3-D PMD sensors*. Advances in Radio Science, 5, 135–146, 2007. [www.adv-radio-sci.net/5/135/2007/](http://www.adv-radio-sci.net/5/135/2007/)
- [223] Z. Xu: *Investigation of 3D-Imaging Systems Based on Modulated Light and Optical RF-Interferometry (ORFI)*. ZESS Forschungsberichte (1999).
- [224] M. Plaue: *Technical Report: Analysis of the PMD Imaging System*. Interdisciplinary Center for Scientific Computing, Univ. of Heidelberg. December 2006.
- [225] H. Rapp: *Experimental and theoretical investigation of correlating ToF-camera systems*. Master thesis, University of Heidelberg, September 2007.
- [226] S. Foix, G. Alenya, C. Torras: *Lock-in Time-of-Flight (ToF) Cameras: A Survey*. IEEE SENSORS JOURNAL, VOL. 11, NO. 3, MARCH 2011.
- [227] M. Lindner: *Calibration and Realtime Processing of Time-of-Flight Range Data*. Dissertation, Universität Siegen, 2010.
- [228] C. M. Bishop: *Pattern Recognition and Machine Learning*. Springer Verlag, 2006. ISBN-10: 0-387-31073-8.
- [229] MOVEDETECT (Sichere Detektion und Lokalisation in kritischen Arealen durch klassifizierende drahtlose Sensornetze). BSI Projekt 684/2010. <https://www.fim.uni-passau.de/en/fim/faculty/chairs/computer-networks-and-communications/research/projects/bsi-movedetect.html>
- [230] <http://www.timedomain.com/p400.php>
- [231] <https://www.pololu.com/catalog/product/1255>
- [232] Hella People Counter: <https://www.hella.com/hella-com/319.html>
- [233] Brickstream SeeMore 3D: <http://www.brickstream.com/products/seemore>
-