

Schlussbericht nach Nr. 8.2 NKBF 98

BMBF-Verbundforschungsvorhaben

**ZE: EADS Deutschland GmbH**  
**Förderkennzeichen: 13N10027**



**Vorhabenbezeichnung:**  
**„Verbesserung der Sicherheit von Verkehrsinfrastrukturen (SiVe)“**

**Teilvorhaben:**  
**„Bildverarbeitungs-basierte Sicherheitstechnologien“**

Laufzeit des Vorhabens: 01.07.2008 - 31.06.2011

kostenneutral verlängert bis 30.09.2011

Projektleiter: Olaf Heinzinger

Autoren: Johannes Schels, Falk Schubert, Klaus Schertler, Jörg Liebelt

EADS Deutschland GmbH,  
handelnd für den Bereich EADS Innovation Works Germany  
Willy-Messerschmitt-Str., 85521 Ottobrunn

Das diesem Bericht zugrundeliegende Vorhaben wurde mit Mitteln des Bundesministeriums für Bildung, und Forschung unter dem Förderkennzeichen 13N10027 gefördert. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt bei den einzelnen Autoren.

## Berichtsblatt

1. ISBN oder ISSN    geplant	2. Berichtsart (Schlussbericht oder Veröffentlichung)    Schlussbericht
3. Titel Verbesserung der Sicherheit von Verkehrsinfrastrukturen: Bildverarbeitungs-basierte Sicherheitstechnologien	
4. Autor(en) [Name(n), Vorname(n)] Schels, Johannes; Schubert, Falk; Schertler, Klaus; Liebelt, Jörg	5. Abschlussdatum des Vorhabens 30.09.2011
	6. Veröffentlichungsdatum    geplant
	7. Form der Publikation    geplant
8. Durchführende Institution(en) (Name, Adresse) EADS Deutschland GmbH, EADS Innovation Works Germany Willy-Messerschmitt-Str., 85521 Ottobrunn	9. Ber. Nr. Durchführende Institution
	10. Förderkennzeichen 13N10027
	11. Seitenzahl    61
12. Fördernde Institution (Name, Adresse) Bundesministerium für Bildung und Forschung (BMBF) 53170 Bonn	13. Literaturangaben    28
	14. Tabellen    0
	15. Abbildungen    45
16. Zusätzliche Angaben	
17. Vorgelegt bei (Titel, Ort, Datum)	
18. Kurzfassung <p>Themenschwerpunkt von SiVe („Verbesserung der Sicherheit von Verkehrsinfrastrukturen“) ist die systematische Analyse, Modellierung, Simulation, Berechnung und Evaluierung von Bedrohungs- und Lastszenarien sowie von Sicherheitssystemen. Neuartige Risiken und Bedrohungssituationen, die zunehmende Komplexität der Sicherheitssysteme und das bislang praktizierte reaktive Risiko- und Sicherheitsmanagement motivieren eine objektive Bewertung von Restrisiken und die daran orientierte werkzeugunterstützte Optimierung von Sicherheitssystemen.</p> <p>An Flughäfen sind Videokameras für Überwachungsaufgaben installiert. Die automatische Auswertung dieser Bilddatenströme birgt großes Potential, durch die automatische Erkennung von Risikoindikatoren die Sicherheit zu erhöhen. Im Rahmen dieses Teilvorhabens „Bildverarbeitungs-basierte Sicherheitstechnologien: Technologieeinsatz und Gefahrenfrüherkennung“ wurden aktuelle Forschungsarbeiten zur Videobilddatenauswertung vorangetrieben, für Sicherheitsaufgaben an Flughäfen angepasst und die Wirksamkeit dieser Verfahren in speziellen Technologieausprägungen quantifiziert. Es wurden folgende drei Themenschwerpunkte bearbeitet: die Verbesserung von Überwachungsbilddaten zur Erkennung statischer Anomalien, um beispielsweise unter schwierigen Lichtverhältnissen abgestellte Objekte erkennen zu können; die Möglichkeiten zur Abwehr der Blendung von Piloten im Landeanflug durch handelsübliche Laserpointer; sowie die Gewinnung und Nutzung von Kontextwissen zur Erkennung von Objekten für die bildbasierte Suche nach Verdächtigen in Überwachungskameradaten. Auf repräsentativen Test- sowie auf realen Feldtestdaten konnte die Wirksamkeit der entwickelten Verfahren demonstriert werden.</p>	
19. Schlagwörter Verkehrsinfrastruktur, Sicherheit, Bildverarbeitung, Maschinensehen, Bildverbesserung, Laserblendung	
20. Verlag	21. Preis

## Document Control Sheet

1. ISBN or ISSN    tbd	2. type of document (e.g. report, publication)    final report
3. Title Improving the Security of Traffic Infrastructures using Image Processing Technologies	
4. author(s) (family name, first name(s)) Schels, Johannes; Schubert, Falk; Schertler, Klaus; Liebelt, Jörg	5. end of project 30.09.2011
	6. publication date    tbd
	7. form of publication    tbd
8. Durchführende Institution(en) (Name, Adresse) EADS Deutschland GmbH, EADS Innovation Works Germany Willy-Messerschmitt-Str., 85521 Ottobrunn	9. originator's report no.
	10. reference no.    13N10027
	11. no. of pages    61
12. sponsoring agency (name, address) Bundesministerium für Bildung und Forschung (BMBF) 53170 Bonn	13. no. of references    28
	14. no. of tables    0
	15. no. of figures    45
16. supplementary notes	
17. presented at (title, place, date)	
18. abstract  Focus of SiVe (Improving the security of traffic infrastructures) is the systematic analysis, modeling, simulation, computation and evaluation of threat and load scenarios as well as security systems. New risks and threat situations, the increasing complexity of security systems and the commonly employed reactive risk and security management justify an objective assessment of remaining risks and a task-oriented, tool-assisted optimization of these security systems.  Cameras for security surveillance are ubiquitous in airports. The automatic assessment of these video streams can potentially contribute to automatically identifying risk indicators and thereby increase security. As part of the research project "Image processing based security technologies: technology deployment and early threat identification", recent research on image analysis was adapted to the security tasks on airports and its effectiveness in specific scenario deployments was assessed. A particular focus was placed on the following three topics: improving the image quality of surveillance cameras in order to detect static anomalies such as abandoned suitcases under difficult light conditions; potential countermeasures against intentional blending of pilots during the final approach using off-the-shelf laser pointers; the derivation and use of context information for improved object detection during the image-based retrieval of suspects in surveillance camera images. On representative as well as field test data, the effectiveness of the developed methods was demonstrated.	
19. keywords Traffic infrastructure, security, image processing, computer vision, image enhancement, laser blending	
20. publisher	21. price

# Inhalt

<b>I.</b>	Kurze Darstellung.....	7
<b>I.1.</b>	Aufgabenstellung.....	7
<b>I.1.</b>	Voraussetzungen, unter denen das Vorhaben durchgeführt wurde.....	7
<b>I.2.</b>	Planung und Ablauf des Vorhabens.....	7
<b>I.3.</b>	Wissenschaftlicher und Technischer Stand bei Vorhabensbeginn.....	8
<b>I.4.</b>	Zusammenarbeit mit anderen Stellen .....	8
<b>II.</b>	Eingehende Darstellung .....	10
<b>II.1.</b>	Verwendung der Zuwendung und erzielte Ergebnisse .....	10
<b>II.1.1.</b>	Rollbrückenszenario: Verbesserung von Überwachungsbilddaten und Erkennung statischer Anomalien.....	10
<b>II.1.2.</b>	Laserblendungsszenario: Erkennung dynamischer Anomalien .....	23
<b>II.1.3.</b>	Kofferszenario: Gewinnung von Szenen-Wissen zur Interpretation und Darstellung erkannter Anomalien .....	29
<b>II.2.</b>	Wichtigste Positionen des zahlenmäßigen Nachweises .....	51
<b>II.3.</b>	Notwendigkeit und Angemessenheit der geleisteten Arbeit .....	51
<b>II.4.</b>	Verwertbarkeit des Ergebnisses.....	51
<b>II.4.1.</b>	Schutzrechtsanmeldungen .....	51
<b>II.4.2.</b>	Weitergehende Nutzung der Ergebnisse.....	52
<b>II.5.</b>	Fortschritt auf dem Gebiet des Vorhabens bei anderen Stellen .....	53
<b>II.6.</b>	Veröffentlichungen des Ergebnisses .....	53
<b>III.</b>	Erfolgskontrollbericht.....	57
<b>III.1.</b>	Beitrag des Ergebnisses zu den förderpolitischen Zielen.....	57
<b>III.2.</b>	Wissenschaftlich-technisches Ergebnis des Vorhabens.....	57
<b>III.3.</b>	Fortschreibung des Verwertungsplans.....	57
<b>III.3.1.</b>	Schutzrechtsanmeldungen .....	57
<b>III.3.2.</b>	Wirtschaftliche Erfolgsaussichten nach Projektende.....	57
<b>III.3.3.</b>	Wissenschaftliche Erfolgsaussichten nach Projektende .....	58
<b>III.3.4.</b>	Wissenschaftliche und wirtschaftliche Anschlussfähigkeit.....	58
<b>III.4.</b>	Arbeiten, die zu keiner Lösung geführt haben .....	58
<b>III.4.1.</b>	Durchwurferkennung als dynamische Anomaliedetektion .....	58
<b>III.4.2.</b>	Versuche zur Laserblendungsdetektion im Kontext ziviler Luftfahrt.....	58
<b>III.4.3.</b>	Anwendung der Bildverbesserung auf Rollbrücken-Feldmessungsdaten .....	58
<b>III.5.</b>	Präsentationsmöglichkeiten für mögliche Nutzer .....	59
<b>III.6.</b>	Einhaltung der Ausgaben- und Zeitplanung.....	59

## Abbildungen

Abbildung 1: HDR-Fusion .....	12
Abbildung 2: SR-Fusion.....	12
Abbildung 3: Kombination SR/HDR .....	13
Abbildung 4: Aufnahmen mit unterschiedlichen Belichtungszeiten (links: kurze, rechts: lang).	14
Abbildung 5: HDR-Bild welches aus zwei Bildern unterschiedlicher Belichtung erzeugt wurde.	14
Abbildung 6: Detektion kurzfristiger Veränderungen bezüglich des multimodalen Hintergrundmodelles.....	15
Abbildung 7: Zeitliche Verfolgung von Hintergrundveränderungen.....	15
Abbildung 8: Detektion langfristiger Veränderungen.....	16
Abbildung 9: Prototyp-Software .....	17
Abbildung 10: Demonstrationssetup .....	17
Abbildung 11: LR Input.....	18
Abbildung 12: HDR/SR Ergebnis .....	18
Abbildung 13: LR Input Zoom 1 .....	18
Abbildung 14: HDR/SR Ergebnis Zoom 1.....	19
Abbildung 15: LR Input Zoom 2 .....	19
Abbildung 16: HDR/SR Ergebnis Zoom 2.....	19
Abbildung 17: Lage der Rollbrücke, die für die Feldtest-Aufnahmen ausgewählt wurde .....	21
Abbildung 18: Beispiel-Bilder der Überwachungskameras von Rollbrücke N2. ....	21
Abbildung 19: Beispiel-Bild für das Szenario "abgestelltes Auto" (Auto in Rot markiert).....	21
Abbildung 20: Beispiel-Bild für das Szenario "abgestellter Koffer" (Koffer in Rot markiert).....	22
Abbildung 21: Beispiel-Bild für das Szenario "angebrachter Haftmarker" (Marker in Rot markiert).....	22
Abbildung 22: Laserblendung.....	24
Abbildung 23: Optra (Optra.com).....	25
Abbildung 24: Vergleich Laserlicht.....	26
Abbildung 25: Gitterdispersion .....	27
Abbildung 26: Laboraufbau.....	28
Abbildung 27: Verarbeitungskette .....	28
Abbildung 28: Synthetische Bilddaten .....	31
Abbildung 29: Verwendete Gepäckstücke für die Feldmessungen zur "Wiedererkennung von Gepäckstücken" .....	31
Abbildung 30: Positionen der Kameras für die Feldmessungen zur "Wiedererkennung von Gepäckstücken" .....	32
Abbildung 31: Beispiel-Bilder für das Szenario "Wiedererkennung von Gepäckstücken" .....	33
Abbildung 32: Komponenten der Objekterkennung .....	33
Abbildung 33: CAD-Gepäckstückmodelle.....	34
Abbildung 34: Ohne örtlichen Kontext sind die beiden ausgewählten Bildbereiche schwer zu interpretieren; dies ist ein typisches Problem bei der Detektion wenig charakteristischer Objektklassen.....	35

Abbildung 35: Örtlicher Kontext zwischen einer Person und einem Trolley: die 2D-Detektion der Person (durchgezogene graue Linie) erlaubt die Bestimmung eines Bildbereichs in 2D, in dem die Suche nach einem Trolley aussichtsreich ist (gestrichelte rote Linie). Allerdings berücksichtigt dies nicht die perspektivischen Veränderungen des Bildes, die durch die Kamera induziert werden können (hier beispielsweise eine starke Kameraneigung). Im Gegensatz dazu beschreibt der projizierte 3D-Kontext den Suchbereich sehr viel genauer (durchgezogene rote Linie).....	36
Abbildung 36: Objektdetektoren, näherungsweise Kameraselbstkalibration und 3D-Objektkontext ermöglichen die Filterung von Einzelbildern verschiedener Überwachungskameras ohne zeitliche Konsistenzanforderungen. In der Folge können komplexe attributsbasierte Suchanfragen beantwortet werden, die Objektklassen, Erscheinungsbild und metrische Angaben enthalten können.....	38
Abbildung 37: Teile-basierte Objektklassendetektoren werden auf synthetischen Daten trainiert.....	39
Abbildung 38: Darstellung der geschätzten Bodenebene für eine Kamera. ....	41
Abbildung 39: 3D Ortskontext für Gepäckstücke, die von einer Person getragen werden können. Falls der Menschendetektor Poseinformation zur Verfügung stellt, kann die Auftretenswahrscheinlichkeit der verschiedenen Gepäckstückarten (rot: Trolley, grün: Koffer, blau: Rucksack) als 3D Gaussellipsoide modelliert werden (links), deren 2D-Projektionen sich einfach berechnen lassen. Im vereinfachten Fall von 2D-Menschendetektionen ohne Poseangaben werden die Ellipsoide zu rotationsinvarianten Tori (rechts).....	42
Abbildung 40: Gefilterte Detektionen nach jedem der beschriebenen Verarbeitungsschritte: Reine 2D-Detektionen (links) von Menschen (grün) und Rucksäcken (rot); die Schwierigkeit der Detektion wenig charakteristischer Objektklassen wird deutlich. Nach 3D-Selbstkalibration können Detektionen, die mit der geschätzten 3D-Szene nicht konsistent sind, entfernt werden (Mitte), aber die Erkennungsrate ist noch immer zu gering. Nach Anwendung des 3D-Kontextmodells anhand der detektierten Menschen verbessert sich die Genauigkeit des Rucksackdetektors erheblich (rechts).....	45
Abbildung 41: Einfluss des vorgeschlagenen 3D-Kontextmodells auf die Genauigkeit der Objektklassendetektion für drei Kofferarten Trolley (links), Koffer (Mitte) und Rucksack (rechts) auf dem gesamten Testdatensatz aller Kameras. Dargestellt sind die reinen 2D-Detektionsergebnisse (grün), die Ergebnisse nach Filterung anhand der rekonstruierten Szene (blau) und das Endergebnis des 3D-Objektkontextmodells (rot).....	45
Abbildung 42: Verbesserung der Genauigkeit der Suchanfragen bei Rang 25 für verschiedene Suchanfragen bei Verwendung des 3D-Objektkontextmodells (rot); komplexe Suchanfragen mit metrischen Angaben können mit dem reinen 2D-Ansatz (grün) nicht beantwortet werden. ....	46
Abbildung 43: Beispiele für Suchergebnisse auf die Anfrage „Person durchschnittlicher Größe mit rotem Trolley“.....	48
Abbildung 44: Beispiele für Suchergebnisse auf die Anfrage „Person durchschnittlicher Größe mit dunklem Rucksack, hellem Hemd und dunkler Hose“.....	49
Abbildung 45: Beispiele für Suchergebnisse auf die Anfrage „grosse Person mit dunklem Koffer, hellem Hemd und dunkler Hose“.....	50

# I. KURZE DARSTELLUNG

## I.1. AUFGABENSTELLUNG

Themenschwerpunkt von SiVe („Verbesserung der Sicherheit von Verkehrsinfrastrukturen“) ist die systematische Analyse, Modellierung, Simulation, Berechnung und Evaluierung von Bedrohungs- und Lastszenarien sowie von Sicherheitssystemen. Neuartige Risiken und Bedrohungssituationen, die zunehmende Komplexität der Sicherheitssysteme und das bislang praktizierte reaktive Risiko- und Sicherheitsmanagement motivieren eine objektive Bewertung von Restrisiken und die daran orientierte werkzeugunterstützte Optimierung von Sicherheitssystemen.

Am Flughafen München sind etwa 1600 Videokameras für Überwachungsaufgaben installiert. Die automatische Auswertung dieser Bilddatenströme birgt ein großes Potential, durch die automatische Erkennung von Risikoindikatoren die Sicherheit zu erhöhen. Im Rahmen dieses Teilvorhabens „Bildverarbeitungs-basierte Sicherheitstechnologien: Technologieeinsatz und Gefahrenfrüherkennung“ sollen aktuelle Forschungsarbeiten zur Videobilddatenauswertung vorangetrieben und für Sicherheitsaufgaben an Flughäfen angepasst werden (AP 2.2.1, 2.2.5). Ziel ist dabei die Forschung im Bereich von bildverarbeitungs-basierten Sicherheitstechnologien, die sich unspezifisch auf Anomalien richten und auf neue Bedrohungen angepasst werden, um damit die Ausgangsbasis für spätere Produktentwicklungen im Überwachungsbereich zu schaffen.

## I.2. VORAUSSETZUNGEN, UNTER DENEN DAS VORHABEN DURCHGEFÜHRT WURDE

Das Vorhaben hatte eine Laufzeit von drei Jahren und drei Monaten und wurde durch den Fördergeber mit einer Zuwendung in Höhe von 1.476.500 EUR ausgestattet. Das Teilvorhaben „Bildverarbeitungs-basierte Sicherheitstechnologien“ wurde so ausgelegt, dass die drei Teilbereiche Erkennung von statischen Anomalien, Erkennung von dynamischen Anomalien und Gewinnung von Szenenwissen bearbeitet werden sollen. Als Projektpartner aus Anwendersicht konnte die Flughafen München GmbH gewonnen werden. Entsprechend des Projektantrags sollte die detaillierte Ausgestaltung der Arbeitsanteile in enger Abstimmung mit dem Anwender erst nach Projektbeginn erfolgen. Dabei wurde für jedes der im Projektantrag beschriebenen Forschungsthemen ein Szenario erarbeitet, das einerseits als Anwendungsszenario für die Forschungsarbeiten geeignet ist und andererseits einem realen Bedarf im Rahmen des Sicherheitskonzeptes des Flughafens München entspricht. Der Flughafen München installierte hierzu Kameras an besonders geeigneten Orten für die Bildverarbeitungsverfahren und stellte im Rahmen eines Feldtests geeignetes Videobildmaterial bereit.

## I.3. PLANUNG UND ABLAUF DES VORHABENS

Es wurden drei Arbeitsgebiete identifiziert, die vom Flughafen München identifizierten Szenarien entsprechen; diese wurden parallel vorangetrieben.

Im sogenannten „Rollbrückenszenario“ sollte die Verbesserung von Überwachungsbilddaten und Erkennung statischer Anomalien untersucht werden, die sich daraus ergeben, dass in einem auf einer längeren Zeitskala als statisch angenommenen Bildbereich (z.B. Gebäudeinfrastruktur) neue statische Elemente (z.B. abgestellte Koffer oder Infrastruktur-Manipulationen) erkannt werden.

Im „Laserblendungsszenario“ wurde die Erkennung dynamischer Anomalien untersucht; der Begriff der „dynamischen Anomalie“ wurde dabei auf die Blendung von Piloten im Landeanflug durch handelsübliche Laserpointer bezogen.

Im sogenannten „Kofferszenario“ ging es um die Gewinnung und Nutzung von Szenen-Wissen zur Interpretation und Darstellung erkannter Anomalien. Dabei wurde ein automatisiertes Auswertungsverfahren für Bilddaten von Überwachungskameras erarbeitet, welches die Arbeit der Sicherheitsbeauftragten am Flughafen München erleichtern kann.

Unabhängig von der Verfügbarkeit der Feldtestdaten konnten theoretische Vorarbeiten auf simulierten Testdaten kontinuierlich validiert werden, sodass die zeitlichen Abhängigkeiten zu anderen Teilvorhaben und Projektpartnern minimiert wurden. Nach dem Kickoff am 2008-07-22 fanden neben regelmäßigen Projektmeetings am 2008-08-07 eine erste eintägige Flughafen-Führung statt, in der das bestehende Überwachungskameraszenario am Flughafen München erläutert sowie Szenarien diskutiert wurden, welche mit dem Stand der Technik zu Vorhabensbeginn nicht zufriedenstellend gelöst werden konnten. Am 2009-05-08 wurden die zu bearbeitenden Szenarien bei einem weiteren Termin am Flughafen detailliert festgelegt. Am 2010-04-21 und 2010-04-29 fanden die Feldtests am Flughafen München statt, bei denen Testdaten aufgenommen wurden. In insgesamt 14 Rollbrückenkameras (Brücken N1, N2, S5, S6) wurden ca. 250000 Einzelbilder (ca.400 GB) für das Rollbrückenszenario und in 3 Überwachungskameras in den Terminals 1 und 2 wurden weitere 128402 Bilder (108 GB) für das Kofferszenario aufgenommen. Die Ergebnisse wurden am 2011-09-21 auf dem Abschlussmeeting präsentiert und trafen seitens des Flughafens München auf positive Rückmeldung.

#### **I.4. WISSENSCHAFTLICHER UND TECHNISCHER STAND BEI VORHABENSBEGINN**

Kamerasysteme sind seit langem in Sicherheitssystemen im Einsatz. Die Bilder dieser Kameras werden alarmgesteuert oder über manuelle Auswahl in die Überwachungszentralen aufgeschaltet. Eine Speicherung oder eine (teil-)automatische (assistierte) Auswertung der Bilddaten in intelligenten Kameras oder zentralen Auswerterechnern findet nicht statt. Die Modellierung und Simulation von in Sicherheitsszenarien eingesetzten Bildverarbeitungsverfahren ist nur teilweise methodisch entwickelt. Die Wirksamkeit dieser Verfahren und insbesondere spezieller Technologieausprägungen konnte vor Projektbeginn nicht ausreichend quantifiziert werden.

Zu Beginn der Arbeiten an den drei Teilthemengebieten wurde jeweils der Stand der Forschung dokumentiert; die detaillierten Angaben hierzu finden sich in II.1.1, Abschnitt Verwandte Arbeiten, in II.1.2, Stand der Technik, und in II.1.3, Verwandte Arbeiten.

#### **I.5. ZUSAMMENARBEIT MIT ANDEREN STELLEN**

Als Projektpartner aus Anwendersicht konnte die Flughafen München GmbH gewonnen werden. Entsprechend des Projektantrags wurde in enger Abstimmung mit dem Flughafen München für jedes der im Projektantrag beschriebenen Forschungsthemen ein Szenario erarbeitet, das einerseits als Anwendungsszenario für die Forschungsarbeiten geeignet ist und andererseits einem realen Bedarf im Rahmen des Sicherheitskonzeptes des Flughafens München entspricht. Der Flughafen stellte im Rahmen von Feldtests Datenmaterial zur Verfügung, um die Bewertung der Verfahren auf realistischen Szenarien zu ermöglichen und selbst die Nutzbarkeit der entwickelten Verfahren bewerten zu können.

## II. EINGEHENDE DARSTELLUNG

### II.1. VERWENDUNG DER ZUWENDUNG UND ERZIELTE ERGEBNISSE

#### II.1.1. *ROLLBRÜCKENSZENARIO: VERBESSERUNG VON ÜBERWACHUNGSBILDDATEN UND ERKENNUNG STATISCHER ANOMALIEN*

##### **Einleitung und Szenariobeschreibung**

Statische Anomalien in Videobildmaterial sind Anomalien, die sich daraus ergeben, dass in einem auf einer längeren Zeitskala als statisch angenommenen Bildbereich (z.B. Gebäudeinfrastruktur) neue statische Elemente (z.B. abgestellte Koffer oder Infrastruktur-Manipulationen) erkannt werden. Hierzu zählen zum Beispiel die Erkennung statischer Elemente in der Bildszene auf unterschiedlichen Zeitskalen und die Klassifikation der statischen Elemente in zeitliche Kategorien wie kurzzeitig abgestellte Objekte (z.B. Koffer) und langfristig vorhandene Objekte (z.B. Gebäudeinfrastruktur).

Durch den Sicherheitsbeauftragten der FMG wurden sogenannte „Rollbrücken“ als potentiell Sicherheitsrisiko hervorgehoben. Diese Brücken verbinden die Landebahnen mit den Terminals. Einige dieser Brücken führen über öffentlich zugängliche Strassen und bieten aufgrund ihres hohen Flugzeugverkehrs ein geeignetes Ziel für terroristische Angriffe (z.B. parkendes Fahrzeug mit Sprengstoff). Zur Absicherung sind deshalb mehrere Kameras unter den Brücken installiert, die manuell überwacht werden. Ziel ist dabei Objekte zu identifizieren, die kein vorbeifahrendes Auto darstellen (z.B. parkendes Auto, vorbeilaufende Person). Zur effizienten Sicherung dieser Aufgabe wäre eine automatische Bildanalyse hilfreich, die die Veränderungen der Szene auswertet und das Sicherheitspersonal beim Erscheinen von unerlaubten Objekten alarmiert. Weiterhin sind die Kameras, speziell in diesem Anwendungsszenario, starken Beleuchtungsschwankungen ausgesetzt, was die manuelle und automatische Bildauswertung erschwert. An einem hellen Tag ist die Lichteinstrahlung im Eingangs- bzw. Ausgangsbereich der Brücke sehr hoch wohingegen der Bereich unterhalb der Brücke relativ dazu wenig beleuchtet ist. Da die eingesetzten Kameras zur praktischen Nutzung einen weiten Blickwinkel besitzen, sind sowohl sehr lichtstarke als auch lichtschwache Szenenbereiche in einem Kamerabild sichtbar. Abhängig vom dynamischen Umfang des verwendeten Kamerasensors werden diese nur teilweise im Bild sinnvoll abgebildet (d.h. sehr dunkle bzw. helle Szenenabschnitte werden unter- bzw. übersättigt dargestellt). Dieses Problem wurde bei der Sichtung vom Bildmaterial der aktuell verwendeten Kameras festgestellt. Um hohe Kosten bei der Beschaffung von noch leistungsfähigeren Kameras (z.B. Kameras mit sehr hohem dynamischen Bereich) zu vermeiden, war deshalb eine softwarebasierte Bildverbesserung wünschenswert, die Bilder generiert in denen sowohl sehr helle als auch sehr dunkle Szenenbereiche sinnvoll abgebildet werden.

Antragsgemäss wurden hierzu im Rahmen des Projektes Methoden der Bildverbesserung und Bildrekonstruktion zur Erzeugung von hoch aufgelösten Detailbildern entwickelt und die derart verbesserten Bilddaten zur Erkennung von statischen Anomalien genutzt. Als Anwendung für die zu entwickelnden Bildverbesserungsverfahren wurde die automatische Erkennung von abgestellten Fahrzeugen, stehenbleibenden Personen oder abgelegten größeren Gegenständen

identifiziert. Durch die Bildverbesserungsmaßnahmen sollte eine erhöhte Erkennungsgenauigkeit erzielt werden. Um die Verbesserungen quantitativ bewerten zu können war es Ziel, ein System für eine automatische Langzeitbewegungsdetektion zu entwickeln, welches die installierten Rollbrückenkameras nutzt. Innerhalb dieses Systems sollten optional die Bildverbesserungsverfahren zuschaltbar sein. Das System sollte in der Lage sein, automatisch über einen längeren Zeitraum Veränderungen im Hintergrund der Szene festzustellen und einen Alarm auszugeben, falls eines der kritischen Ereignisse erkannt wird.

### **Verwandte Arbeiten**

Es existieren verschiedene Methoden, um aus niedrig aufgelösten Bildsequenzen eine qualitätsverbesserte Bildsequenz zu erzeugen. Diese Methoden reichen von der klassischen Bildverarbeitung auf Einzelbildern (Helligkeitsnormalisierung, Kontrastverbesserung, Schärfung, Rauschunterdrückung, ...) bis hin zu Methoden, welche mehrere Bilder einer Bildsequenz (Bildstapel) fusionieren [25,26,27,28]. Hierzu zählen die Methoden der Superresolution (SR) und der High-Dynamic-Range Rekonstruktion (HDR).

Der Nachteil dieser Methoden besteht darin, dass diese jeweils nur auf einen beschränkten Teil der von der Sensorik gewinnbaren Bildinformationen angewendet werden. So arbeiten z.B. die klassischen Bildverarbeitungsmethoden nur auf Einzelbildern. Die Fusionsmethoden arbeiten hingegen entweder nur auf im Bildstapel enthaltenen redundanten Informationen im Sinne einer Mehrfachmessung (Superresolution) oder ausschließlich auf komplementären Informationen durch Variation von Sensorparametern (Belichtungsserie im HDR-Verfahren). Das Potential der gemeinsamen Nutzung von sowohl redundanten als auch komplementären Bildinformationen wird in diesen Methoden nicht ausgenutzt.

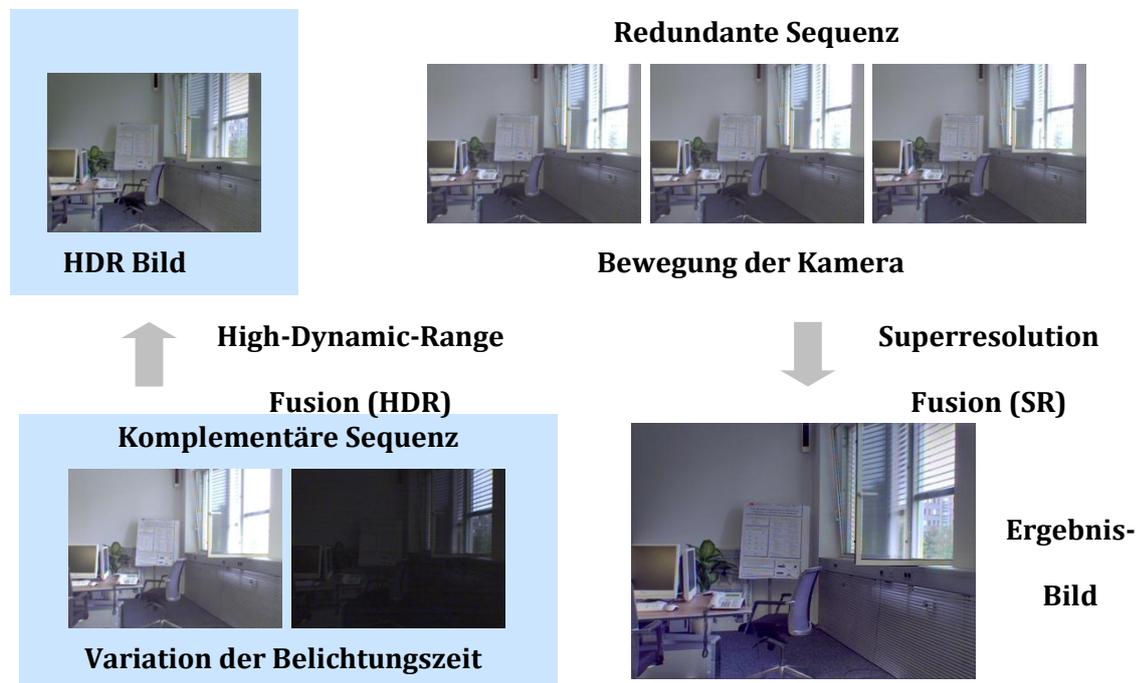
### **Theoretische Vorarbeiten**

Anhand der Literaturrecherche zu existierenden Methoden für High-Dynamic-Range-Fusion zur Vergrößerung des Dynamikumfangs, Superresolution zur Verbesserung der Auflösung, Bewegungsdetektion, Hintergrundmodellierung und deren Implementierung auf parallelen Architekturen und einer Analyse der Vor- und Nachteile der existierenden Methoden wurde ein integrierter Lösungsansatz für das vorliegende Szenario erarbeitet. Dieses Bildverbesserungssystem basiert auf der Fusion von mehreren Bildern, die durch aktive Steuerung des Aufnahmesensors (z.B. Variation der Belichtungszeiten – HDR Fusion – oder Bewegung des Sensors – SR Fusion) erzeugt wurden.

Bei der HDR Fusion werden anhand eines Optimierungsprozesses jeweils die Bildanteile der komplementären Eingabebilder extrahiert, die die geringste Sättigung aufweisen. Diese Bildanteile werden zu einem HDR Bild zusammengefasst. In folgender Abbildung ist der Ablauf der Fusion schematisch dargestellt.



ergänzender Weise (HDR Fusion verbessert den Dynamikumfang, die SR Fusion verbessert die Pixelauflösung). Es liegt daher nahe beide Fusionsmethoden zu geeignet zu kombinieren. Folgende Abbildung zeigt den, im Rahmen dieses Projektes entwickelten, Ansatzes zur Kombination beider Fusionsmethoden.



**Abbildung 3: Kombination SR/HDR**

Die Forschungsergebnisse wurden in Form einer Publikation auf einer Fachkonferenz validiert. Weitere Details können dieser Publikation entnommen werden.

Der bisherige entwickelte Fusionsprototyp war für den Einsatz in Echtzeitsystemen zu rechenintensiv. Um dennoch eine lauffähige Implementierung zu gewährleisten, wurden die Berechnungen für die HDR Fusion teilweise parallelisiert. Dazu wurden Teile der Software mittels CUDA zu Berechnung auf der Graphikkarte (GPU) implementiert. Ein Anteil bestand dabei in der schnellen Lösung der 2D Poisson-Gleichung zur Bildverbesserung; durch die GPU-basierten Parallelisierungen konnte für diese Funktion ein Beschleunigungsfaktor von 4 ggü. einer effizienten CPU -Implementierung auf einzelnen Bildern von Überwachungskameras (704x576) erreicht werden.



**Abbildung 4: Aufnahmen mit unterschiedlichen Belichtungszeiten (links: kurze, rechts: lang).**



**Abbildung 5: HDR-Bild welches aus zwei Bildern unterschiedlicher Belichtung erzeugt wurde.**

Nach ersten erfolgreichen Arbeiten zur Verbesserung des Dynamikumfangs und Auflösung, wurde Methoden zur Bewegungsdetektion untersucht, die speziell in der Lage sind, langfristige Veränderungen im Hintergrund zu erkennen. Dazu wurde zunächst eine intensive Literaturrecherche durchgeführt.

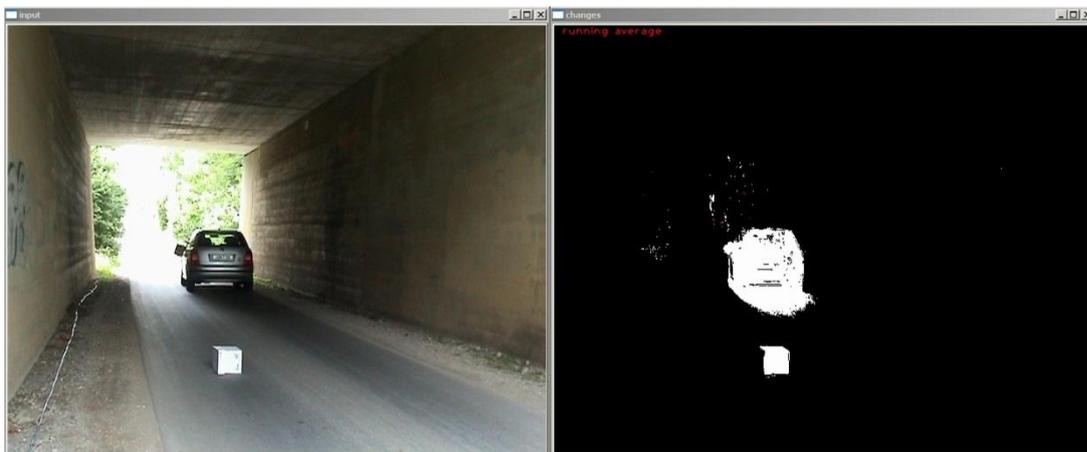
Die Methode muss in der Lage sein, leichte, wiederholende Veränderungen des Hintergrundes (z.B. Baum, der sich im Wind leicht bewegt, Helligkeitsschwankungen über einen längeren Zeitraum) von relevanten Bewegungen (z.B. fahrendes Auto, laufende Person) zu unterscheiden. Deshalb wurde die Methoden der „multi modalen Hintergrundbewegungsdetektion“ ausgewählt. Die unkritischen Veränderungen der Hintergrundpixel werden dabei als Überlagerung mehrerer Gaußverteilungen modelliert. Kleinere Schwankungen, die durch Sensorrauschen oder kurzfristige Bewegungen (z.B. vorbeifahrendes Auto) verursacht werden, können dadurch als multimodalen Hintergrund erfasst werden. Veränderungen, die von diesem adaptiven Hintergrundmodell über einen längeren Zeitraum abweichen, werden als Unregelmäßigkeit detektiert und lösen einen Alarm aus.

Folgende Abbildungen zeigen Ergebnisse der automatischen Detektion von langfristigen Änderungen des Hintergrundes (ein vorbeifahrendes Auto wirft 2 größere Objekte aus dem Fenster). In der folgenden Abbildung werden alle registrierten Bewegungen einer Testsequenz markiert. Wiederholende Veränderungen der Hintergrundpixel, die zum Beispiel durch Rauschen verursacht werden, sind dabei unterdrückt.



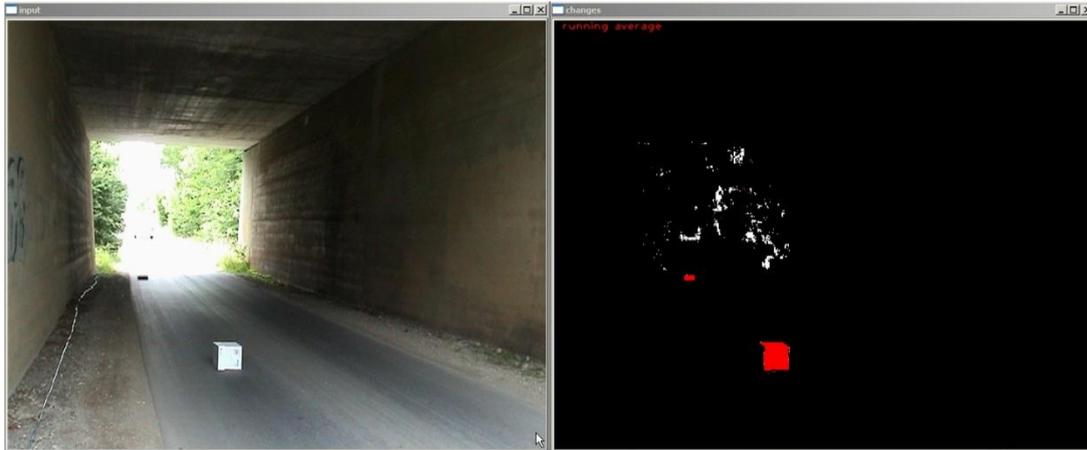
**Abbildung 6: Detection kurzfristiger Veränderungen bezüglich des multimodalen Hintergrundmodelles.**

Bereiche die langfristig von dem typischen Hintergrund abweichen, werden als potentiell kritische Objekte zeitlich verfolgt. Als Beispiel ist in folgender Abbildung eine abgeworfene Kiste registriert worden.



**Abbildung 7: Zeitliche Verfolgung von Hintergrundveränderungen.**

Überschreitet die Verweildauer kritischer Objekte eine gewählte Zeitgrenze, wird ein Alarm ausgelöst. In folgender Abbildung wurden dazu die kritischen Objekte rot markiert.



**Abbildung 8: Detektion langfristiger Veränderungen**

### **Fusionssoftware-Prototyp**

Es wurde eine Software entwickelt, die Bilder, die mit unterschiedlichen Kameraparametern aufgenommen wurden (z.B. Variation der Belichtungszeit) fusioniert. In der folgenden Abbildung 9 ist ein Screenshot der Bedienungsfläche dargestellt. Zur Demonstration werden dabei die benötigten Eingabebilder (z.B. Variation der Kameraparameter) mittels einer Kamera (z.B. Webcam) erstellt und auf einem angeschlossenen Rechner fusioniert. Die Erzeugung erfolgt einzelbildweise auf Benutzeranfrage. In Abbildung 10 ist dieser Prozess schematisch dargestellt. Die Ergebnisse, die mit Hilfe der Prototype-Software erzeugt wurden, veranschaulichen die Anwendung auf realem Bildmaterial. In Abbildung 11 ist eines der niedrig aufgelösten Eingabebilder (LR Input) mit geringem Dynamikumfang dargestellt. Abbildung 12 zeigt das verbesserte Bild (HDR-SR Result) mit doppelter Auflösung und höherem Dynamikumfang. Die übersättigten Bildbereiche (z.B. Fensters) im Eingabebild lassen kaum Szenendetails erkennen (siehe Abbildung 13). Diese Bereiche sind im Ergebnisbild deutlich besser sichtbar (siehe Abbildung 14). Ein Vergleich von feinen Strukturen (z.B. Text auf Poster) im Eingabebild (siehe Abbildung 15) und im Ergebnisbild (siehe Abbildung 16) zeigt ebenfalls deutlich die Bildverbesserungen.

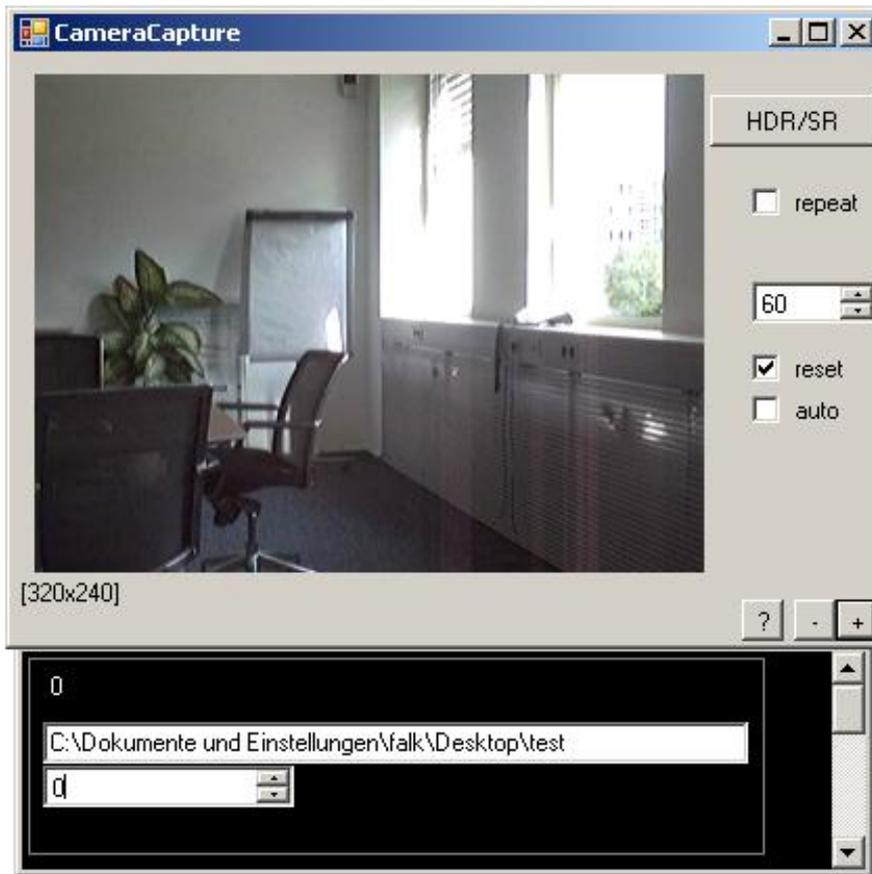


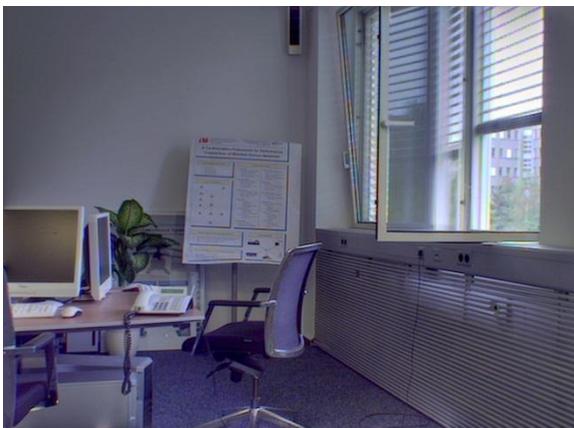
Abbildung 9: Prototyp-Software



Abbildung 10: Demonstrationssetup



**Abbildung 11: LR Input**



**Abbildung 12: HDR/SR Ergebnis**



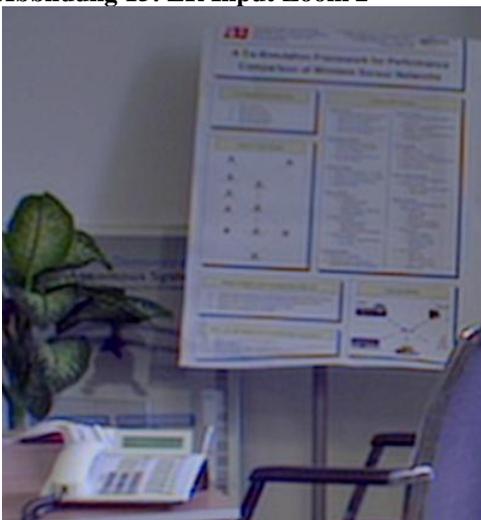
**Abbildung 13: LR Input Zoom 1**



**Abbildung 14: HDR/SR Ergebnis Zoom 1**



**Abbildung 15: LR Input Zoom 2**



**Abbildung 16: HDR/SR Ergebnis Zoom 2**

## **Feldmessungen Rollbrücken**

Die entwickelten Bildverarbeitungsmethoden müssen auf Ihre Praxistauglichkeit getestet werden. Weiterhin sollen Kenngrößen ermittelt werden, welche die Genauigkeit und Robustheit der Algorithmen charakterisieren. Diese Kennzahlen sollen dann in das Bewertungssystem einfließen. Dazu müssen reale Bilddaten aufgezeichnet werden, die den geplanten Anwendungen entsprechen (z.B. reale Aufnahmen von abgestellten Koffern unter einer Rollbrücke). Weiterhin ist eine Annotation (z.B. Markierung von Koffern in den Bilddaten zur Validierung der Algorithmen) dieser Bilddaten notwendig, damit eine quantitative Auswertung ermöglicht wird. Die Bilddaten wurden im Rahmen von Feldmessungen aufgezeichnet, die in intensiver Abstimmung mit dem Flughafen München und den örtlichen Behörden durchgeführt wurden. Die Aufzeichnungen fanden in den Kalenderwochen 16 und 17 in 2010 statt.

Die aufgezeichneten Bilddaten sollen so nah wie möglich das geplante Anwendungsszenario widerspiegeln. Das heißt verschiedene abgestellte Objekte, die an unterschiedlichen Bereichen der Rollbrücke platziert sind, sollen automatisch detektiert werden.

Für die Aufnahmen im Feldtest wurden dazu 3 verschiedene Objekttypen ausgewählt, die in unterschiedlichen Bereichen der Rollbrücke typischerweise platziert werden.

1. Koffer (verschiedene Größen und Farben), die an unterschiedlichen Stellen auf dem Gehweg unter der Brücke platziert werden.
2. Haftmarker (verschiedene Größen und Farbe), die an unterschiedlichen Stellen an den Brückenpfeilern angebracht wurden.
3. Auto, welches am Straßenrand geparkt wurde.

Bei einem ersten Aufnahmetermin wurden alle verfügbaren Kameras (10 Stück) unter den Rollbrücken N1, N2, S5 und S6 aufgezeichnet. Dabei wurden noch keine realen Szenarien (z.B. abgestellter Koffer unter der Rollbrücke) simuliert. Die umfangreichen Videos (210GB Bilddaten) wurden analysiert und die Brücke (N2, siehe Lageplan in Abbildung 17) mit den schwierigsten Belichtungsbedingungen ausgesucht.

Bei einem zweiten, späteren Aufnahmetermin wurden verschiedenen Szenarien mit abgestellten Gegenständen durchgeführt. Eine Person stellte dabei die jeweiligen Objekte in verschiedenen Bereichen unter der Brücke für ca. 1 Minute ab. Die Aufgabe des Detektionsalgorithmus besteht nun darin diese abgestellten Objekte zur erkennen und einen Alarm auszugeben, falls die Verweildauer einen Grenzwert (z.B. 30 Sekunden) überschreitet. Die Ereignisse unter Brücke N2 wurde dabei mit allen 4 Kameras (siehe Abbildung 18 für Beispiel-Aufnahmen von diesen Kameras) aufgezeichnet. Die resultierenden umfangreichen Daten (160 GB Bilddaten) bieten eine reichhaltige Basis für eine Auswertung der Detektionsalgorithmen für die abgestellten Objekte. Beispiel-Bilder für die verschiedenen Szenarien sind in Abbildung 19, Abbildung 20 und Abbildung 21 gezeigt.

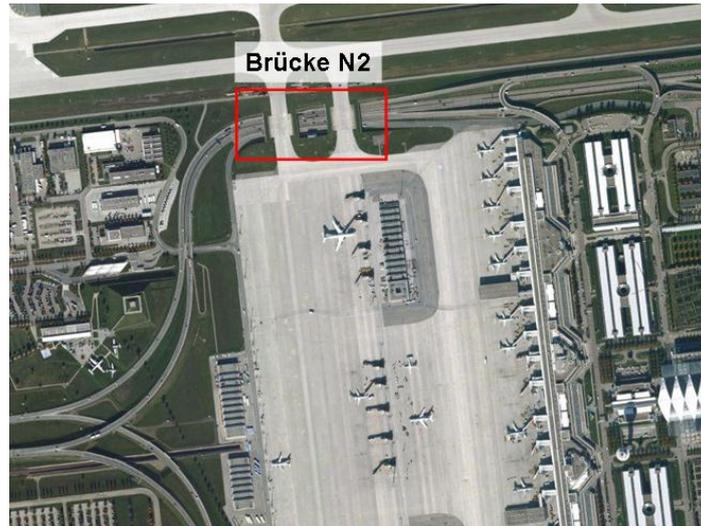


Abbildung 17: Lage der Rollbrücke, die für die Feldtest-Aufnahmen ausgewählt wurde



Abbildung 18: Beispiel-Bilder der Überwachungskameras von Rollbrücke N2.



Abbildung 19: Beispiel-Bild für das Szenario "abgestelltes Auto" (Auto in Rot markiert).



**Abbildung 20: Beispiel-Bild für das Szenario "abgestellter Koffer" (Koffer in Rot markiert).**



**Abbildung 21: Beispiel-Bild für das Szenario "angebrachter Haftmarker" (Marker in Rot markiert).**

### **Anwendung auf Daten der Feldmessungen**

Die Anwendung des entwickelten Verfahrens auf den Daten der Rollbrücken-Feldmessung scheiterte an technischen Unzulänglichkeiten der vorhandenen Kameras und

Sicherheitsbedenken bezüglich des Austauschs der installierten Kameras durch geeignete Hardware.

Die vor Ort installierten Kameras erlauben keine gezielte Ansteuerung der Belichtungszeiten, um in schneller Folge dieselbe Szene in unterschiedlicher Helligkeit abbilden zu können. Wie in zuvor beschrieben, ist dies jedoch zur HDR-Verarbeitung unverzichtbar.

Der Austausch dieser Kameras durch geeignete Hardware scheiterte jedoch an Sicherheitsbedenken. Einerseits hätte für den Zeitraum des Austauschs die Überwachung der Rollbrücken nicht mehr sichergestellt werden können, wodurch sich möglicherweise Auswirkungen auf den Flugbetrieb ergeben hätten; andererseits liegen die Rollbrücken außerhalb des Zuständigkeitsbereichs der Flughafensicherheit, sodass alle Arbeiten vor Ort mit Polizeikräften zu koordinieren sind. Es wurde in Absprache mit der FMG entschieden, dass der zusätzliche Nutzen des Austauschs der Kameras zur Evaluierung auf Feldmessungsdaten gegenüber den bereits durchgeführten Tests auf vergleichbar relevanten Brückendaten nicht gerechtfertigt gewesen wäre.

### **II.1.2. LASERBLENDUNGSSZENARIO: ERKENNUNG DYNAMISCHER ANOMALIEN**

#### **Einleitung und Szenariobeschreibung**

Es kann aufgrund geschichtlicher Betrachtungen des Potentials der menschlichen kriminellen Energie davon ausgegangen werden, dass trotz Betrachtung potentieller, noch nicht vorgefallener Anschläge oder krimineller Akte eine vollständige Abdeckung nicht möglich ist. Deshalb wurden im Rahmen des Projektes ebenfalls Methoden betrachtet, die unspezifisch auf Verhaltensänderungen sowie Abnormalitäten fokussiert sind. Solche Abnormalitäten können sich beispielsweise in optisch wahrnehmbaren dynamischen Anomalien zeigen. Diese erlauben nicht in jedem Fall eine spezifische Zuordnung von konkreten Problemlösungen oder Gegenmaßnahmen, können jedoch in Kombination mit anderen Informationsquellen und Alarmierungsmechanismen als effektives Warnsystem zur Gefahrenfrüherkennung dienen.

Die ursprünglich im Projektantrag als potentiell Anwendungsszenario genannte Auswertung von dynamischen Anomalien in Bewegungsmustern makroskopischer Personenströme („Rücklaufverhinderung“) wurde durch den Sicherheitsbeauftragten der FMG als weniger relevant eingeschätzt, da der Flughafen München kurz nach Projektbeginn bereits die Installation einer mechanischen Lösung zur Verhinderung des Rücklaufs von Personen und Gegenständen aus dem öffentlich zugänglichen in den beschränkten Passagierbereich begonnen hatte. Stattdessen musste festgestellt werden, dass seit wenigen Jahren, aber in 2009 sehr stark zunehmend, Piloten im Landeanflug mit überwiegend grünen Lasern geblendet werden. Dabei kann es auch aus mehreren Kilometern Entfernung zu so starker Blendung kommen, dass die Piloten kurzzeitig erblinden, was einen Abbruch der Landung (go around) zur Folge hat. Dieser ist mit erheblichen Kosten, Verspätungen und Risiken verbunden. Bleibende gesundheitliche Schäden der Piloten wurden bisher nicht bekannt, sind aber mit stärkeren Lasern, z.B. aus Lasershows, nicht auszuschließen. Auch wenn die wenigen aufgeklärten Fälle in Deutschland bisher von mehr oder weniger ahnungslosen Tätern ausgeführt wurden, die „einfach mal den Laser ausprobiert haben“, wird eine derartige Blendung als Gefährdung des Luftverkehrs, in anderen Ländern sogar zunächst als terroristischer Akt verfolgt. Eine Untersuchung möglicher Gegenmaßnahmen erscheint dringend geboten.

Nach Absprache mit dem Projektträger und der FMG wurde deshalb entschieden, die Zielsetzung des Arbeitsanteils anzupassen und die bei der EADS verfügbaren bildverarbeitungs-basierten Laserwarnsysteme für den Infrarotbereich hinsichtlich der Anwendung im sichtbaren Bereich und den Anforderungen des zivilen Luftverkehrs zu untersuchen. Diese Warnsysteme ermöglichen neben der Bestimmung der Laser-Wellenlänge auch eine Bestimmung der Position der Laserquelle, z.B. für Schutzmechanismen oder zur Strafverfolgung. Ziel war eine Abschätzung von Machbarkeit, Kosten und Nutzen des Einsatzes derartiger Technologien im zivilen Flugverkehr.

### **Bedrohungslage**

In Deutschland werden fast täglich Laserangriffe auf Piloten registriert. Allein 2009 nahm die Deutsche Luftsicherung pro Monat 20 bis 30 ähnlicher Fälle auf. In den USA wurden im ersten Halbjahr 2009 921 derartiger Angriffe registriert. Häufig kommen dabei einfache Laser-Pointer zum Einsatz, die günstig über das Internet zu erwerben sind. Dazu zählen beispielsweise grüne Laser-Pointer der Wellenlänge 532nm mit 700mw Leistung und Reichweiten im Bereich von 10km. Bei der Fokussierung eines derartigen Laser-Pointers im sichtbaren Bereich (VIS) auf die Netzhaut kommt es zu photochemischen Prozessen und Wärmewirkung (z.B. Ödeme ab 40°, Kollagen-denaturierung ab 80°). Die entstehenden Schäden am Auge sind im Allgemeinen irreparabel.



**Abbildung 22: Laserblendung**

### **Stand der Technik**

Aus der gegebenen Bedrohungslage heraus stellt sich die Frage nach dem Stand der Technik von Systemen zur Detektion der Bestrahlung durch Laser. Eines der wenigen kommerziell verfügbaren Systeme zur Detektion von Lasern im optischen Wellenlängenbereich ist der Laser Event Recorder der Firma Optra (s. Abbildung 23). Es handelt sich um ein tragbares System zum Zweck der Lokalisierung des Angreifers, der Beweissammlung und der Identifikation einer bestehenden Laserbedrohung. Nachteil des Systems ist die Notwendigkeit der manuellen Ausrichtung des Suchers in Richtung des Lasers. Dies scheint weder für Piloten noch für das Flugzeugpersonal ein akzeptabler Weg zu sein. Zudem fehlt die Möglichkeit der Integration des Systems in übergeordnete Systeme.



**Abbildung 23: Optra (Optra.com)**

## **Funktionen**

Ein geeignetes Laserdetektionssystem kann die Grundlage zur Realisierung folgender Funktionen im Kontext des Laserblendungsszenarios bilden:

- Warnung
- Automatisierter Schutz
- Gegenmaßnahme
- Strafverfolgung

Die Warnung des Piloten vor einer detektierten Laserbedrohung stellt die elementarste Funktion dar. Alleinstehend kann diese Funktion allerdings nur in Szenarien sinnvoll erscheinen, in denen die Bedrohungsdetektion in einem ausreichenden zeitlichen Vorlauf vor der eigentlichen Blendung des Piloten erfolgt (z.B. wenn zuerst Flügel oder Rumpf des Flugzeuges von der Laserstrahlung beaufschlagt werden). Nur so hätte der Pilot ggf. die Möglichkeiten Selbstschutzmaßnahmen (z.B. die Nutzung einer Laserschutzbrille) zu ergreifen. Ein automatisierter Schutzmechanismus (z.B. durch automatische Abdunkelung von Scheiben oder Brillen) hingegen, kann die Wahrscheinlichkeit der Blendung des Piloten durch kürzere Reaktionszeiten weiter reduzieren. Hierbei erschweren allerdings operationelle und zulassungstechnische Fragestellungen die Nutzung eines solchen Mechanismus. Die Funktionen von Gegenmaßnahmen können ausschließlich auf einen abschreckenden Effekt – beispielsweise durch gezielte Beleuchtung des Angreifers vom Flugzeug aus – bauen. Die Funktion der Strafverfolgung erfordert zumindest die Lokalisierung des Angreifers sowie ggf. die Bestimmung der verwendeten Laser-Art (Laserwellenlänge). Die entsprechenden Informationen könnten an zuständige Behörden weitergeleitet werden. Die Verwendung von Spiegelplattformen zur Bildakquisition ermöglicht bei bekannter Lokalisierung des Angreifers ggf. die schnelle detaillierte Beweiserfassung in Form von Beweisfotos.

## **Anforderungen**

Aus den oben genannten prinzipiell denkbaren Funktionen (Warnung, Schutz, Gegenmaßnahme und Strafverfolgen) ergeben sich Anforderungen an das Laserdetektionssystem. Dieses muss sich zu deren Realisierung durch folgende Eigenschaften auszeichnen:

- kurze Warnzeiten
- genaue Richtungsbestimmung

- genaue Wellenlängenbestimmung
- hohe Winkelabdeckung

Insbesondere sind hierbei kurze Warnzeiten von Relevanz, da nur dadurch ein schnelles und eventuell automatisches Einleiten von potentiellen Schutz- und Gegenmaßnahmen möglich ist. Die Bestimmung der Laserwellenlänge, sowie die genaue Richtungsbestimmung der Laserquelle (und damit die Bestimmung der Position des Angreifers am Boden) sind für Aspekte der Strafverfolgung von höchster Bedeutung. Eine weitreichende Winkelabdeckung über den gesamten Bereich des Cockpits muss zudem gewährleistet sein, um unter verschiedenen topologischen Bedingungen in den Start- und Landephase des Flugzeuges eine Absicherung – nicht nur des Cockpits – zu gewährleisten.

### Neue Ansätze zur Laserwarnung

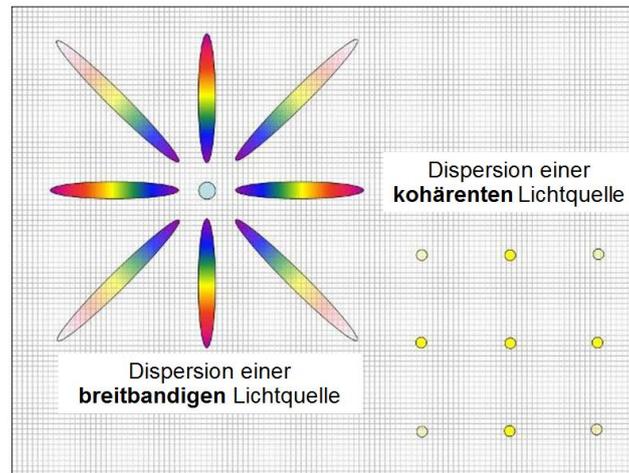
Gegenwärtig entwickelt EADS eine neue Generation von integrierten Laserwarnern zum Blendschutz. Diese zeichnen sich insbesondere durch die Eigenschaften der genauen Richtungs- und Wellenlängenbestimmung aus.

Zur Realisierung dieser Eigenschaften macht man sich die - durch die Kohärenz eines Laserstrahls bedingte - Wellenlängen-Schmalbandigkeit (Monochromasie) des Lasers im Vergleich zu herkömmlichen Lichtquellen zu Nutze (s. Abbildung 24).

Herkömmliche Lichtquellen (z.B. Glühbirne, Sonne)	Laserlicht
breites Farbspektrum	Monochromasie
unpolarisiertes Licht	lineare Polarisation (nicht generell bei allen Lasern)
Emission in alle Raumrichtungen	Gerichtetheit
schnellste Lichtänderungen > 0,1ms	Pulsdauern < 0,1ms
keine feste Phasenbeziehung der emittierten Wellenzüge	Kohärenz

**Abbildung 24: Vergleich Laserlicht**

Bildet man eine monochromatische Lichtquelle über ein Beugungsgitter (kurz Gitter) ab, so entsteht ein für das Gitter charakteristisches Beugungsmuster in Form einer Punktstruktur in äquidistanten Abständen (z.B. eine 3x3 Gitterstruktur von Lichtpunkten) im Bild. Eine breitbandige Lichtquelle hingegen erzeugt ausgewaschene Linienförmige Strukturen (s. Abbildung 25). Die Kohärenz bzw. Monochromasie des Lasers dient somit als Unterscheidungskriterium zu anderen Lichtquellen und damit der robusten Detektion des Lasers.



**Abbildung 25: Gitterdispersion**

### **Ansatz zur Richtungsbestimmung des Lasers:**

Die Richtungsbestimmung des Lasers ist in diesem Ansatz durch die algorithmische Bestimmung der Lage der sogenannten Nullten-Ordnung der charakteristischen Gitterstruktur möglich. Diese ist im Beispiel einer 3x3 Gitterstruktur der zentrale Lichtpunkt. Die Position der Nullten-Ordnung im Bild ist in diesem Ansatz in gleicher Weise als räumliche Position der Lichtquelle zu interpretieren wie die Position eines Objektes in einem herkömmlichen Bild (z.B. in einer Fotografie) dessen räumliche Position widerspiegelt.

### **Wellenlängenbestimmung des Lasers:**

Eine Wellenlängenvermessung des Lasers ist in diesem Ansatz durch die algorithmische Bestimmung der relativen Abstände der Gitterpunkte in der charakteristischen Gitterstruktur möglich. Hierbei besteht ein näherungsweise linearer Zusammenhang zwischen Gitterabstand und Wellenlänge. Ein größerer Gitterabstand deutet demnach auf eine größere Wellenlänge hin. Beides kann zueinander kalibriert werden.

### **Laborversuche**

Um den beschriebenen Ansatz der Laserdetektion mit Richtungs- und Wellenlängenbestimmung zu untersuchen, wurden bei EADS Laborversuche mit optischen Lasern und verschiedenen Gittertypen vorgenommen.

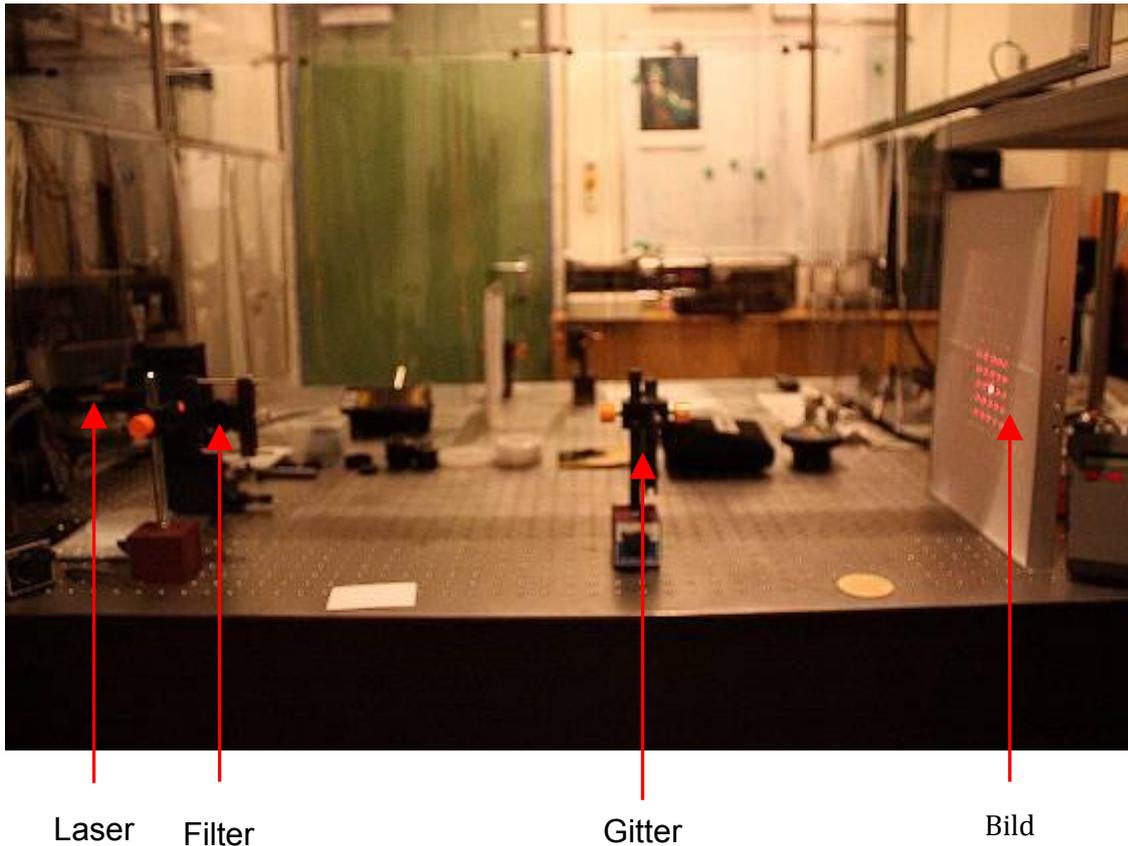


Abbildung 26: Laboraufbau

Die für die Bestimmungen der Lage und des Gitterabstandes der Lasersignatur untersuchten Algorithmen der Bildverarbeitung werden auf einem Rechner ausgeführt. Dieser ist mit einer Kamera verbunden welche mit einem Gitter (und eventuellen Wellenlängenfiltern) ausgestattet ist (s. Abbildung 27).

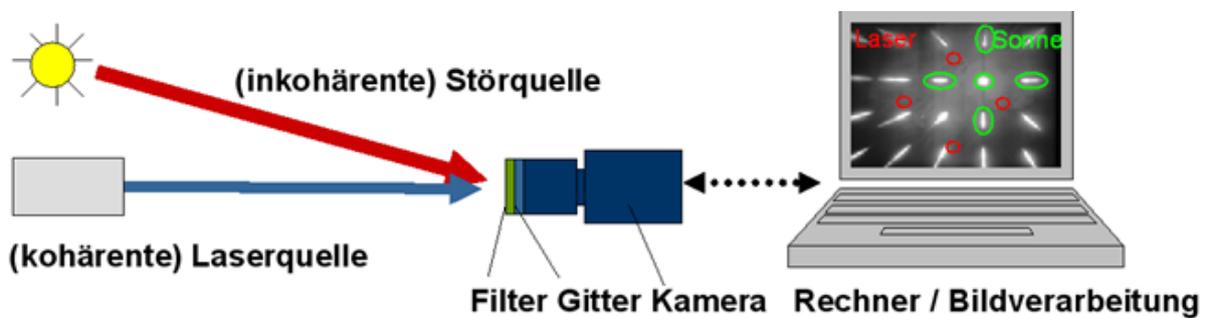


Abbildung 27: Verarbeitungskette

Die Laborversuche bestätigten dabei die erwarteten Signaturen der Laserquellen. Zudem konnte gezeigt werden, dass die punktförmige Signatur des Lasersignals algorithmisch durch Bildverarbeitungsverfahren schnell und robust von der länglichen Signatur der Sonne unterscheidbar ist (s. Abbildung 27). Dies ist die entscheidende Voraussetzung für robuste und automatisierte Detektionen der Laserbedrohungen.

### **Zusammenfassung**

Neue Ansätze zur Laserdetektion bieten die Möglichkeit sowohl die Richtung eines eintreffenden Lasers als auch seine Wellenlänge genau zu bestimmen. Die von EADS nachgewiesene automatisierte algorithmische Bestimmung dieser Größen ist grundlegende Voraussetzung zur Realisierung von nachgeordneten Funktionen wie Warnung, automatisierter Schutz, Gegenmaßnahmen oder Strafverfolgung.

Eine detailliertere Betrachtung dieser Funktionen konnte im Rahmen des Projektes nicht vorgenommen werden, da dies zum Teil zulassungsrechtliche Fragestellungen berührt. Die generelle Anwendbarkeit des beschriebenen Laserdetektionsprinzips im optischen Wellenlängenbereich – als Voraussetzung zur Realisierung weitergehender Funktionen – wurde in Laborversuchen und algorithmischen Analysen nachgewiesen.

## ***II.1.3. KOFFERSZENARIO: GEWINNUNG VON SZENEN-WISSEN ZUR INTERPRETATION UND DARSTELLUNG ERKANNTER ANOMALIEN***

### **Einleitung und Szenariobeschreibung**

Die oben beschriebenen Methodiken zur Erkennung von statischen und dynamischen Anomalien basieren auf Verfahren zur automatischen Analyse von Bildmaterial von Überwachungskameras. Erhaltene Erkennungs-Aussagen dieser Methodiken sind daher auch nur in der Bildebene interpretierbar. Dies bedeutet, dass beispielsweise die Position eines abgestellten Koffers (statische Anomalie) zwar in der Bildebene erkannt wurde (z.B. mit der Aussage „im Bild oben-links“), dessen Position im Raum aber nicht bekannt ist. Um den Rückschluss auf die Positionen des Objekts im Raum und relativ zu anderen Objekten oder Personen zu ermöglichen ist zusätzliches Szenen-Wissen, d.h. Wissen über die räumliche und kontextuelle Struktur der beobachteten Szene notwendig.

Als zur Erkennung von statischen und dynamischen Anomalien querschnittliches Thema sollte daher die Möglichkeit der automatischen algorithmischen Szenenrekonstruktion aus gegebener Überwachungssensorik untersucht und die Wechselwirkung von Szenenrekonstruktion und Objekterkennung ausgewertet werden. Das so gewonnene Szenenwissen sollte dazu mit den Methodiken zur Anomalieerkennung verknüpft werden, um damit sowohl die automatische Interpretation der Anomalien zu erleichtern, als auch deren Darstellung und Interpretierbarkeit für das Sicherheitspersonal während der Alarmierung zu verbessern.

In Abstimmung mit den Vertretern der FMG wurde ein Szenario identifiziert, in dem der Einsatz von automatisierten Auswertungsverfahren für Bilddaten von Überwachungskameras die Arbeit der Sicherheitsbeauftragten der FMG sinnvoll und effizient unterstützen könnte. Nach Aussage der FMG besteht regelmäßig die Notwendigkeit, auf Anfrage von Sicherheitsbehörden oder nach Hinweisen von Sicherheitskräften der FMG in den Überwachungskameradaten

flughafenweit nach als verdächtig identifizierten Personen und Gegenständen zu suchen. Typischerweise handelt es sich dabei um Personen mit Gepäckstücken, deren Eigenschaften (Grösse, Kleidungs- und Kofferfarbe, Form, besondere Merkmale etc.) von Zeugen oder Mitarbeitern aus der Erinnerung heraus „in etwa“ beschrieben werden können.

Als Zielsetzung wurde festgelegt, im Rahmen dieses Teilprojektes der FMG ein prototypisches System zur Verfügung stellen zu können, dass folgende Fähigkeiten besitzt:

- Erstellung eines „Phantom-Modells“ einer gesuchten Person mit Gepäckstück, ausgehend von einer Datenbank mit 3D-Modellen in „typischen“ Ausprägungen: mittels vom System automatisch erkannter Parameter soll dem Anwender die Möglichkeit gegeben werden, Eigenschaften dieser „typischen“ Basismodelle zu mischen und eine möglichst genaue 3D-Repräsentation der Suchanfrage zu erstellen;
- Training des Systems auf das Erscheinungsbild und die 3D-Geometrie der gesuchten Person mit Gepäckstück, wobei vom Anwender ausgewählte Variationen bestimmter Freiheitsgrade, wie bspw. der Farbgebung, mitberücksichtigt werden sollen;
- Suche in den Bilddaten ausgewählter Überwachungskameras nach gesuchten Personen mit Gepäckstücken, die der vom Anwender des Systems angegebenen Beschreibung entsprechen; hierbei sollen die zu erwartenden Freiheitsgrade wie Blickwinkel, Beleuchtung, Teilverdeckung etc. vom System „robust“ ausgeglichen werden können.
- Aufbereitung und Darstellung der Suchergebnisse zur weiteren Auswertung durch den Anwender.

### **Synthetische Datenerzeugung**

Entsprechend der Aufgabenstellung sollte die Nutzung von realen Trainingsdaten vermieden und durch synthetisch erzeugte Computergraphik ersetzt werden. Die hierzu notwendigen Vorarbeiten bestanden in der Erstellung der Komponenten für eine Graphikbibliothek zur Darstellung von 3D-Computergrafikmodellen mit folgendem Funktionsumfang:

- Einlesen von Geometrie, Farb-, Textur- und Materialeigenschaften von 3D-Modellen aus Dateien im offenen Collada-Standard,
- Darstellung von eingelesenen Modellen mittels der Graphik-Schnittstelle OpenGL unter Verwendung von sog. Shadern zur hardwarenahen und flexiblen Parametrisierung der Darstellungseigenschaften,
- Auswahl und Implementierung eines realitätsnahen Beleuchtungsmodells nach dem „Blinn-Phong-Modell“ unter Berücksichtigung von verschiedenen Beleuchtungsquellen unterschiedlicher Ausprägung (diffus, spekulär, mehrere Lichtquellen),
- Auswahl und Implementierung eines realitätsnahen Schattenwurfmodells nach dem „Soft Shadow“-Prinzip unter Verwendung von durch die Hardware erstellten Tiefendaten und einem „Random Sampling“ zur Simulation von Lichtstreuungseffekten.

Die folgende Abbildung zeigt ein Beispiel des Rendervorgangs.



**Abbildung 28: Synthetische Bilddaten**

### **Feldmessungen Kofferszenario**

Zur Evaluierung des Verfahrens sind reale Aufnahmen aus dem Innenbereich des Flughafens München notwendig, die Personen mit definierten Gepäckstücken enthalten und entsprechend mit Grundwahrheitsannotationen versehen werden müssen. Die aufgezeichneten Bilddaten sollen so nah wie möglich das geplante Anwendungsszenario widerspiegeln, das heißt verschiedene Personen und Gepäckstücke sollen automatisch in den Bildern verschiedener Überwachungskameras detektiert werden. Dazu wurden für die Aufnahmen im Feldtest 6 bekannte Gepäckstücke aus verschiedenen Gepäckstückklassen (Rucksack, Aktenkoffer, Rollkoffer) ausgewählt. In der folgenden Abbildung sind die verwendeten Gepäckstücke zu sehen.



**Abbildung 29: Verwendete Gepäckstücke für die Feldmessungen zur "Wiedererkennung von Gepäckstücken".**

Mit Hilfe dieser Gepäckstücke wurden zahlreiche reale Szenarien aufgezeichnet. Insgesamt umfasst das aufgezeichnete Datenmaterial 110 GB, was eine reichhaltige Datenbasis für eine Bewertung der entwickelten Detektionsalgorithmen darstellt. Nach intensiver Abstimmung mit dem Flughafen München wurden 3 verschiedene Überwachungskameras aus Terminal 1 und Terminal 2 ausgewählt:

- Eine Kamera am Eingang zu Terminal 2
- Eine Kamera am einem Laufband in Terminal 1
- Eine Kamera am einem Rolltreppenaufgang in Terminal 2

Die Positionen der Kameras wurden so ausgewählt, dass sie entlang möglicher Wegrouen von Passagieren liegen (z.B. Passagier verlässt Terminal 2 und begibt sich in Terminal 1). Die Lage der Kameras als auch Beispielbilder ist in folgender Abbildung zu sehen. Die folgenden Beispielbilder der Kameras zeigen, dass die Umgebung der einzelnen Wegpunkte variiert stark (z.B. unterschiedliche Bodenmusterung und Anzahl von Personen im Bild), wodurch eine aussagekräftige Evaluierung auf Grund dieser Bilddaten möglich ist.

Die Bilddaten aus den Feldmessungen wurden anschliessend zur Vorbereitung der Evaluierung manuell annotiert. Hierbei wurde die Position und Klasse von Personen und Gepäckstücken in den Einzelbildern der aufgenommenen Videos markiert.

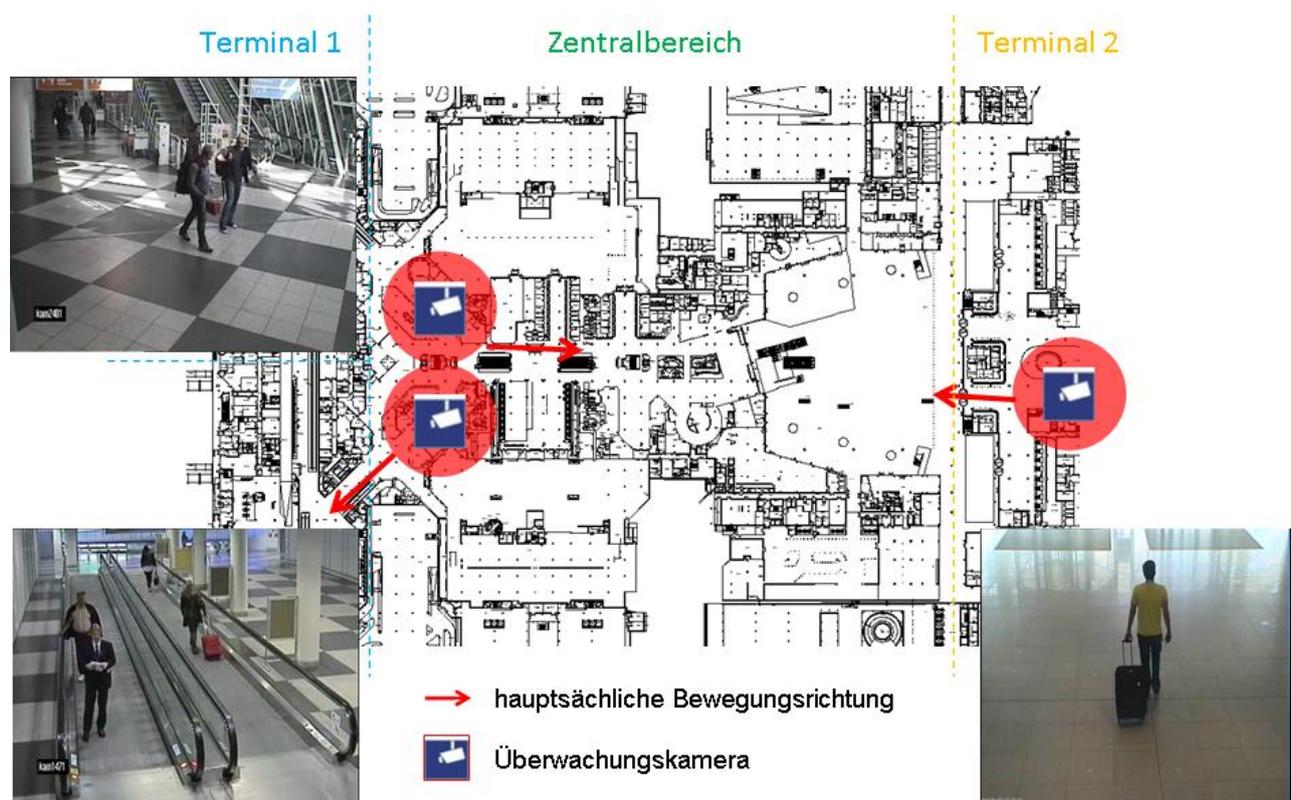


Abbildung 30: Positionen der Kameras für die Feldmessungen zur "Wiedererkennung von Gepäckstücken".

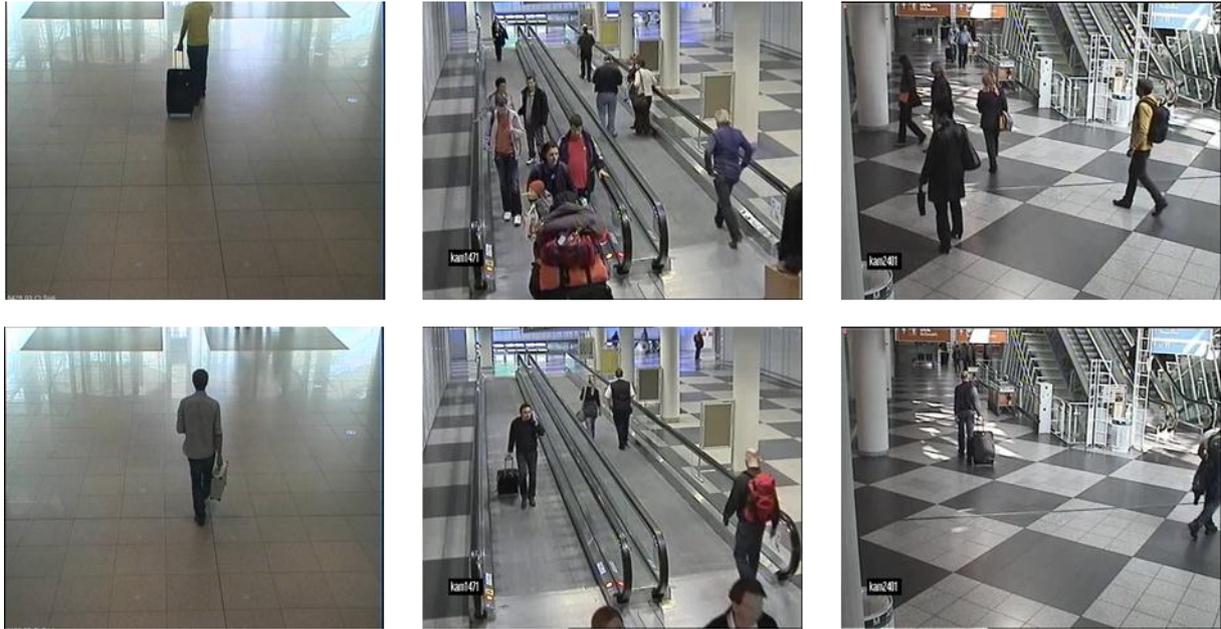


Abbildung 31: Beispiel-Bilder für das Szenario "Wiedererkennung von Gepäckstücken".

### Theoretische Vorarbeiten zur Objektklassenerkennung

Die im Rahmen von SiVe zu lösende Aufgabe der Erkennung von Personen und Gepäckstücken kann als Spezialisierung der allgemeinen Objektdetektion verstanden werden. Moderne Ansätze zur Objekterkennung bestehen typischerweise aus den folgenden Komponenten:

- Aufteilung und Transformation der Bilder in eine robuste hochdimensionale Datenrepräsentation;
- Auswahl einer geeigneten aufgabentypischen Untermenge der zur Verfügung stehenden Daten als Trainingsmaterial;
- Anwendung von statistischen Lernverfahren auf der Trainingsmenge zum Zweck der Identifikation statistisch relevanter Gemeinsamkeiten innerhalb der zu erkennenden Objektklasse;
- Algorithmische Umsetzung der durch das Lernverfahren gewonnenen Relevanzkriterien zur schnellen Relevanzbewertung von neuen Testdaten;

Die folgende Abbildung illustriert die oben beschriebenen Schritte.



Abbildung 32: Komponenten der Objekterkennung

Durch EADS IW wurde für SiVe ein neuartiges Konzept zur Objekterkennung entwickelt, welches für die Anforderungen in besonderer Weise geeignet ist. Die Arbeit wurde in der folgenden Veröffentlichung ausführlich beschrieben:

*Liebelt et al., „Multi-View Object Class Detection With A 3D Geometric Model“, IEEE International Conference for Computer Vision and Pattern Recognition 2010.*

Im besonderen Unterschied zu anderen Arbeiten zur Objekterkennung, die auf die Verfügbarkeit realer Trainingsbilddatenbanken angewiesen sind, erlaubt das von EADS IW entwickelte Verfahren, Trainingsdaten unter Verwendung von Methoden der Computergraphik aus CAD-Modellen synthetisch zu erzeugen. Die Grundlagen hierzu wurden im Vorfeld von SiVe durch EADS IW im Rahmen der folgenden Publikation erarbeitet:

*Liebelt et al., „Viewpoint-Independent Object Class Detection using 3D Feature Maps“, IEEE International Conference for Computer Vision and Pattern Recognition 2008.*

Die Vorteile dieses Ansatzes liegen in der insbesondere für Sicherheitsanwendungen relevanten Möglichkeit, aufgabentypische Objekt- und Kontextkonfigurationen in beliebiger Menge zu erzeugen, ohne auf langwierige Sammlung und Annotation realer Videodatenbanken angewiesen zu sein. Langfristig kann ein solches Vorgehen geeignet sein, um dem Sicherheitspersonal am Flughafen die Möglichkeit zu geben, das Detektionsverfahren schnell auf das Auffinden neuer Gepäckstücke hin auszulegen, für die zum Zeitpunkt der Suchanfrage noch keine realen Bilder vorhanden sind. Die nachfolgende Abbildung zeigt Beispiele einiger CAD-Gepäckstückmodelle.



**Abbildung 33: CAD-Gepäckstückmodelle**

In den Zwischenberichten 2010 und 2011 wurden die theoretischen Vorüberlegungen und praktischen Implementierungs- und Parallelisierungsansätze ausführlich beschrieben, die im Rahmen dieses Abschlussberichts nicht erneut aufgegriffen werden sollen. Stattdessen wird im folgenden Abschnitt das Gesamtsystem beschrieben, welches die Aufgabe des Kofferszenarios löst.

### **Selbstkalibrierender 3D-Kontext zur Suche nach Personen mit Gepäckstücken in Bildsequenzen**

Im Folgenden wird die Suche nach Bildern eines Überwachungskameranetzwerk mittels attributsbasierter Suchanfragen beschrieben. Der Ansatz basiert auf Detektoren für verschiedene Objektklassen, die es ermöglichen, nach Kombinationen von Personen und charakteristischen Gepäckstücken zu suchen. Der Ansatz ist unabhängig von den Bildraten der

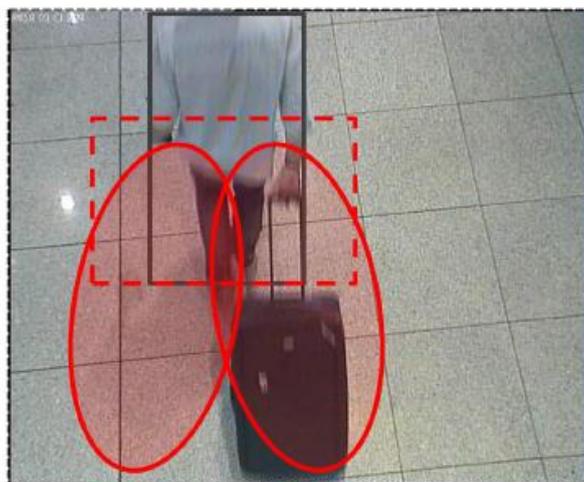
Kameras, da er weder auf Objektverfolgung noch auf Annahmen bezüglich des Bildhintergrunds basiert. Zudem wird der Ansatz vollständig mit synthetisch erzeugten Bilddaten trainiert und erfordert folglich weder reale Trainingsbildaufzeichnungen noch manuelle Bildannotationen. Eine näherungsweise Selbstkalibration jeder Kamera anhand einiger im Bild detektierter Personen erlaubt es, Kontextbeziehungen zwischen Objekten in einem 3D-Koordinatensystem auszunutzen und dadurch die Leistungsfähigkeit der einzelnen Objektdetektoren, die auf unauffälligen Objektklassen häufig hohe Fehlerraten erzeugen, erheblich zu verbessern. Der Ansatz wurde auf Testdaten verschiedener Überwachungskameras des Flughafens München evaluiert und konnte die Fähigkeit des Systems nachweisen, Suchanfragen mit Attributen wie Erscheinungsbild und metrischem Kontext über verschiedene Kameras hinweg erfolgreich zu beantworten.



**Abbildung 34: Ohne örtlichen Kontext sind die beiden ausgewählten Bildbereiche schwer zu interpretieren; dies ist ein typisches Problem bei der Detektion wenig charakteristischer Objektklassen.**

Die bildbasierte Suche nach Personen in umfangreichen Aufzeichnungen von Überwachungskameranetzwerken stellt insbesondere dann eine Herausforderung dar, wenn Zeugenaussagen nur ungenaue Beschreibungen liefern. Es ist überaus schwierig, Personen anhand von näherungsweise Angaben wie „mittelgross, trägt ein dunkles Hemd“ in Bildsequenzen wiederzufinden, die überfüllte Innenräume mit vergleichsweise niedrigen Videoauflösungen abbilden. Folglich ist jede zusätzliche Angabe wichtig, um die Suchergebnisse weiter zu verfeinern; hierzu zählen beispielsweise Beschreibungen der Gepäckstücke, die eine gesuchte Person mit sich trägt (z.B. „trägt einen roten Rucksack“). Andererseits liefern unabhängig trainierte Objektdetektoren für derartige Objektklassen bislang noch keine besonders guten Ergebnisse, wie in verschiedenen Testreihen, z.B. VOC2010 [7], nachgewiesen wurde. Dieses Verhalten ist auch nicht überraschend, wenn man beispielsweise die wenigen Bildmerkmale betrachtet, die sich in Abbildung 34 für einen Rucksack ergeben, ohne dabei den gesamten Szenenkontext zu berücksichtigen. Folgerichtig wurde die Verwendung von Szenenkontext in verschiedenen Veröffentlichungen vorgeschlagen, um das gleichzeitige

Auftreten von Objektklassen auszunutzen, welche sich unabhängig voneinander nur schwer detektieren lassen [6]. Andererseits ist die Anwendbarkeit derartiger Kontextverfahren in realen Szenarien häufig eingeschränkt, da sich kleine Objekte schlecht detektieren lassen und der reine 2D-Bildkontext für Szenarien mit starken perspektivischen Veränderungen nicht geeignet ist (siehe Abbildung 35). Höherwertige 3D-Kontextverfahren erfordern in der Regel Vorannahmen über die Geometrie der aufgenommenen Szene [11] oder eine manuelle Kalibration, was für große Überwachungskameranetzwerke nicht praktikabel ist. Zudem eignen sich die relativ niedrigen Aufzeichnungsraten derartiger Anwendungsfälle nicht für den Einsatz von Verfolgungsmethoden zur Schätzung von Fluchtpunkten anhand von Bewegungstrajektorien [14], und die manuelle Annotation von Trainingsdaten unter Berücksichtigung aller möglichen Kontextangaben zwischen verschiedenen Objektklassen ist ebenfalls nicht praktisch umsetzbar.



**Abbildung 35: Örtlicher Kontext zwischen einer Person und einem Trolley: die 2D-Detektion der Person (durchgezogene graue Linie) erlaubt die Bestimmung eines Bildbereichs in 2D, in dem die Suche nach einem Trolley aussichtsreich ist (gestrichelte rote Linie). Allerdings berücksichtigt dies nicht die perspektivischen Veränderungen des Bildes, die durch die Kamera induziert werden können (hier beispielsweise eine starke Kameraneigung). Im Gegensatz dazu beschreibt der projizierte 3D-Kontext den Suchbereich sehr viel genauer (durchgezogene rote Linie).**

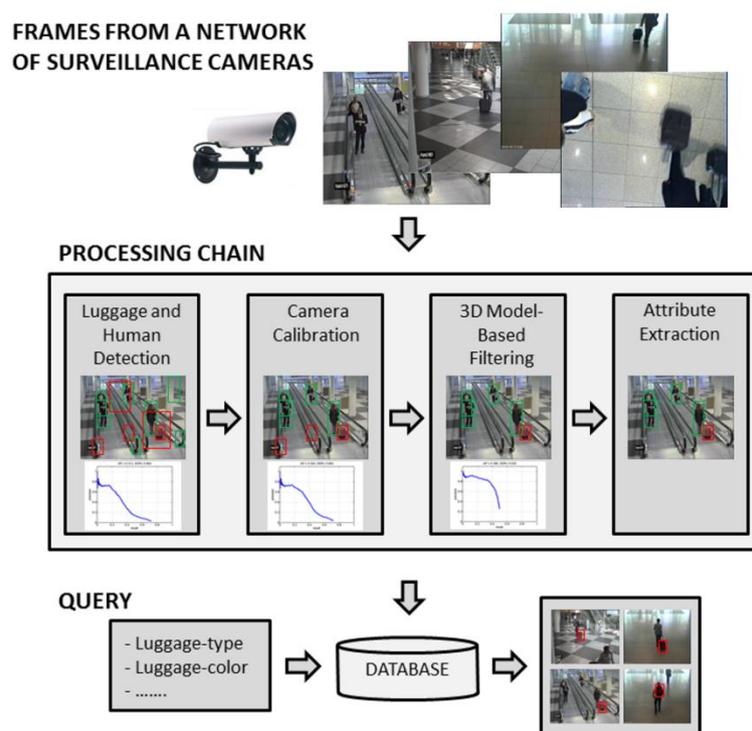
In dieser Arbeit beschreiben wir einen Ansatz, der ausschließlich auf synthetisch erzeugten Bildern trainiert wird und dadurch vom realen Einsatzrahmen unabhängig ist und universeller verwendbar wird. Er verzichtet auf die Verwendung von Objektverfolgungsmethoden und schätzt die näherungsweise 3D-Szenengeometrie für jede Kamera anhand einiger detektierter Menschen, ohne dass ein manueller Kalibrationsschritt notwendig ist. In der Folge kann diese 3D-Szenengeometrie genutzt werden, um die Auftretensabhängigkeiten zwischen Personen und Gepäckstücken in einem 3D-Koordinatensystem zu modellieren und dadurch die Detektionsgenauigkeiten von individuell schwierig zu detektierenden Gepäckstückklassen wie Rucksäcken und Koffern zu verbessern. Zudem können nun Suchanfragen verschiedene Angaben zu Objektklassen, Erscheinungsbild und metrischen Größen kombinieren (z.B. „eine große Person, die ein dunkles Hemd und einen roten Rucksack trägt und einen Trolley zieht“). Der beschriebene Ansatz wurde auf einem realistischen Datensatz aus ca. 6000 Einzelbildern und kurzen Bildfolgen verschiedener Überwachungskameras des Flughafens München evaluiert und konnte gegenüber herkömmlichen Ansätzen mit individuellen Detektoren für jede

Objektklasse eine erhebliche Verbesserung erzielen. Nachfolgend geben wir zunächst eine kurze Übersicht über bestehende Arbeiten zu den Themen der attributbasierten Bildsuche, der kontextabhängigen Objekterkennung und der Kamerakalibration. Ein Überblick des Ansatzes wird in Absatz „Überblick“ gegeben, eine detaillierte Beschreibung der Verarbeitungskette ist in Abschnitt „Verarbeitungskette“ zu finden. Die Versuchsergebnisse werden im Abschnitt „Experimentelle Untersuchung“ zusammengefasst.

## **Verwandte Arbeiten**

Diese Arbeit befasst sich mit einem Ansatz zur attributbasierten Suche nach einzelnen Bildern von Überwachungskameras. Verschiedene verwandte Arbeiten behandeln dieses Thema ebenfalls: [9] beschreiben Verfolgung und Bewegungsklassifikation, um attributbasiert Fahrzeuge in kalibrierten Kamerabildern zu suchen. [10] verwenden einen mehrschichtigen Adaboost-Klassifikator zur Suche nach Fahrzeugen, der teilweise auf synthetischen Daten trainiert wird. [20,23] beschreiben Methoden, die für einzelne Körperteile Klassifikatoren trainieren, um Menschen durch detaillierte Suchanfragen finden zu können. Obwohl viele andere Arbeiten nicht direkt auf die Bildsuche abzielen, können ihre Grundansätze gegebenenfalls dazu beitragen. So beschreiben [2] die Verfolgung von Fussgängern und die robuste Verbindung ihrer Trajektorien; [22] nutzen bewegungssegmentierte Silhouetten und [5] nutzen Silhouetten im zeitlichen Verlauf, um zu bestimmen, ob eine Person ein Objekt trägt. Im Gegensatz dazu erfordert der hier vorgestellte Ansatz keine konsekutiven Bildfolgen. Statt der Verfolgung oder Bewegungssegmentierung kommt er ausschliesslich mit Klassifikatoren für Menschen und Gepäckstücke in Einzelbildern aus und ist dadurch unabhängig von Aufzeichnungsraten, Hintergrund- und Bewegungsmodellen. In den letzten Jahren sind durch latente teilbasierte Objektdetektoren [8] in Zusammenspiel mit Gradientenhistogrammen als Bildbeschreibungen [4] beeindruckende Detektionsergebnisse für viele Objektklassen erzielt worden. In [13,19] konnte weiterhin gezeigt werden, dass durch den Einsatz von synthetischen Trainingsdaten die Abhängigkeit von realen Trainingsdatenbanken verringert und eine systematische Veränderung von Blickwinkeln und Bildaufnahmebedingungen während des Trainings erreicht werden kann. In Bezug auf die Detektion von Menschen konnten mit synthetischen Daten zu realen Daten vergleichbare Ergebnisse erzielt werden [15,17]. In der vorliegenden Arbeit bauen wir entsprechend auf diesen Ergebnissen auf und trainieren den Ansatz aus [8] ausschliesslich auf synthetischen Menschenmodellen. Für andere, weniger charakteristisch texturierte oder geformte Objektklassen sind die dadurch erzielbaren Ergebnisse jedoch deutlich schlechter [7]. Andererseits sind für Szenarien aus dem Bereich der Überwachung gerade solche weniger charakteristischen Klassen wie beispielsweise Koffer von besonderer Bedeutung. Um die Erkennungsraten in diesen Fällen zu verbessern, greift die vorliegende Arbeit auf Kontextbeziehungen im 3D-Ortsraum zwischen Detektoren charakteristischer Klassen, wie beispielsweise Personen, und Detektoren weniger charakteristischer Klassen, wie beispielsweise Gepäckstücken, zurück. Die Verwendung von Kontextbeziehungen zwischen verschiedenen Objektklassen und auch zwischen Szene und Objektklassen ist bereits in der Vergangenheit vorgeschlagen worden; eine Übersicht wird in [6] gegeben. [21] verwenden Kontext im 3D-Ortsraum für kalibrierte Szenen, um Menschendetektion und Menschenverfolgung zu verbessern, indem die Detektoren an die geometrischen Anforderungen der Szenenumgebung angepasst werden. Im Gegensatz dazu erfordert der vorliegende Ansatz keine vorausgehende Kalibration und ist folglich ähnlicher zu [11], die ein 3D-Kameramodell schätzen, um geometrisch unwahrscheinliche Detektionen zu filtern, oder [1,3,24] die die partielle 3D-Geometrie der Szene bestimmen und damit die zu

erwartende Sichtbarkeit von Objekten in den Detektionsprozess einfließen lassen. Andererseits benötigen [11,18] Vorwissen über die Szenengeometrie und viele klassische Ansätze zur Selbstkalibration aus Objektdetektionen beruhen auf Objektverfolgungsmethoden (z.B. [12]), der Berechnung von Fluchtlinien (z.B. [14]) oder nehmen identische Objektgrößen an (z.B. [16]). Im Unterschied dazu schätzt unser Ansatz extrinsische und intrinsische Kameraparameter mit einem wahrscheinlichkeitstheoretischen Verfahren derart, dass die Verteilung der Größen der detektierten Menschen mit der als bekannt angenommenen realen menschlichen Größenverteilung übereinstimmt. Obwohl dieses Verfahren weniger präzise ist, bietet es dennoch einen hinreichenden Rahmen zur Auswertung von örtlichen 3D-Kontextbeziehungen zwischen Menschen und verschiedenen Gepäckstücken.



**Abbildung 36: Objektdetektoren, näherungsweise Kameraselbstkalibration und 3D-Objektkontext ermöglichen die Filterung von Einzelbildern verschiedener Überwachungskameras ohne zeitliche Konsistenzanforderungen. In der Folge können komplexe attributsbasierte Suchanfragen beantwortet werden, die Objektklassen, Erscheinungsbild und metrische Angaben enthalten können.**

## Überblick

Abbildung 36 gibt eine Übersicht über das in der vorliegenden Arbeit beschriebene Verfahren. Es wird ein realistisches Szenario angenommen, in dem Einzelbilder aus verschiedenen unkalibrierten Pan-Tilt-Zoom-Überwachungskameras in einer Datenbank gespeichert werden, um zu forensischen Zwecken nach Personen mit bestimmten Attributen wie Erscheinungsbild, Größen und Gepäck durchsucht werden zu können. Um Netzwerkauslastung und Speicherbedarf zu reduzieren, werden nur Einzelbilder aufgenommen, weswegen Verfolgungsverfahren anhand zeitlicher Kohärenz nicht anwendbar sind. Jedes Einzelbild wird markiert, um anzugeben, ob der Kameraoperator die Kamera seit dem letzten aufgenommenen Bild bewegt hat. Der vorliegende Ansatz arbeitet wie folgt:

- Die Einzelbilder jeder Kamera des Netzwerkes werden in Untermengen aufgeteilt, die keine zwischenzeitliche Bewegung der Kamera aufweisen. Objektklassendetektoren für Personen und verschiedene Gepäckstückarten werden auf jedes der Einzelbilder jeder Untermenge angewandt; siehe Abschnitt „Objektklassendetektion“. Falls wenigstens eine Person in der Untermenge detektiert wurde, wird der nächste Prozessschritt ausgeführt.
- Eine näherungsweise Selbstkalibration der intrinsischen und extrinsischen Kameraparameter der aktuellen Untermenge wird ausgeführt; siehe Abschnitt „3D Selbstkalibration“. Diese Kameraparameter sind innerhalb derselben Untermenge konstant. Eine Maximum-Likelihood-Schätzung bestimmt dabei die Parameter derart, dass die Verteilung der Größen der detektierten Personen nach erfolgter 3D-Rekonstruktion mit einer bekannten realen Größenverteilung von Menschen möglichst gut übereinstimmt. Diese Schätzung ermöglicht die Filterung von Objekterkennungen von Personen und Gepäckstücken, die zu der gemeinsamen 3D-Szenengeometrie unplausibel sind. (siehe Abbildung 40, Mitte).
- Durch Kombination der rekonstruierten 3D-Szenengeometrie mit einem statischen 3D-Kontextmodell für Personen, die verschiedene Arten von Gepäckstücken tragen (siehe Abbildung 39) kann nun die Genauigkeit der Gepäckdetektionen erheblich verbessert werden.
- Metadaten aus den einzelnen Verarbeitungsschritten werden nun in einer Datenbank gespeichert, um als Attribute effizient durchsucht werden zu können; siehe Abschnitt „Attributsextraktion“ und die Suchergebnisse in Abschnitt „Experimentelle Untersuchung“ und Abbildung 43, Abbildung 44, Abbildung 45.



**Abbildung 37: Teile-basierte Objektklassendetektoren werden auf synthetischen Daten trainiert.**

## **Verarbeitungskette**

In diesem Abschnitt wird die Verarbeitungskette des Ansatzes erläutert.

### *Objektklassendetektion*

Die vorliegende Arbeit baut auf dem aktuellen Erkennungsansatz von [8] auf, der diskriminative deformierbare Teilemodelle für Gradientenhistogramme auf mehreren Skalen lernt. In [8] wird

ein Teilemodell eines Objektes in eine globale Beschreibung und mehrere latente Teilebeschreibungen unterteilt, die das Objekt als eine flexible Konstellation verschiedener Komponenten darstellen. Anstatt reale Trainingsdaten zu nutzen, werden in der vorliegenden Arbeit die Ideen von [13,19] aufgegriffen, indem Objektklassendetektoren mit synthetischen Computergrafikmodellen trainiert werden, welche keinerlei manuelle Annotationen erfordern und es ermöglichen, Trainingsdaten in großem Umfang zu erstellen. Abbildung 37 zeigt einige Trainingsbeispiele für die einzelnen Objektklassen; siehe auch Abschnitt „Trainingsdaten“ für Details zur Erstellung dieser Trainingsbeispiele.

### 3D Selbstkalibration

Ausgehend von einer Anzahl von Einzelbildern einer Kamera ohne zwischenzeitliche Bewegung liefert ein 3D Selbstkalibrationsschritt den Rahmen zur Interpretation des 3D-Kontextes zwischen Objekten. Dieser Kalibrationsschritt basiert auf einem Bayes-Ansatz, um die Verteilung der Größen der detektierten Personen nach erfolgter 3D-Rekonstruktion mit einer bekannten realen Größenverteilung von Menschen möglichst gut übereinstimmen zu lassen; es handelt sich dabei um eine Art Verallgemeinerung der Idee in [11]. Dabei wird ein vereinfachtes Kameramodell angenommen, um von einer im Bild detektierten Person auf deren Größe in der realen Szene zu schliessen. Die intrinsischen und extrinsischen Parameter dieses Kameramodells werden mit einem Maximum-Likelihood-Verfahren bestimmt. Nachfolgend werden die verwendeten Variablen eingeführt: Weltkoordinaten eines Objektes  $X_w = (X_w, Y_w, Z_w)$ , dessen Kamerakoordinaten  $X_c = (X_c, Y_c, Z_c)$ , Höhe der Kamera im Weltkoordinatensystem  $Y_w$ ; Bildkoordinaten  $u = (u, v)$ , Kameraneigung  $\theta$ , Brennweite  $f$ , optischer Aufpunkt der Kamera  $(u_c, v_c)$ .

#### Kameramodell

Es wird ein vereinfachtes Kameramodell angenommen, das keine Drehung der Kamera um die Blickachse zulässt und einen in der Bildebene zentrierten Aufpunkt aufweist. Die Bodenebene liegt in  $Y_w = 0$ , wobei alle detektierten Personen im aktuellen Einzelbildsatz angenommene Fusspunkte auf dieser Bodenebene besitzen. Der Ansatz ist unabhängig vom Kameragierwinkel und könnte auch mit einem sich innerhalb des Einzelbildsatzes verändernden Gierwinkel funktionieren, solange die Bodenebenenannahme erfüllt bleibt. Es wird ein perspektivisch projektives Modell mit einem Bildseitenverhältnis von 1 angenommen. In homogenen Koordinaten hat die Transformation von Welt- zu Kamerakoordinaten die Form

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos \theta & \sin \theta & -Y_w^c \cos \theta \\ 0 & -\sin \theta & \cos \theta & Y_w^c \sin \theta \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix}. \quad (1)$$

Die Transformation von Kamera- in Bildkoordinaten, welche die perspektivische Projektion  $\Phi$  festlegt, hat die Form

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \frac{1}{Z_c} \begin{bmatrix} f & 0 & u_c & 0 \\ 0 & f & v_c & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix}. \quad (2)$$

Aus den beiden Gleichungen folgt für  $Y_w$

$$Y_w = \frac{Z_w(v \cos \theta - v_c \cos \theta - f \sin \theta)}{f \cos \theta + v \sin \theta - v_c \sin \theta} + \frac{Y_w^c(v \sin \theta - v_c \sin \theta + f \cos \theta)}{f \cos \theta + v \sin \theta - v_c \sin \theta}. \quad (3)$$

Unter der Annahme, dass die Fußpunkte aller detektierten Menschen auf der gemeinsamen Bodenebene liegen, kann man aus dem obersten (Kopf-) und dem untersten (Fuß-) Bildpunkt einer Menschendetektion  $v_t$  und  $v_b$  den Abstand zur Kamera  $Z_w$  in  $v_b$  ableiten:

$$Z_w = \frac{-Y_w^c(v_b \sin \theta - v_c \sin \theta + f \cos \theta)}{v_b \cos \theta - v_c \cos \theta - f \sin \theta}. \quad (4)$$

Damit folgt für die Größe  $Y_w^h$  des Menschen in der Szene für gegebene Fuß- und Kopfpunkte der Detektion im Bild

$$Y_w^h = \frac{\frac{-Y_w^c(v_b \sin \theta - v_c \sin \theta + f \cos \theta)}{v_b \cos \theta - v_c \cos \theta - f \sin \theta} (v_t \cos \theta - v_c \cos \theta - f \sin \theta)}{f \cos \theta + v_t \sin \theta - v_c \sin \theta} + \frac{Y_w^c(v_t \sin \theta - v_c \sin \theta + f \cos \theta)}{f \cos \theta + v_t \sin \theta - v_c \sin \theta}. \quad (5)$$

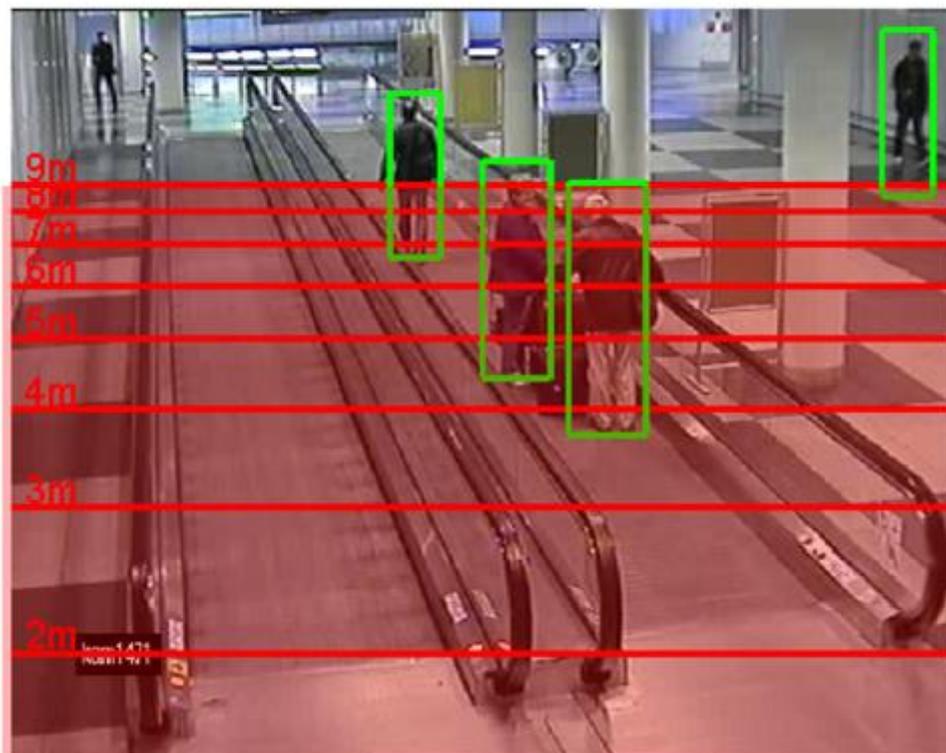


Abbildung 38: Darstellung der geschätzten Bodenebene für eine Kamera.

### Maximum-Likelihood-Schätzung

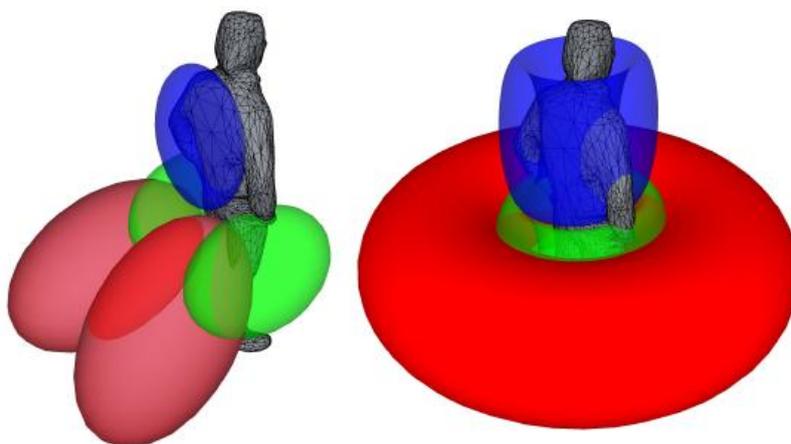
Unter der Annahme, dass der Aufpunkt der Kamera im Bildzentrum liegt, verbleiben nun drei freie Kameraparameter  $\alpha$ : Kameraneigung  $\theta$ , Höhe der Kamera im Weltkoordinatensystem  $Y_{wc}$ , und Brennweite  $f$ . Der Kamerawinkel muss für die gegebene Aufgabe nicht geschätzt werden. Ausgehend von  $N$  verlässlichen, d.h. hochbewerteten Menschendetektionen  $\omega = [(v_t, v_b)_1, \dots, (v_t, v_b)_N]$  innerhalb eines Einzelbildsatzes einer Kamera lassen sich diese drei Kameraparameter bestimmen aus

$$\hat{\alpha} = \underset{\alpha}{\operatorname{argmax}} p(\omega|\alpha). \quad (6)$$

Zur Lösung wird weiter vorausgesetzt, dass die Grössen der Menschen einer Gaussverteilung  $N$  mit bekanntem Mittelwert  $\mu$  und Varianz  $\sigma$  folgen. Logarithmieren der obigen Gleichung liefert die optimalen Kameraparameter  $\alpha$  so, dass die Verteilung der rekonstruierten Menschengrössen  $Y_{w_i}^h$  der vorangenen Verteilung der realen Grössen  $N$  übereinstimmt, d.h.

$$\hat{\alpha} = \underset{\alpha}{\operatorname{argmax}} \left[ \frac{N}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_i^N (Y_{w_i}^h(\alpha) - \mu)^2 \right]. \quad (7)$$

Dies wiederum erlaubt die Berechnung metrischer Grössen für jede weitere Objektdetektion. Obwohl diese Schätzungen näherungsweise sind und stark vom Vorhandensein einer hinreichenden Anzahl genauer Menschendetektionen und der Gültigkeit der vorangenen Grössenverteilung abhängen, erlauben sie es dennoch, Detektionsfehler herauszufiltern, die zu der geschätzten 3D-Szene nicht konsistent sind. Zudem liefern sie den Rahmen für den im Folgenden beschriebenen 3D-Kontext zwischen Objekten. Zudem können die metrischen Grössenangaben für die Suchanfragen verwendet werden. Abbildung 38 zeigt eine Visualisierung der rekonstruierten Bodenebene einer Kamera.



**Abbildung 39: 3D Ortskontext für Gepäckstücke, die von einer Person getragen werden können. Falls der Menschendetektor Poseinformation zur Verfügung stellt, kann die Auftretenswahrscheinlichkeit der verschiedenen Gepäckstückarten (rot: Trolley, grün: Koffer, blau: Rucksack) als 3D Gaussellipsoide modelliert werden (links), deren 2D-Projektionen sich einfach berechnen lassen. Im vereinfachten Fall**

**von 2D-Menschen-detectionen ohne Poseangaben werden die Ellipsoide zu rotationsinvarianten Tori (rechts).**

### *3D-Kontext zwischen Objekten*

Ziel ist die Modellierung des örtlichen Kontextes in einem 3D-Koordinatensystem zwischen Menschen und verschiedenen Gepäckstücken. Da die Möglichkeiten der Interaktion zwischen Menschen und Gepäckstücken begrenzt sind und die Verfügbarkeit von 3D-Szeneninformationen aus dem Rekonstruktionsschritt es erlaubt, diese Abhängigkeiten in 3D zu beschreiben, wird im Folgenden ein statisches 3D-Kontextmodell vorgeschlagen (siehe Abbildung 39, links). Unter der Annahme, dass der Detektionsschritt eine Schätzung der Menschengeschwindigkeit liefert, kann die Auftretenswahrscheinlichkeit jeder Art von Gepäckstück als eine Gaußverteilung in 3D mit Mittelwert  $\mu_{3D}$  und Kovarianz  $C_{3D}$  beschrieben werden, sodass sich für ein Gepäckstück an der Position  $X_w$  die Auftretenswahrscheinlichkeit wie folgt berechnet:

$$p(\mathbf{X}_w | v, \mu_{(3D)}, C_{(3D)}) = \mathcal{N}(\mathbf{X}_w | P_v \mu_{(3D)}, R_v C_{(3D)}) \quad (8)$$

Hierbei entspricht für eine Menschenpose  $v$  die Transformation  $P_v = [R_v | t_v]$  mit Rotation  $R_v$  und Translation  $t_v$ . Aus dem Kalibrierungsschritt ergibt sich mit der Projektion  $\Phi_\alpha$  für gegebene Kameraparameter  $\alpha = [\theta, f, Y_{cw}]$  die Auftretenswahrscheinlichkeit an einer Position  $u$  im Bild als

$$p(u | v, \mu_{(3D)}, C_{(3D)}, \alpha) = \mathcal{N}(u | \Phi_\alpha(P_v \mu_{(3D)}), \Phi_\alpha(R_v C_{(3D)})) \quad (9)$$

Um die Berechnung der projizierten Kovarianz zu vereinfachen, kann die Projektion  $\Phi_\alpha$  unter Annahme einer lokalen Affinität mit einer Taylorerweiterung an  $P_v \mu_{(3D)}$  angenähert werden:

$$\Phi_\alpha(R_v C_{(3D)}) \approx J_{\Phi_\alpha}(P_v \mu_{(3D)}) \cdot R_v C_{(3D)} \cdot J_{\Phi_\alpha}^t(P_v \mu_{(3D)}) \quad (10)$$

Hierbei ist  $J_{\Phi_\alpha}$  die Jacobimatrix der Projektion  $\Phi_\alpha$ . Falls die Menschen-detection keine Richtungsschätzung liefert, können die Gepäckstückauftretenswahrscheinlichkeiten als rotationsinvariante Tori mit Mittelpunkt im Menschentorso modelliert werden (Abbildung 39, rechts). Diese können ebenfalls als Menge von Gaußverteilungen angenähert werden, deren Projektionen sich entsprechend berechnen lassen. Als Ergebnis dieses Schrittes können detektierte Gepäckstücke mittels ihrer Auftretenswahrscheinlichkeiten in Bezug zur nächstgelegenen Menschen-detection neubewertet werden.

### *Attributsextraktion*

Für jedes kontextbewertete Detektionspaar eines Menschen und eines Gepäckstücks werden nun semantische Attribute extrahiert, die den Detektionen beigeordnet werden (siehe Abbildung 36). Diese Attribute können zur Beantwortung von Suchanfragen wie „eine große Person mit hellem Oberteil und dunkler Hose, die einen roten Trolley zieht“ genutzt werden. Die Suchergebnisse werden dem Nutzer entsprechend ihrer Bewertung sortiert ausgegeben. Folgende Attribute werden extrahiert:

- Koffertyp: Hierbei handelt es sich um die Ausgabe des Objektdetektors für Gepäckstücke (Trolley, Rucksack, Koffer).
- Kofferfarbe: Die dominante Farbe jedes Gepäckstücks wird extrahiert, in Anlehnung an [10]: der HSL-Farbraum wird in 5 Farbbereiche unterteilt (rot,grün,blau,weiss,schwarz). Die dominante Farbe jedes Gepäckstücks wird berechnet, indem jede Detektion in den HSL-Farbraum konvertiert und jedes Pixel dem ähnlichsten der fünf Farbbereiche zugewiesen wird. Der Farbbereich mit den meisten Zuweisungen ist dann die dominante Farbe dieser Detektion. Die Farbzuzuweisung ist nur begrenzt robust gegenüber Beleuchtungsveränderungen, ist allerdings hinreichend zur Beantwortung einfacher Suchanfragen mit relativen Farbattributen wie beispielsweise hell oder dunkel.
- Menschengröße: Wir unterteilen die Fläche der Menschendetektion in einen oberen und einen unteren Teilbereich, die separat wie oben beschrieben farbzugewiesen werden. Folglich können Suchanfragen Farbangaben zur Kleidung von Ober- und Unterkörper eines Menschen enthalten.

### **Experimentelle Untersuchung auf den Feldmessungen**

Nachfolgend werden die Ergebnisse beschrieben, die mit einer Matlabimplementierung des obigen Verfahrens auf einem realistischen Datensatz erzielt wurden. Es wird die synthetische Trainingsmenge und der reale Testdatensatz beschrieben und der Einfluss des vorgeschlagenen Verfahrens auf Objektklassendetektion und Suchanfragengenauigkeit untersucht.

#### *Datensatz*

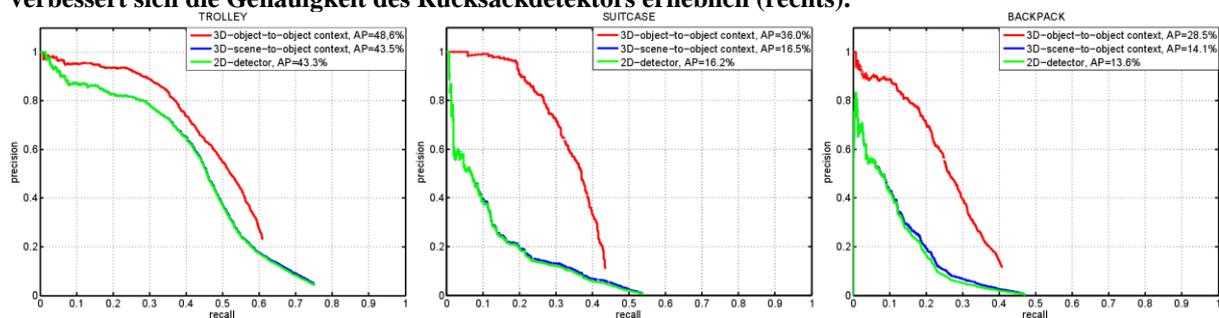
Aus einem Flughafenüberwachungskameranetzwerk wurden 6418 Einzelbilder von drei Kameras unter verschiedenen Pan-Tilt-Zoom-Einstellungen aufgenommen. Obwohl der Datensatz auch einige Bildfolgen enthält, liegt keine durchgehende zeitliche Konsistenz der Bilder vor. Zur Evaluierung wurde der Testdatensatz manuell mit Angaben zur Grundwahrheit bzgl. Position und Objektklasse annotiert; er enthält 3217 Trolleys, 995 Koffer und 2051 Rucksäcke. Zur Rekonstruktion wurde eine Größenverteilung der Menschen von  $1.8m \pm 0.1m$  angenommen.

#### *Trainingsdaten*

Die Objektdetektoren jeder Klasse wurden auf synthetisch erzeugten Bilddaten trainiert. Es wurden verschiedene texturierte CAD-Modelle des Anbieters turbosquid.com erworben, um die verschiedenen Erscheinungsbilder der Objektklassen abzubilden; insgesamt wurden 32 Menschen in verschiedenen Posen, 2 Trolleys, 5 Rucksäcke und 9 Koffer verwendet. Die Modelle wurden vor realen rein negativen, nicht im Testsatz enthaltenen Bilddaten aus dem Datensatz von [4] und 30 weiteren Bildern von typischen Flughafeninnenräumen gerendert. Abbildung 37 zeigt einige Beispiele. Trainingsannotationen werden automatisch aus den projizierten Begrenzungsrahmen der CAD-Modelle generiert. Das Training entspricht der Standardprozedur aus [8].



**Abbildung 40: Gefilterte Detektionen nach jedem der beschriebenen Verarbeitungsschritte: Reine 2D-Detektionen (links) von Menschen (grün) und Rucksäcken (rot); die Schwierigkeit der Detektion wenig charakteristischer Objektklassen wird deutlich. Nach 3D-Selbstkalibration können Detektionen, die mit der geschätzten 3D-Szene nicht konsistent sind, entfernt werden (Mitte), aber die Erkennungsrate ist noch immer zu gering. Nach Anwendung des 3D-Kontextmodells anhand der detektierten Menschen verbessert sich die Genauigkeit des Rucksackdetektors erheblich (rechts).**

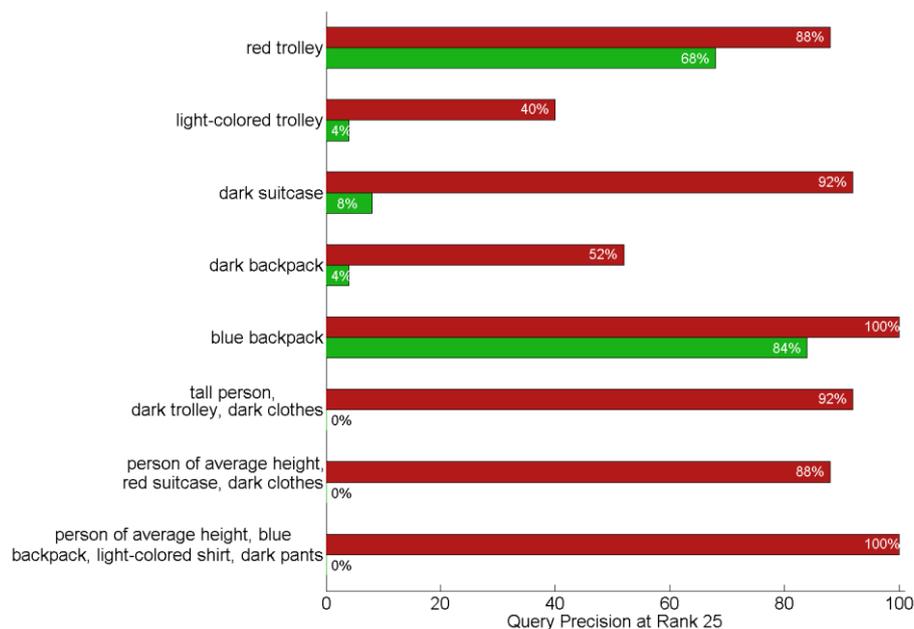


**Abbildung 41: Einfluss des vorgeschlagenen 3D-Kontextmodells auf die Genauigkeit der Objektklassendetektion für drei Kofferarten Trolley (links), Koffer (Mitte) und Rucksack (rechts) auf dem gesamten Testdatensatz aller Kameras. Dargestellt sind die reinen 2D-Detektionsergebnisse (grün), die Ergebnisse nach Filterung anhand der rekonstruierten Szene (blau) und das Endergebnis des 3D-Objektmodell (rot).**

### Objektklassenerkennung

Es wurde untersucht, welchen Einfluss das vorgeschlagene 3D-Kontextmodell auf die Verbesserung der einfachen voneinander unabhängig angewandten 2D-Objektdetektoren hat, vor und nach Filterung mit metrisch rekonstruierten 3D-Szeneninformationen. Abbildung 41 vergleicht Genauigkeit (Precision) und Abdeckung (Recall) für die drei Gepäckklassen bei einfacher 2D-Detektion (grüne Kurve), nach metrischer 3D-Szenenfilterung (blau) und bei Einbeziehung des 3D Objektmodell (rot). Da der verwendete 2D-Detektor derzeit keine Poseinformation zu den erkannten Menschen liefert, wurde das rotationsinvariante Kontextmodell aus Abschnitt „3D-Kontext“ eingesetzt. Beispielhafte Rucksackdetektionen in einem Einzelbild sind in Abbildung 40 dargestellt: links der einfache 2D-Detektor, in der Mitte nach Filterung mit 3D-Szeneninformationen, und rechts mit dem vollständigen 3D-Kontextmodell. Da Rucksäcke die am wenigsten diskriminativen Objekte aller drei Gepäckklassen sind, ist der reine 2D-Detektor vergleichsweise ungenau. Selbst nach Filterung der metrisch unplausiblen Detektoren anhand der 3D-Szenenrekonstruktion ist die Genauigkeit immer noch unzureichend für die vorliegende Suchaufgabe. In Abbildung 41 wird die Genauigkeit ggü. Abdeckungsrate für jede der Objektklassen auf dem gesamten Testdatensatz verglichen. Es ist eine deutliche Verbesserung der durchschnittlichen Genauigkeit aller drei Objektklassen erkennbar, die zwischen 5% und 20% liegt. Die Trolleyklasse ist am charakteristischsten und weist entsprechend einen weniger starken Genauigkeitsgewinn durch

das Kontextmodell auf. Bei den anderen beiden Gepäcktypen ist der Gewinn allerdings wesentlich für den Erfolg der Suchaufgabe. Allerdings reduziert sich die Abdeckung im 3D-Kontextschritt etwas, da der Menschendetektor einige Menschen aufgrund starker Verdeckung nicht erkennt und die zugehörigen Gepäckstücke entsprechend nicht kontextabhängig verarbeitet werden können. Nur durch die metrische Filterung anhand der rekonstruierten Szene ohne das Kontextmodell erhöht sich die Genauigkeit nicht erheblich, da die Menge an szenenkonsistenten Fehldetektionen unverändert hoch bleibt.



**Abbildung 42: Verbesserung der Genauigkeit der Suchanfragen bei Rang 25 für verschiedene Suchanfragen bei Verwendung des 3D-Objektmodell (rot); komplexe Suchanfragen mit metrischen Angaben können mit dem reinen 2D-Ansatz (grün) nicht beantwortet werden.**

### Suchanfragen

Um den Einfluss des vorgeschlagenen Ansatzes auf die Suchaufgabe zu bewerten, wurde die Suchgenauigkeit bei Rang 25 (P(25)) für verschiedene Suchanfragen verschiedener Komplexität bestimmt, für die sichergestellt werden konnte, dass wenigstens 25 suchrelevante Einzelbilder im Testsatz enthalten sind. Dieses Bewertungskriterium wird auch in standardisierten Bildsuch-Vergleichstests eingesetzt (imageclef.org). Es ist anzumerken, dass im vorliegenden Testdatensatz Menschen mehrfach durch dieselbe Kamera aufgenommen werden oder in mehr als nur einer Kamera zu sehen sein können. Für einfache Suchanfragen vergleichen wir die Suchgenauigkeit der 2D-Detektoren und des Kontextmodells; anspruchsvollere kombinierte Suchanfragen mit metrischen Angaben können nur mittels des Kontextmodells beantwortet werden und es sind entsprechend keine Vergleiche zur reinen 2D-Detektion möglich. Da die Grundwahrheit nur die Kofferposition und -Art angibt, steht keine Annotation zur Bewertung von Farb- oder Grössenanfragen zur Verfügung; folglich wurde P(25) manuell bestimmt. Alle Suchergebnisse werden entsprechend der Bewertungszahlen sortiert ausgegeben. Abbildung 42 zeigt P(25) für einige Testsuchen; Abbildung 43, Abbildung 44, Abbildung 45 zeigen Beispiele für Bilder als Antworten auf einige komplexe Suchanfragen. Ergebnisbilder derselben Person in aufeinanderfolgenden Bildern werden hier nicht gezeigt, um die Varianz der Suchergebnisse

besser darzustellen. Obwohl die Farbzuzuweisung in einigen Beleuchtungsszenarien fehlschlägt, kann der 3D-Kontextansatz die Suchergebnisse gegenüber den einzelnen 2D-Detektoren erheblich verbessern. Natürlich hängt die Genauigkeit der Suchantwort davon ab, wie charakteristisch ein Gepäckstück oder eine Menschenbeschreibung ist und wie zuverlässig die Detektion von Menschen als Suchkontext für Gepäckstücke funktioniert; es ist jedoch offensichtlich, dass ohne das vorgeschlagene 3D-Kontextmodell die Suchergebnisse im vorliegenden Anwendungsfall nicht praktisch nutzbar wären.

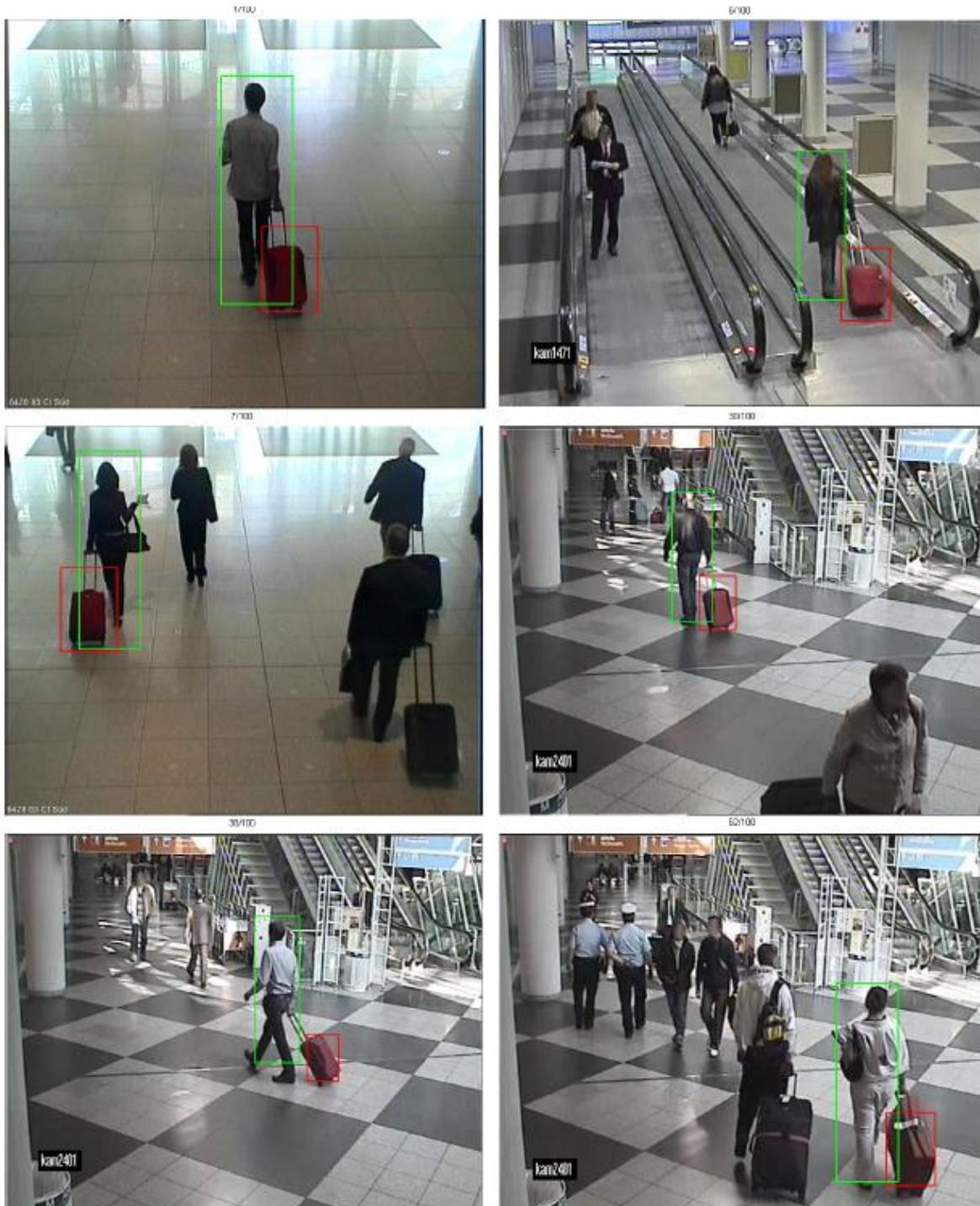
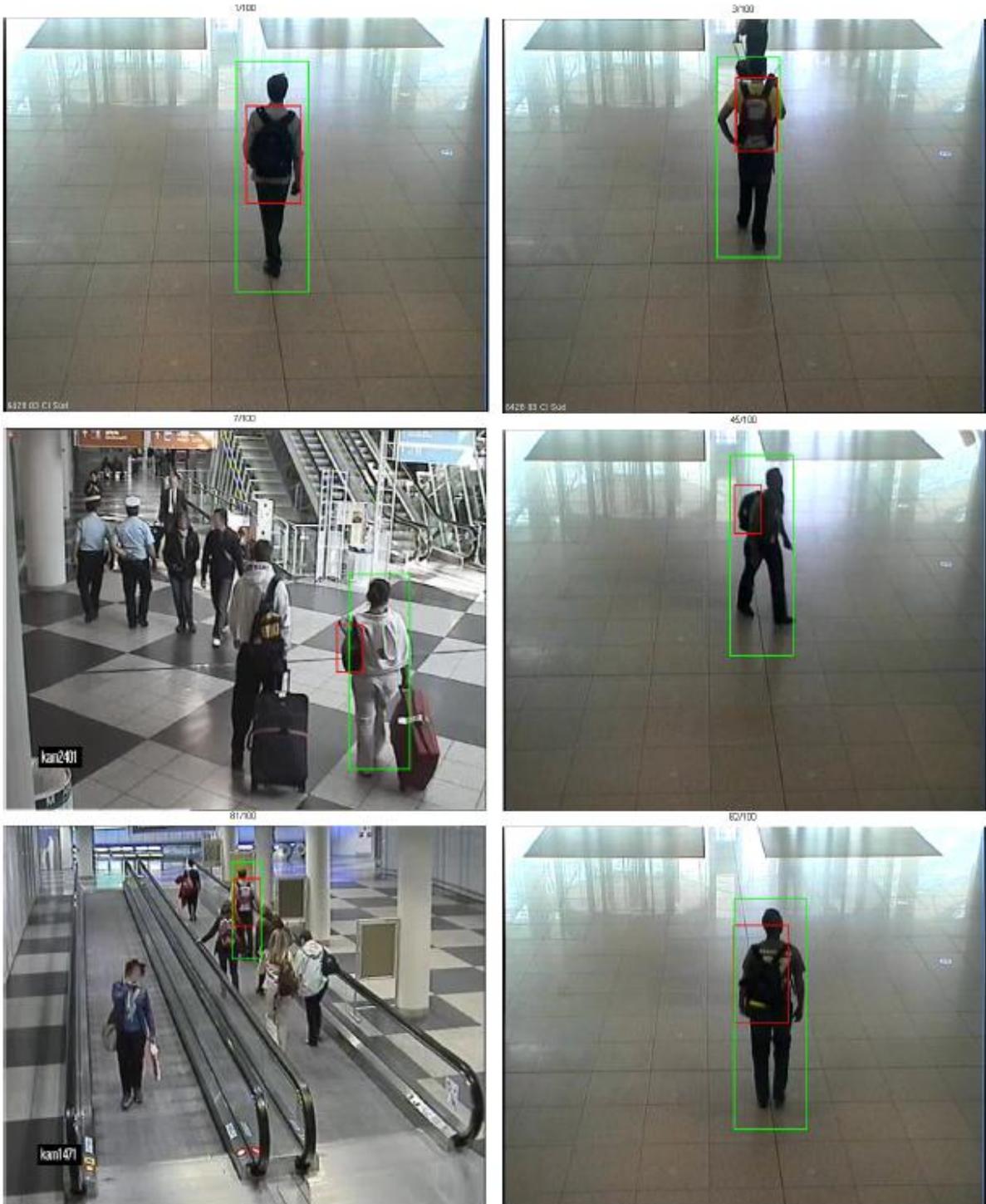
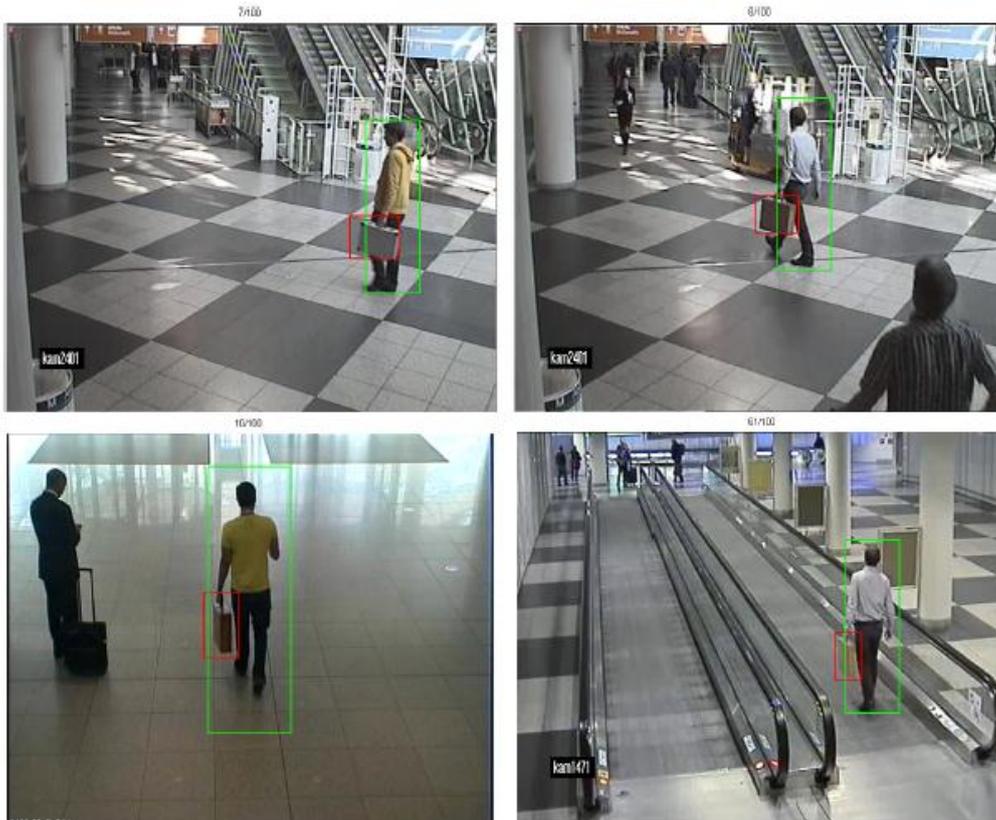


Abbildung 43: Beispiele für Suchergebnisse auf die Anfrage „Person durchschnittlicher Grösse mit rotem Trolley“.



**Abbildung 44: Beispiele für Suchergebnisse auf die Anfrage „Person durchschnittlicher Größe mit dunklem Rucksack, hellem Hemd und dunkler Hose“.**



**Abbildung 45: Beispiele für Suchergebnisse auf die Anfrage „grosse Person mit dunklem Koffer, hellem Hemd und dunkler Hose“.**

### **Schlussbemerkungen**

In der vorliegenden Arbeit wurde ein Ansatz zur Suche nach Einzelbildern in einem Netzwerk von Überwachungskameras vorgestellt; die Suche kann dabei auf komplexe attributsbasierte bildinhaltsbezogene Suchanfragen aufbauen, um Personen und deren Gepäckstücke wiederzufinden. Es wurde gezeigt, dass ein einfaches Modell zur Beschreibung des örtlichen 3D-Kontextes des Auftretens verschiedener Objektklassen zusammen mit einer Selbstkalibration der Szene eine erhebliche Verbesserung der Objektdetektion erreichen kann. Dies trifft insbesondere für Objektklassen zu, die sich durch ein wenig charakteristisches Erscheinungsbild auszeichnen, wie es beispielsweise für einige Arten von Gepäckstücken der Fall ist, die ohne Szenenkontext sonst nicht zuverlässig detektierbar sind.

In Bezug auf die Aufgabenstellung des Kofferszenarios ermöglicht das beschriebene Verfahren dem Sicherheitspersonal, anhand von 3D-CAD-Modellen und komplexen attributsbasierten Beschreibungen nach Personen mit Gepäckstücken in den Bilddaten beliebiger Überwachungskameras zu suchen. Die Versuchsergebnisse auf den Feldmessungsdaten zeigen, dass die Nutzbarkeit eines derartigen Ansatzes in der praktischen Anwendung am Flughafen München zumindest hinsichtlich der erzielbaren Präzision bereits gegeben ist. Andererseits sind weitere Arbeiten notwendig, um die Abdeckung (Recall) des Verfahrens unter den anspruchsvollen Beleuchtungs- und Verdeckungsverhältnissen des Flughafens weiter zu verbessern.

## **II.2. WICHTIGSTE POSITIONEN DES ZAHLENMÄßIGEN NACHWEISES**

Siehe auch „Zahlenmässiger Verwendungsnachweis“ vom 11.4.2012; die Gesamtkosten erfüllen den beantragten Gesamtförderungsrahmen.

Über die gesamte Laufzeit aller Teilvorhaben der EADS fielen 13.490 EUR Reisekosten an, die primär für die Beteiligung an den regelmäßigen Projektmeetings und –Workshops und die Feldtests und Begehungen am Flughafen, sowie für Reisen der Erstautoren zum Vortrag der veröffentlichten Konferenzbeiträge (ca. 2.000EUR) verwendet wurden.

Wesentliche Material- und Fremdleistungskosten beinhalten Fachliteratur, 3D-CAD-Modelle zum Training der Objekterkennungsverfahren (ca. 1.300EUR), Framegrabber-Hardware zur Anbindung und Aufzeichnung der Netzwerkkameras am Flughafen München (ca. 450EUR) sowie Entwicklungskosten für die Spiegelplattform und deren Ansteuerungselektronik (ca. 46.540EUR).

## **II.3. NOTWENDIGKEIT UND ANGEMESSENHEIT DER GELEISTETEN ARBEIT**

Der Projektantrag sah im Rahmen der identifizierten Forschungsthemen eine Festlegung der Detailzielsetzung jedes Arbeitspakets auch nach Projektbeginn vor, um die Abstimmung der Arbeiten mit den Anforderungen des Flughafens München zu ermöglichen. Diese Anpassungen sind in diesem Bericht sowie den Zwischenberichten ausführlich dargelegt.

Die im Rollbrückenszenario erarbeiteten theoretischen Ergebnisse erfüllen die Aufgabenstellung des ursprünglichen Projektantrags und die mit der FMG abgestimmte Detailzielsetzung. Die Evaluierung dieser Arbeiten konnte aufgrund der beschriebenen technischen Rahmenbedingungen und Sicherheitsbedenken nicht auf den Feldmessungsdaten durchgeführt werden. Die stattdessen auf vergleichbar relevanten Daten durchgeführten Tests erfüllen jedoch die Zielsetzung des Szenarios.

Im Kofferszenario konnte die antragsgemäße und mit der FMG abgestimmte Zielsetzung ebenfalls erfüllt werden. Im ursprünglichen Antrag bezogen sich die zu detektierenden Anomalien auf einzelne Objekte; diese Definition wurde unter Berücksichtigung der Vorstellungen der FMG auf die deutlich komplexere Kombination von Personen mit Gepäckstücken erweitert.

Die Arbeiten zum Laserblendungsszenario entstanden aus einer mit dem Projektträger abgestimmten Änderung der Zielsetzung heraus, die hauptsächlich durch den geänderten Bedarf der FMG im Bereich „Rücklaufverhinderung“ begründet ist. Diesem geänderten Bedarf wurde Rechnung getragen; die Untersuchungen zum Einsatz von Laserwarnern im zivilen Luftfahrtbereich erfüllen die geänderte Zielsetzung.

## **II.4. VERWERTBARKEIT DES ERGEBNISSES**

### **II.4.1. SCHUTZRECHTSANMELDUNGEN**

Teile der im Rollbrückenszenario des Teilvorhabens „Bildverarbeitungs-basierte Sicherheitstechnologien“ entwickelten Technologien wurden am 25.07.2008 beim Deutschen Patentamt zum Patent angemeldet:

Autoren	K. Schertler F. Schubert	EADS IW EADS IW
Titel	Verfahren und Einrichtung zur Erzeugung von fehlerreduzierten hochauflösenden und kontrastverbesserten Bildern	
Patent	DE 10 2008 034 979	
Einreichung	2008	
Erteilung	2011	
Abstract	Verfahren zur Erzeugung von fehlerreduzierten hochauflösenden und kontrastverbesserten Bildern aus Bildsequenzen mit Bildern niedriger Auflösung, die von einem Bildsensor mit einstellbaren Aufnahmeparametern gewonnen werden, wobei die Bildsequenzen der Bilder (K, R) niedriger Auflösung einer kombinierten Superresolution-Verarbeitung (SR) und High-Dynamic-Range-Rekonstruktionsverarbeitung (HDR) unterzogen werden, um fehlerreduzierte hochauflösende und kontrastverbesserte Bilder auf der Grundlage von in den Bildern niedriger Auflösung enthaltener redundanter und komplementärer Bildinformation zu erhalten, dadurch gekennzeichnet, dass in einem ersten Fusions-schritt die Bilder einer ersten Sequenz (K) von Bildern (K1, K2, ... KNK) niedriger Auflösung durch High-Dynamic-Range-Rekonstruktionsverarbeitung (HDR), bei welcher komplementäre Informationsanteile der Bilder (K1, K2, ... KNK) niedriger Auflösung zur Erhöhung des Kontrasts erarbeitet werden, zu einem oder mehreren Bildern (HDR) mit erhöhtem Kontrast fusioniert werden, dass in einem zweiten Fusions-schritt das oder die Bilder (HDR) mit erhöhtem Kontrast zur Superresolution-Verarbeitung (SR) verwendet werden, um ein oder mehrere fehlerreduzierte Bilder (HDR-SR) erhöhter Auflösung zu erhalten, und dass die Verarbeitung...	
URL	<a href="http://worldwide.espacenet.com/publicationDetails/originalDocument?FT=D&amp;date=20110707&amp;DB=worldwide.espacenet.com&amp;locale=en_EP&amp;CC=DE&amp;NR=102008034979B4&amp;KC=B4&amp;ND=4">http://worldwide.espacenet.com/publicationDetails/originalDocument?FT=D&amp;date=20110707&amp;DB=worldwide.espacenet.com&amp;locale=en_EP&amp;CC=DE&amp;NR=102008034979B4&amp;KC=B4&amp;ND=4</a>	

#### II.4.2. WEITERGEHENDE NUTZUNG DER ERGEBNISSE

EADS Deutschland GmbH möchte sich in der Zukunft strategisch verstärkt auf dem zivilen Sicherheitsmarkt engagieren. Die Themen Transportsicherheit sowie Sicherheit von kritischen Infrastrukturen wurden als interessante Marktsegmente erkannt. Um in Zukunft den Kunden bessere Produkte, eine kompetente Beratung und auf den Kunden zugeschnittene Dienstleistungen anbieten zu können, ist eine stetige Weiterentwicklung der Kernkompetenzen unverzichtbar. Hierfür sind detaillierte Kenntnisse der Prozesse der zukünftigen Kunden notwendig. Als Systemanbieter ist EADS auf ein Netzwerk kompetenter Partner angewiesen.

EADS Deutschland GmbH konnte als Ergebnis der Forschung im Vorhaben detaillierte Kenntnisse der Prozesse der einzelnen Hoheitsträger am Flughafen sowie der Leistungsfähigkeit der verwendeten Technologien gewinnen.

Die in AP 2 entwickelten Technologien im Bereich Bildverarbeitung besitzen ein breites sicherheitstechnisches Anwendungsspektrum, das nicht auf die Anwendung an Flughäfen beschränkt ist. Sämtliche Überwachungsszenarien mit hohem Personenaufkommen (öffentliche Plätze, Bahnhöfe, Schulhöfe, Sportstätten, usw.) sind gleichfalls im Fokus der Anwendung. Darüber hinaus kann an eine Übertragung der Methodik auf Verkehrsströme (Straßen, Autobahnen) gedacht werden. Die erarbeiteten Ergebnisse können also zu künftigen Produktentwicklungen in zahlreichen Sicherheitsanwendungen führen.

Eine Miniaturisierung und Integration der Verfahren innerhalb von Überwachungskameras („lokale Intelligenz“) eröffnet darüber hinaus weiteres wirtschaftliches Potential in Bezug auf die effizientere Nutzung von Kommunikationsressourcen in großen verteilten Video-Überwachungsnetzwerken.

EADS wird die im Rahmen des Kofferszenarios erarbeiteten Ergebnisse in neue Projektanträge einbringen, um den Reifegrad der Arbeiten weiter zu erhöhen und die Praxistauglichkeit in verschiedenen Einsatzgebieten zu untersuchen. Zum Zeitpunkt der Erstellung des Schlussberichts wird beispielsweise die Möglichkeit der Integration dieser Arbeiten in einen dedizierten Forschungsflughafen geprüft.

Firmenintern wird die Weiterentwicklung und Nutzung der entwickelten Ansätze im Rahmen von Produkten zum Infrastrukturschutz und der Grenzsicherung angestrebt.

Im Rahmen des Projektes konnte weiterhin das Netzwerk zu Forschungs- und Industriepartnern ausgebaut werden.

## II.5. FORTSCHRITT AUF DEM GEBIET DES VORHABENS BEI ANDEREN STELLEN

Seit Vorhabensbeginn wurden in der akademischen Forschung zum Teil erhebliche Fortschritte im Bereich der Bildverarbeitung und des Maschinensehens erzielt, die im laufenden Vorhaben berücksichtigt wurden. In vielen Fällen konnte auch die praktische Anwendbarkeit der entwickelten Methoden verbessert werden, da sich die Nutzung paralleler kostengünstiger Hardwareplattformen wie der Grafikkarten durch die Einführung von vereinheitlichten Programmierschnittstellen verbreitet durchgesetzt hat. Relevante Arbeiten zu den im Rahmen des Vorhabens bearbeiteten Themen sind in den einzelnen Abschnitten und im Literaturverzeichnis dargestellt (II.1.1, Abschnitt Verwandte Arbeiten, II.1.2, Stand der Technik, und II.1.3, Verwandte Arbeiten).

## II.6. VERÖFFENTLICHUNGEN DES ERGEBNISSES

Arbeitsanteile des Vorhabens wurden auf internationalen Konferenzen veröffentlicht, die im Folgenden zusammen mit (englischen) Kurzfassungen aufgelistet sind. Weitere Veröffentlichungen auf Grundlage der direkten Vorhabensergebnisse sind nicht geplant.

Autoren	F. Schubert K. Schertler K. Mikolajczyk	EADS IW EADS IW U Surrey
Titel	A Hands-On Approach to High-Dynamic-Range and Superresolution Fusion	
Konferenz	Workshop on the Applications of Computer Vision (WACV)	

Jahr	2009
Abstract	This paper discusses a new framework to enhance image and video quality. Recent advances in high-dynamic-range image fusion and superresolution make it possible to extend the intensity range or to increase the resolution of the image beyond the limitations of the sensor. In this paper, we propose a new way to combine both of these fusion methods in a two-stage scheme. To achieve robust image enhancement in practical application scenarios, we adapt state-of-the-art methods for automatic photometric camera calibration, controlled image acquisition, image fusion and tonemapping. With respect to high-dynamic-range reconstruction, we show that only two input images can sufficiently capture the dynamic range of the scene. The usefulness and performance of this system is demonstrated on images taken with various types of cameras.
URL	<a href="http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=5403080">http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=5403080</a>

Autoren	J. Liebelt C. Schmid	EADS IW INRIA Grenoble
Titel	Multi-View Object Class Detection with a 3D Geometric Model	
Konferenz	IEEE Conference on Computer Vision and Pattern Recognition (CVPR)	
Jahr	2010	
Abstract	<p>This paper presents a new approach for multi-view object class detection. Appearance and geometry are treated as separate learning tasks with different training data. Our approach uses a part model which discriminatively learns the object appearance with spatial pyramids from a database of real images, and encodes the 3D geometry of the object class with a generative representation built from a database of synthetic models. The geometric information is linked to the 2D training data and allows to perform an Approximate 3D pose estimation for generic object classes.</p> <p>The pose estimation provides an efficient method to evaluate the likelihood of groups of 2D part detections with respect to a full 3D geometry model in order to disambiguate and prune 2D detections and to handle occlusions. In contrast to other methods, neither tedious manual part annotation of training images nor explicit appearance matching between synthetic and real training data is required, which results in high geometric fidelity and in increased flexibility. On the 3D Object Category datasets CAR and BICYCLE [15], the current state-of-the-art benchmark for 3D object detection, our approach outperforms previously published results for viewpoint estimation.</p>	
URL	<a href="http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=5539836">http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=5539836</a>	

Autoren	J. Schels J. Liebelt K. Schertler R. Lienhart	EADS IW EADS IW EADS IW U Augsburg
Titel	Building a Semantic Part-Based Object Class Detector From Synthetic 3D Models	
Konferenz	IEEE International Conference on Multimedia and Expo (ICME)	
Jahr	2011	
Abstract	<p>This paper presents a new approach for multi-view object class detection based on part models. While most existing approaches have in common that they use real images for training, our approach requires only a database of synthetic 3D models to represent both the appearance and the geometry of an object class. We use semantically equivalent object points on 3D models to build part models</p>	

	and encode the local appearance of the parts by a discriminative learning method that applies AdaBoost to histograms of gradients. The geometric configuration of the parts is represented by spatial distributions which are also directly derived from the 3D models. For recognizing an object in an image, our model provides object hypotheses which are re-ranked with global appearance models. The 2D localization is evaluated on the PASCAL 2006 data set for cars and bicycles, showing that its performance can compete with state-of-the-art detection results.
URL	<a href="http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=6011850">http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=6011850</a>

Autoren	J. Schels J. Liebelt K. Schertler R. Lienhart	EADS IW EADS IW EADS IW U Augsburg
Titel	Synthetically Trained Multi-View Object Class and Viewpoint Detection for Advanced Image Retrieval	
Konferenz	ACM International Conference on Multimedia Retrieval (ICMR)	
Jahr	2011	
Abstract	This paper proposes a novel approach to multi-view object class and viewpoint detection for the retrieval of images showing one or several objects from a given viewpoint, a viewpoint range or any viewpoint in image databases. All detectors are trained exclusively on a few synthetic 3D models without any manual bounding-box, viewpoint or part annotation, making object class and viewpoint detection a scalable learning task. Previous work on this topic relies on the detection of object parts for each individual viewpoint, ignoring the responses of part detectors specific to other viewpoints. Instead, we explicitly exploit appearance ambiguities caused by spurious detections of parts under more than one viewpoint by combining all detector responses in a joint spatial pyramid encoding. We achieve state-of-the-art results in multi-view object class detection and viewpoint determination on current benchmarking data sets and demonstrate increased robustness to partial occlusion.	
URL	<a href="http://dl.acm.org/citation.cfm?id=1991999">http://dl.acm.org/citation.cfm?id=1991999</a>	

Autoren	J. Schels J. Liebelt R. Lienhart	EADS IW EADS IW U Augsburg
Titel	Self-Calibrating 3D Context for Retrieving People with Luggage	
Konferenz	IEEE International Workshop on Visual Surveillance (VS2011) in conjunction with Intl. Conference on Computer Vision (ICCV)	
Jahr	2011	
Abstract	We outline the retrieval of images from a network of security cameras by means of an attribute-based query. Our approach is based on detectors for several object classes which enable combined queries to retrieve people based on characteristic pieces of luggage. The approach works independently of camera recording frame rates since it does not rely on tracking or background assumptions, and it requires neither real training images nor manual annotations since it is entirely trained on synthetic data. By performing an approximate 3D auto-calibration for each camera from a few detected humans and exploiting object-level context in a 3D coordinate system, we can significantly improve the precision of otherwise weakly performing detectors for inconspicuous object classes. We evaluate our approach on data from an	

	airport security camera network and demonstrate the system's ability to respond to combined appearance and 3D metric contextual attribute queries over multiple cameras.
URL	<a href="http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=6130483">http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=6130483</a>

Autoren	J. Schels J. Liebelt R. Lienhart	EADS IW EADS IW U Augsburg
Titel	Learning an Object Class Representation on a Continuous Viewsphere	
Konferenz	IEEE Conference on Computer Vision and Pattern Recognition (CVPR)	
Jahr	2012	
Abstract	We propose an approach to multi-view object class detection and approximate 3D pose estimation. It relies on CAD models as positive training examples and discriminatively learns photometric object parts such that an optimal coverage of intra-class and viewpoint variation is guaranteed. In contrast to previous work, the approach shows a significantly reduced training set dependency while avoiding any manual training supervision or annotation, since it is capable of deriving all relevant information exclusively from the provided set of 3D CAD models and an arbitrary set of 2D negative images. In entirely circumventing semantic or view-based representations, part symmetries and co-occurrences between viewpoints can be efficiently exploited. This, in turn, leads to a significantly lower complexity whilen still achieving state-of-the-art performance on two current benchmark data sets for two different object classes.	
URL	to appear	

## **III. ERFOLGSKONTROLLBERICHT**

Im Erfolgskontrollbericht kann auf Abschnitte des Schlussberichts (Nrn. I. und II.) verwiesen werden.

### **III.1. BEITRAG DES ERGEBNISSES ZU DEN FÖRDERPOLITISCHEN ZIELEN**

Das vorliegende Vorhaben wurde im Rahmen des Programms „Forschung für die zivile Sicherheit“ gefördert, das die Bundesregierung am 24. Januar 2007 als Bestandteil der Hightech-Strategie für Deutschland beschlossen hat. Im Mittelpunkt des Sicherheitsforschungsprogramms steht die Verbesserung des Schutzes der Bürgerinnen und Bürger. Das Ziel ist, gesellschaftlichen Bedrohungen durch Terrorismus, organisierte Kriminalität, Naturkatastrophen oder technische Großunfälle entgegenzuwirken. Charakteristischer Anspruch an die Vorhaben des Programms ist das anwendungsorientierte Arbeiten innerhalb der Projekte durch Einbeziehung der gesamten Innovationskette von der Forschung über die Industrie bis hin zu den Endnutzern.

Eine ausführliche Beschreibung des vorwiegend fachlichen Mehrwerts der Vorhabensergebnisse findet sich in I.1 und II.1, Aussagen zur Verwertungsfähigkeit im Rahmen der förderpolitischen Ziele finden sich in II.4. Die im vorliegenden Vorhaben beschriebenen Ansätze bildverarbeitungs-basierte Sicherheitstechnologien fügen sich auch in den förderpolitischen Anspruch der Einbeziehung der gesamten Innovationskette ein. Durch innovative Forschung, die sich in internationalen Veröffentlichungen (II.6) und Schutzrechtsanmeldungen (III.3.1) widerspiegelt, wurde ein Bedarf adressiert, der aus Anwendersicht des Flughafens München mit bestehender Technologie zu Projektbeginn nicht erfüllt werden konnte. Die erarbeiteten Methoden konnten auf realistischem Datenmaterial angewandt und evaluiert werden; auf Grundlage dieser Ergebnisse sollen nach Vorhabensende in enger Abstimmung mit den EADS-Geschäftsbereichen Demonstratoren erstellt werden, die eine Nutzung der Forschungsergebnisse in Produkten vorbereiten. Zudem wurden datenschutzrechtliche Aspekte berücksichtigt, indem beispielsweise im sogenannten „Kofferszenario“ ausschließlich synthetisch erzeugte CAD-Modelle als Trainingsdaten eingesetzt wurden, wodurch die Notwendigkeit der Sammlung großer Mengen realer Aufnahmen von Überwachungskameras zum Zweck des Trainings der Detektionsmethoden überflüssig wird. Weiterhin konnten beispielsweise im „Kofferszenario“ auch Aspekte der Effizienz und kostengünstigen Nutzung bestehender Infrastruktur berücksichtigt werden, da die entwickelten Softwaremethoden eine automatische Kalibration und Szenenrekonstruktion auf bestehender Kamerahardware erlauben und dadurch den Mehrwert vorhandener Überwachungskameranetzwerke deutlich erhöhen.

### **III.2. WISSENSCHAFTLICH-TECHNISCHES ERGEBNIS DES VORHABENS**

Siehe II.1.

### **III.3. FORTSCHREIBUNG DES VERWERTUNGSPLANS**

#### **III.3.1. SCHUTZRECHTSANMELDUNGEN**

Siehe II.4.1.

#### **III.3.2. WIRTSCHAFTLICHE ERFOLGSAUSSICHTEN NACH PROJEKTENDE**

Siehe II.4.2.

### **III.3.3.**      *WISSENSCHAFTLICHE ERFOLGSAUSSICHTEN NACH PROJEKTENDE*

Siehe II.5, II.6.

### **III.3.4.**      *WISSENSCHAFTLICHE UND WIRTSCHAFTLICHE ANSCHLUSSFÄHIGKEIT*

Siehe II.4.2.

## **III.4.**      *ARBEITEN, DIE ZU KEINER LÖSUNG GEFÜHRT HABEN*

### **III.4.1.**      *DURCHWURFERKENNUNG ALS DYNAMISCHE ANOMALIEDETEKTION*

Die ursprünglich im Projektantrag als potentiell Anwendungsszenario genannte Auswertung von dynamischen Anomalien in Bewegungsmustern makroskopischer Personenströme („Rücklaufverhinderung“) wurde durch den Sicherheitsbeauftragten der FMG als weniger relevant eingeschätzt, da der Flughafen München kurz nach Projektbeginn bereits die Installation einer mechanischen Lösung zur Verhinderung des Rücklaufs von Personen und Gegenständen aus dem öffentlich zugänglichen in den beschränkten Passagierbereich begonnen hatte; siehe auch II.1.2, Einleitung. Nach Absprache mit dem Projektträger und der FMG wurde entschieden, die Zielsetzung des Arbeitsanteils zu verändern und auf eine Untersuchung von Laserblendungsdetektion auszurichten.

### **III.4.2.**      *VERSUCHE ZUR LASERBLENDUNGSDETEKTION IM KONTEXT ZIVILER LUFTFAHRT*

Statt der Durchwurferkennung wurden die bei der EADS verfügbaren bildverarbeitungs-basierten Laserwarnsysteme für den Infrarotbereich hinsichtlich der Anwendung im sichtbaren Bereich und den Anforderungen des zivilen Luftverkehrs untersucht, um die zunehmend auftretenden Vorfälle der Blendung von Verkehrsflugzeugen beim Landeanflug adressieren zu können. Ziel war eine Abschätzung von Machbarkeit, Kosten und Nutzen des Einsatzes derartiger Technologien im zivilen Flugverkehr. Abschnitt II.1.2 enthält eine Zusammenfassung dieser Arbeiten und einige Laborversuchsergebnisse. Allerdings konnte eine darüberhinausgehende detailliertere Betrachtung dieser Funktionen im Rahmen des Projektes nicht vorgenommen werden, da dies zum Teil zulassungsrechtliche Fragestellungen berührt hätte. Die generelle Anwendbarkeit des beschriebenen Laserdetektionsprinzips im optischen Wellenlängenbereich – als Voraussetzung zur Realisierung weitergehender Funktionen – wurde in den Laborversuchen und den algorithmischen Analysen jedoch nachgewiesen.

### **III.4.3.**      *ANWENDUNG DER BILDVERBESSERUNG AUF ROLLBRÜCKEN- FELDMESSUNGSDATEN*

Die Anwendung des entwickelten Bildverbesserungsverfahrens auf den Daten der Rollbrücken-Feldmessung scheiterte an technischen Unzulänglichkeiten der vorhandenen Kameras und Sicherheitsbedenken bezüglich des Austauschs der installierten Kameras durch geeignete Hardware; siehe II.1.1. Die vor Ort installierten Kameras erlauben keine gezielte Ansteuerung der Belichtungszeiten, um in schneller Folge dieselbe Szene in unterschiedlicher Helligkeit abbilden zu können. Wie in zuvor beschrieben, ist dies jedoch zur HDR-Verarbeitung unverzichtbar. Der Austausch dieser Kameras durch geeignete Hardware scheiterte jedoch an Sicherheitsbedenken. Einerseits hätte für den Zeitraum des Austauschs die Überwachung der

Rollbrücken nicht mehr sichergestellt werden können, wodurch sich möglicherweise Auswirkungen auf den Flugbetrieb ergeben hätten; andererseits liegen die Rollbrücken außerhalb des Zuständigkeitsbereichs der Flughafensicherheit, sodass alle Arbeiten vor Ort mit Polizeikräften zu koordinieren sind. Es wurde in Absprache mit der FMG entschieden, dass der zusätzliche Nutzen des Austauschs der Kameras zur Evaluierung auf Feldmessungsdaten gegenüber den bereits durchgeführten Tests auf vergleichbar relevanten Brückendaten nicht gerechtfertigt gewesen wäre.

### **III.5. PRÄSENTATIONSMÖGLICHKEITEN FÜR MÖGLICHE NUTZER**

In II.6 sind Veröffentlichungen aufgeführt, die im Rahmen des Vorhabens entstandene Arbeiten detailliert beschreiben. Diese Veröffentlichungen können als Anhaltspunkte für Reifegrad und Nutzbarkeit der erarbeiteten Methoden dienen; sie wurden auf Konferenzen vor internationalem Fachpublikum (WACV; CVPR,ICCV, ICME, ICMR) präsentiert.

Wie weiterhin in II.4.2 beschrieben, werden derzeit Möglichkeiten untersucht, um die Ergebnisse des Vorhabens in Demonstratoren zu integrieren. Hierzu zählen Demonstrationsinstallationen zum integrierten Infrastrukturschutz, die verschiedene Sensorquellen kombinieren und möglichst umfassende Lagebilder liefern sollen, sowie die Einbindung der bildinhaltsbasierten Suche in Überwachungskameradaten in einen dedizierten Forschungsflughafen. Die Prüfung dieser Vorhaben findet in enger Abstimmung mit dem EADS-Geschäftsbereich CASSIDIAN statt.

### **III.6. EINHALTUNG DER AUSGABEN- UND ZEITPLANUNG**

Am 12.4.2011 wurde ein Antrag auf kostenneutrale Verlängerung des Vorhabens um 3 Monate bis zum 30.9.2011 gestellt, der am 16.5.2011 genehmigt wurde. Die Verlängerung wurde unter Anderem notwendig, um die Arbeiten zu dem erst spät in das Vorhaben eingebrachte Thema „Laserblendung landender Luftfahrzeuge“ abschliessen zu können. Wie in III.4.1, II.1.2 erläutert, wurde dieses Thema erst später in das Vorhaben eingebracht, nachdem der Bedarf an einer bildgestützten Rücklauferkennung durch den Flughafen München aus Anwendersicht nicht länger gegeben war.

Die antragsgemäße Gesamtkostenplanung für das Teilvorhaben wurde eingehalten, siehe Dokument „Zahlenmässiger Verwendungsnachweis“ vom 11.4.2012.

## IV. LITERATURANGABEN

- [1] Y. Bao, M. Sun, and S. Savarese. Toward coherent object detection and scene layout understanding. In CVPR, 2010.
- [2] B. Benfold and I. Reid. Stable multi-target tracking in realtime surveillance video. In CVPR, 2011.
- [3] M. Breitenstein, E. Sommerlade, B. Leibe, L. V. Gool, and I. Reid. Probabilistic parameter selection for learning scene structure from video. In BMVC, 2008.
- [4] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In CVPR, 2005.
- [5] D. Damen and D. Hogg. Detecting carried objects in short video sequences. In ECCV, 2008.
- [6] S. K. Divvala, D. Hoiem, J. H. Hays, A. A. Efros, and M. Hebert. An empirical study of context in object detection. In CVPR, 2009.
- [7] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. The PASCAL Visual Object Classes Challenge 2010 (VOC2010) Results, 2010.
- [8] P. Felzenszwalb, D. McAllester, and D. Ramanan. A discriminatively trained, multiscale, deformable part model. In CVPR, 2008.
- [9] R. Feris, J. Petterson, B. Siddiquie, L. Brown, and S. Pankanti. Large-scale vehicle detection in challenging urban surveillance environments. In WACV, 2011.
- [10] R. Feris, B. Siddiquie, Y. Zhai, J. Petterson, L. Brown, and S. Pankanti. Attribute-based vehicle search in crowded surveillance videos. In ICMR, 2011.
- [11] D. Hoiem, A. Efros, and M. Hebert. Putting objects into perspective. In CVPR, 2006.
- [12] I. N. Junejo and H. Foroosh. Trajectory rectification and path modeling for video surveillance. In ICCV, 2007.
- [13] J. Liebelt, C. Schmid, and K. Schertler. Viewpointindependent object class detection using 3D feature maps. In CVPR, 2008.
- [14] F. Lv, T. Zhao, and R. Nevatia. Self-calibration of a camera from video of a walking human. In ICPR, 2002.
- [15] J. Marín, D. Vázquez, D. Gerónimo, and A. M. López. Learning appearance in virtual scenarios for pedestrian detection. In CVPR, 2010.
- [16] B. Micusik and T. Pajdla. Simultaneous surveillance camera calibration and foot-head homology estimation from human detections. In CVPR, 2010.
- [17] L. Pishchulin, A. Jain, C. Wojek, M. Andriluka, T. Thormählen, and B. Schiele. Learning people detection models from few training samples. In CVPR, 2011.

- [18] J. Renno, J. Orwell, and G. Jones. Learning surveillance tracking models for the self-calibrated ground plane. In BMVC, 2002.
- [19] J. Schels, J. Liebelt, K. Schertler, and R. Lienhart. Synthetically trained multi-view object class and viewpoint detection for advanced image retrieval. In ICMR, 2011.
- [20] B. Siddiquie, R. S. Feris, and L. S. Davis. Image ranking and retrieval based on multi-attribute queries. In CVPR, 2011.
- [21] S. Stalder, H. Grabner, and L. V. Gool. Exploring context to learn scene specific object detectors. In CVPR PETS workshop, 2009.
- [22] D. Tao, X. Li, X. Wu, and S. J. Maybank. Human carrying status in visual surveillance. In ICCV, 2006.
- [23] D. A. Vaquero, R. S. Feris, D. Tran, L. Brown, A. Hampapur, and M. Turk. Attribute-based people search in surveillance environments. In WACV, 2009.
- [24] C. Wojek, S. Walk, S. Roth, and B. Schiele. Monocular 3D scene understanding with explicit occlusion reasoning. In CVPR, 2011.
- [25] E. Reinhard, G. Ward, S. Pattanaik, P. Debevec, High Dynamic Range Imaging: Acquisition, Display and Image-based Lighting, Morgan Kaufmann, 2006, ISBN: 0125852630.
- [26] M. Gevrekci, B. K. Gunturk, Super-Resolution Approaches For Photometrically Diverse Image Sequences, ICASSP, 2007.
- [27] T. Pham , Spatiotonal Adaptivity in Super-Resolution of Under-Sampled Image Sequences, PhD Thesis, Delft University of Technology, 2006.
- [28] S. Farsiu, D. Robinson, M. Elad, P. Milanfar, Advances and Challenges in Super-Resolution, IJIST, 2004