

Schlussbericht

Schlussbericht	
Zuwendungsempfänger:	Technische Universität Ilmenau Fachgebiet für Neuroinformatik und Kognitive Robotik
Förderkennzeichen:	13 N 10 797
Vorhabensbezeichnung:	Analyse von Personenbewegungen an Flughäfen mittels zeitlich rückwärts- und vorwärtsgerichteter Videodatenströme (APFel)
Teilvorhaben:	Video- und Laser basiertes Personentracking und Trajektorienanalyse (VILPERT)
Laufzeit des Vorhabens:	01.01.2010 - 31.03.2014
Berichtszeitraum:	01.01.2010 - 31.03.2014

Inhaltsverzeichnis

1	Kurzdarstellung	3
1.1	Aufgabenstellung	3
1.2	Vorhabensbezogene Voraussetzungen	4
1.3	Planung und Ablauf des Vorhabens	5
1.4	Wissenschaftlich-technischer Stand zu Projektbeginn	7
1.4.1	Detektion & Wiedererkennung von Personen in Videodaten	7
1.4.2	Laserbasiertes Personentracking	8
1.4.3	Repräsentation und Analyse von Bewegungstrajektorien	8
1.4.4	Auffälligkeitsanalyse, Rück- und Vorwärtsprädiktion	9
1.4.5	Detektion, Tracking und Analyse von Personengruppen	9
1.5	Zusammenarbeit mit anderen Stellen	9
2	Eingehende Darstellung	9
2.1	Erzielte Ergebnisse	9
2.1.1	Arbeitspaket 1: Spezifikation	10
2.1.2	Arbeitspaket 2: Kamerabezogene Videodatenstromanalyse	10
2.1.3	Arbeitspaket 3: Kameraübergreifende Personenverfolgung	11
2.1.4	Arbeitspaket 4: Erfassung von Passagierströmen	11
2.1.5	Arbeitspaket 5: Analyse von Personentrajektorien	12
2.1.6	Arbeitspaket 8: Videoanalyse im Kontext von Gruppen	12
2.1.7	Arbeitspaket 9: Mengentheoretische Gruppenanalyse	14
2.1.8	Arbeitspaket 7: Implementierung, Test der Demonstratorstufen	14
2.2	Voraussichtlicher Nutzen und Verwertbarkeit der Ergebnisse	15
2.3	Fortschritt auf dem Gebiet des Vorhabens bei anderen Stellen	16
2.4	Erfolgte oder geplante Veröffentlichungen	16
3	Literatur	18

1 Kurzdarstellung

1.1 Aufgabenstellung

Das Ziel von APFeI war die Realisierung eines Systems, welches die Bewegung einer von einem Operator als verdächtig eingestuft und im Bild einer Überwachungskamera markierten Person (Zielperson) über mehrere Videokameras hinweg ermittelt und die dabei automatisch gewonnene Bewegungstrajektorie der Zielperson mit typischen Bewegungsspuren bekannter Passagierströme vergleicht. Das stark methodisch-technisch ausgerichtete Teilvorhaben APFeI-VILPERT beschäftigte sich dazu vornehmlich mit Aufgabenstellungen des video- und laserbasierten Personentrackings sowie der Analyse der ermittelten raum-zeitlichen Bewegungsspuren von Einzelpersonen und Gruppen im Rahmen des Einsatzfeldes (Flughafen). Die zu erarbeitenden Verfahren sollten sowohl zur Rückwärtsanalyse des aktuellen, gespeicherten Videodatenmaterials bezüglich des bisherigen Bewegungsverhaltens der markierten Zielperson als auch zur Vorwärtsprädiktion ihres hypothetischen weiteren Bewegungsverlaufs eingesetzt werden. Alle videobasierten Detektions- und Trackingverfahren sollten eine Rückwärtsanalyse in Hyperechzeit, d.h. vielfach schneller als der reale Zeitablauf, ermöglichen, um dem Operator einen Zeitgewinn bei der Sichtung des Videomaterials und damit einen Zeitvorsprung bei der Wahrnehmung seiner Überwachungsaufgaben zu garantieren. Dabei musste zu jeder Zeit gewährleistet sein, dass sich die Zielperson nicht in Menschenmengen verstecken kann, sondern auch innerhalb von Gruppen weiter getrackt wird.

Schwerpunkte des Teilvorhabens APFeI-VILPERT waren:

- methodische und technische Beiträge zur generischen Personendetektion auf der Basis universeller Personenmodelle zur Erstellung und Adaption eines spezifischen Personenmodells der Zielperson,
- zum Tracken des Personenmodells im Kamerabild,
- zur Fusion der Trackingsegmente bei Verdeckungen,
- sowie zum Gruppentracking und zur Mengentheoretischen Videodatenanalyse.

Darüber hinaus sollten mit diesem Teilvorhaben Beiträge zur Kamerakalibrierung und Weltmodellierung sowie zum kameraübergreifenden Tracken und zur Fusion der Trajektoriensegmente unter Anwendung von erworbenem Wissen über die spezifische Bewegungsdynamik der Personen vor Ort erbracht werden. Neben den videobasierten Verfahren zur Erfassung realer Passagierströme umfasste die Zielstellung auch den Einsatz neuartige laserbasierte Personentrackingverfahren. Weiterhin sollte das Teilvorhaben wesentliche Beiträge zur Ermittlung typischer Bewegungsspuren von Personen und Gruppen im Überwachungsbereich (bspw. aus den Daten der Videokameras oder temporär installierter Laserscanner) sowie zur Analyse von Personenbewegungen durch Vergleich mit typischen Bewegungsspuren leisten. Ziel war weiterhin das Ermöglichen einer Rückwärtsanalyse der gespeicherten Videodaten in Hyperechzeit mit den Rahmen des Teilvorhabens zu erarbeitenden videobasierten Detektions- und Trackingverfahren.

Die kameraübergreifende Verfolgung von Einzelpersonen stellte die Basis des APFeI-Systems dar. Eine Anwendbarkeit auf realen Flughäfen konnte jedoch erst durch die zusätzliche Betrachtung von Gruppen erzielt werden. Diesbezüglich sollten Methoden zur Detektion, Tracking und Prädiktion von Gruppen erarbeitet werden. Die Verfahren setzten dabei auf den jeweiligen Methoden für Einzelpersonen auf, mussten jedoch um neue Merkmale erweitert werden, da in der Regel nicht mehr alle Personen einer Gruppe erkannt werden können. Der dadurch entstehende sehr komplexe und hochdynamische Merkmalsraum sollte für eine Wiedererkennung von Gruppen genutzt werden. Um zu vermeiden, dass Personen in Gruppen unerkannt bleiben können, musste deren Identität mit mengentheoretischen Betrachtungen unter Zuhilfenahme des Ausschlussprinzips auf wenige Möglichkeiten eingegrenzt werden.

1.2 Vorhabensbezogene Voraussetzungen

Für die Umsetzung der Aufgabenstellung ist Expertise in den Bereichen Detektion und Tracking von Personen in visuellen Datenströmen, Lasertracking, Sensorkalibrierung und der Analyse von Bewegungstrajektorien notwendig. Das Fachgebiet Neuroinformatik und Kognitive Robotik (NIKR) der TU Ilmenau verfügt bereits über eine mehr als 10jährige ausgewiesene Forschungsexpertise in den genannten Themenfeldern und im Bereich Kognitiver Systeme. Insbesondere im Bereich der vision-basierten Mensch-Maschine-Interaktion reichen die Arbeiten am Fachgebiet bis ans Ende der 90er Jahre zurück. Sie umfassen innovative Verfahren zur audio-visuellen Personendetektion und zum Personentracking, wie z.B. [Schaffernicht 05], [Martin 06], [Scheidig 06], [Müller 07], [Gross 08] oder Arbeiten, die sich mit der Problematik der detaillierten Analyse von Gesichtsinformationen hinsichtlich Alter, Geschlecht, Identität und emotionalem Zustand eines Interaktionspartners bei der Mensch-Maschine-Interaktion befassen [Wilhelm 04], [Wilhelm 05], [Martin 08], [Stricker 09]. Weitere Arbeiten beschäftigen sich mit der Erkennung von Gesten und Zeigeposen zur Kommandierung mobiler Systeme [Corradini 00], [Gross 06], [Richarz 06], [Richarz 07], [Martin 09] sowie der Analyse von Bewegungstrajektorien bzw. typischen/atypischen Bewegungsmustern von Personen bei der Mensch-Roboter-Interaktion [Müller 08], [Hellbach 08], [Hellbach 09]. Im Bereich der 3D-Raumwahrnehmung und Sensorkalibrierung gibt es eigene Expertise, die durch [Einhorn 07] und [Einhorn 09], verdeutlicht werden kann.

Insgesamt kann das Fachgebiet NIKR auf den Themenfeldern Kognitive Systeme und Robotik seit Mitte der 90er Jahre auf die Einwerbung von 27 Förder- und Drittmittelprojekten, gefördert durch DFG, BMBF, EU, AiF/BMWA, Thüringer Wissenschafts- und Wirtschaftsministerien sowie von Unternehmen der freien Wirtschaft (Honda, DaimlerCrysler, MAN, Vattenfall, Powitec), verweisen. Ausdruck der Forschungsexpertise in den genannten Themenfeldern sind neben den oben genannten Publikationen mehr als 10 erfolgreich abgeschlossene Dissertationen und ca. 100 Diplomarbeiten. Eine Übersicht über die wichtigsten vorhabensrelevanten eigenen Publikationen des Fachgebietes ist im Anhang zu finden.

1.3 Planung und Ablauf des Vorhabens

Die zur Erfüllung der in Abschnitt 1.1 genannten Aufgaben notwendigen Arbeiten sind in 9 Arbeitspakete aufgeteilt. Die hierzu notwendigen F&E-Arbeiten sind in den wissenschaftlich technischen Arbeitspaketen AP1 bis AP5, AP8 und AP9 zusammengefasst (siehe Abb. 1). Die TU Ilmenau hat Beiträge zu jedem dieser Arbeitspakete, sowie zum AP7 (Implementierung und Test der Demonstratorstufen) geliefert.

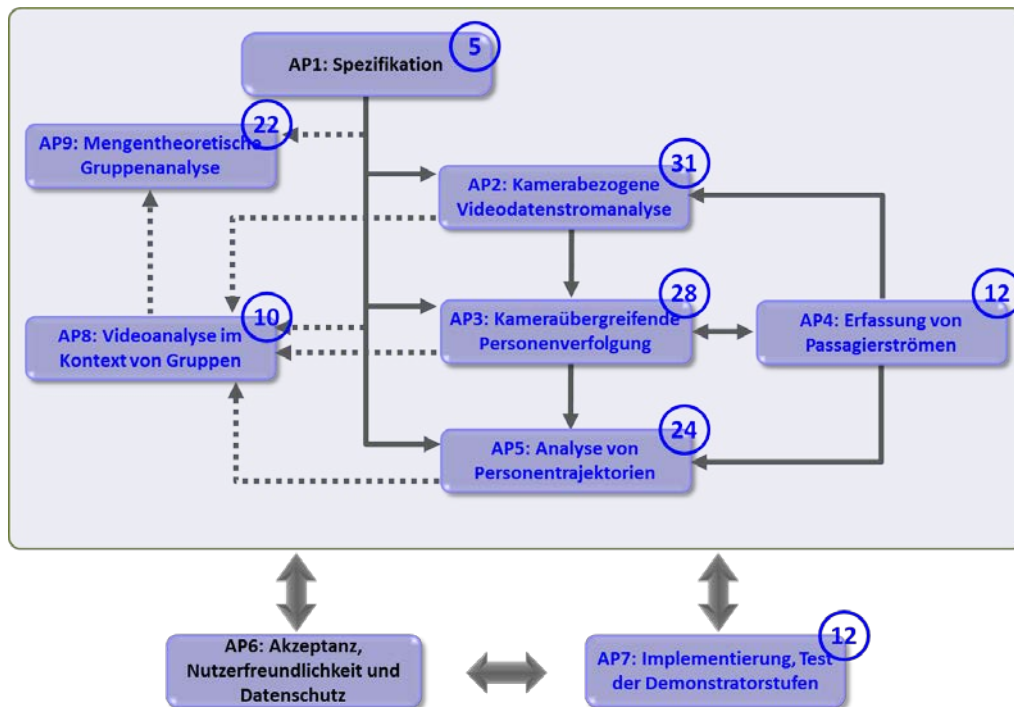


Abbildung 1: Übersicht über die Struktur und Verzahnung der Arbeitspakete (AP). Die APs, zu denen die TU Ilmenau beigetragen hat, sind mit blauer Schrift hervorgehoben. Zusätzlich ist die Verteilung der Personenmonate der TU Ilmenau auf die Arbeitspakete angegeben.

Das Gesamtvorhaben des Verbundprojektes war in drei Phasen aufgeteilt, welche jeweils mit einem Demonstrator abgeschlossen wurden. In der ersten Phase von M0-M18 wurden die erzielten Ergebnisse zu einer ersten Demonstratorstufe zusammengeführt, um frühzeitig Aussagen über die Qualität und zusätzlichen Innovationsbedarf bei den einzelnen Modulen machen zu können (Meilenstein 1). In der zweiten Phase (M19-M39) wurde die erste Demonstratorstufe zum finalen Demonstrator bezüglich der Verfolgung von Einzelpersonen weiterentwickelt. Dabei wurde eine 3-Monatsige zwenungsneutrale Verlängerung in Anspruch genommen. Das erstellte System ist in der Lage Einzelpersonen robust kameraübergreifend zu verfolgen. Das hohe Personenaufkommen an Flughäfen macht jedoch auch die Betrachtung von Gruppen notwendig. Das Vorhaben wurde daher um die APs 8 und 9 ergänzt, welche in einer 12-monatigen Aufstockungsphase bearbeitet wurden (M40-M51). Ein integriertes Gesamtsystem wurde zum Ende des Projektes fertig gestellt. Neben dem Meilenstein in M18 wurden projektintern Check-Points (siehe Projektplan) während der Projektlaufzeit festgelegt, um frühzeitig auf Abweichungen zum Projektplan reagieren zu können (Risikomanagement). Der beiliegende Projektplan gibt einen Überblick über die zeitliche Abfolge der beschriebenen Arbeitspakete. Die für dieses Teilvorhaben relevanten Arbeitspakete sind zwecks besserer Übersichtlichkeit rot hervorgehoben.

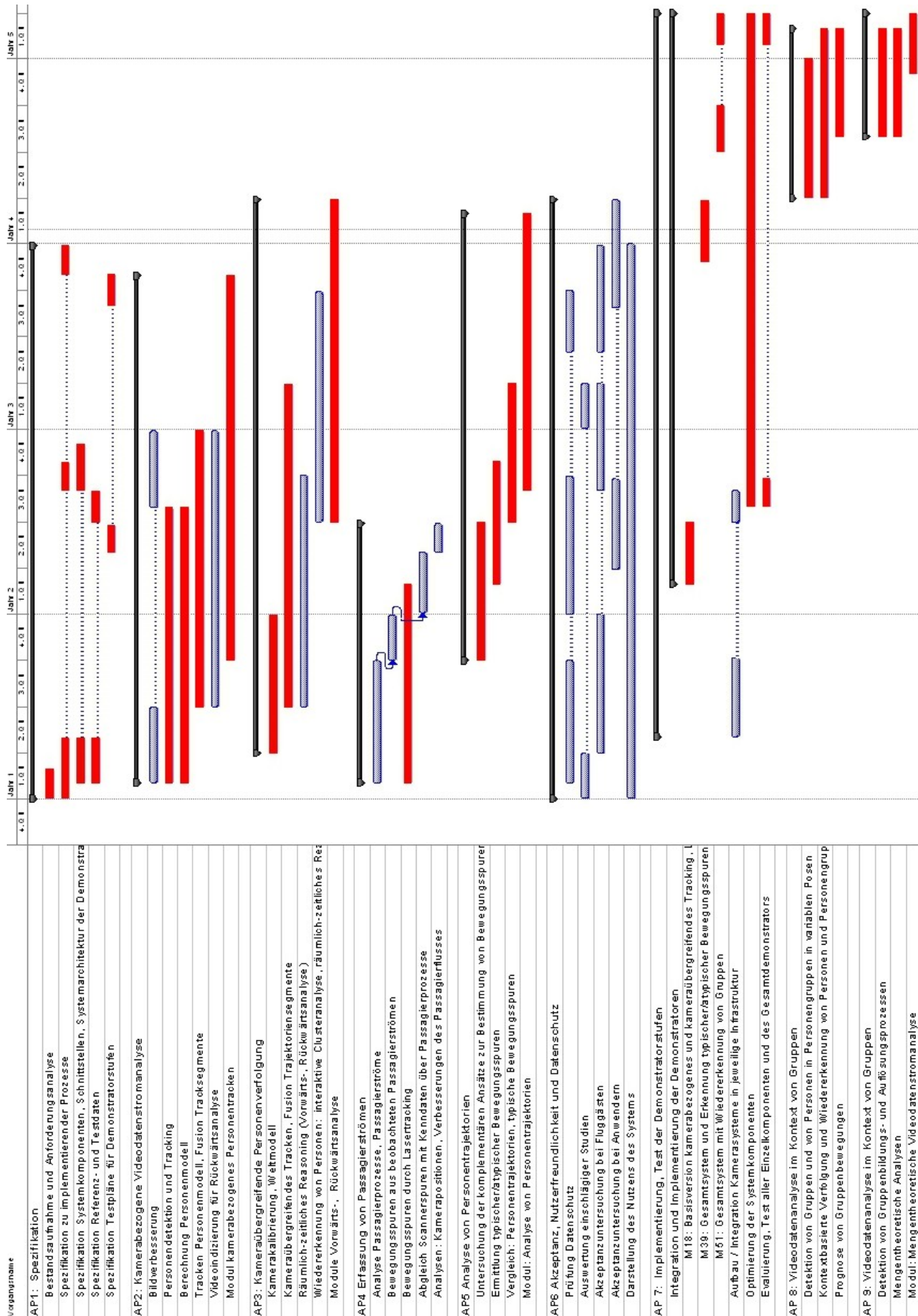


Abbildung 2: Übersicht über die Arbeitspakete, deren Umfang und zeitlichen Verlauf.

1.4 Wissenschaftlich-technischer Stand zu Projektbeginn

Nachfolgend ist der wissenschaftlich-technische Stand der Literatur zu Projektbeginn zu den im Projekt bearbeiteten Themenfeldern aufgeführt. Zur besseren Systematisierung sind die Ergebnisse der Recherche nach Themenfeldern geordnet.

1.4.1 Detektion & Wiedererkennung von Personen in Videodaten

Einen Schwerpunkt des Projekts APFEL stellt das robuste Detektieren und Tracken von Personen sowie das Wiedererkennen einer speziellen Person in einer öffentlichen Einsatzumgebung dar. Zum Thema der videobasierten Personendetektion und personenspezifischen Wiedererkennung gab es bereits zu Projektbeginn eine ungeheure Fülle an Arbeiten, die zum Teil recht problem- und anwendungsspezifische Annahmen treffen, die für APFEL nicht oder nur teilweise übertragbar waren. Um dennoch einen kompakten und aussagefähigen Überblick zu geben und auch die anfangs favorisierten Ansätze und Methoden einordnen zu können, wurde die nachfolgende grafische Systematisierung erstellt.

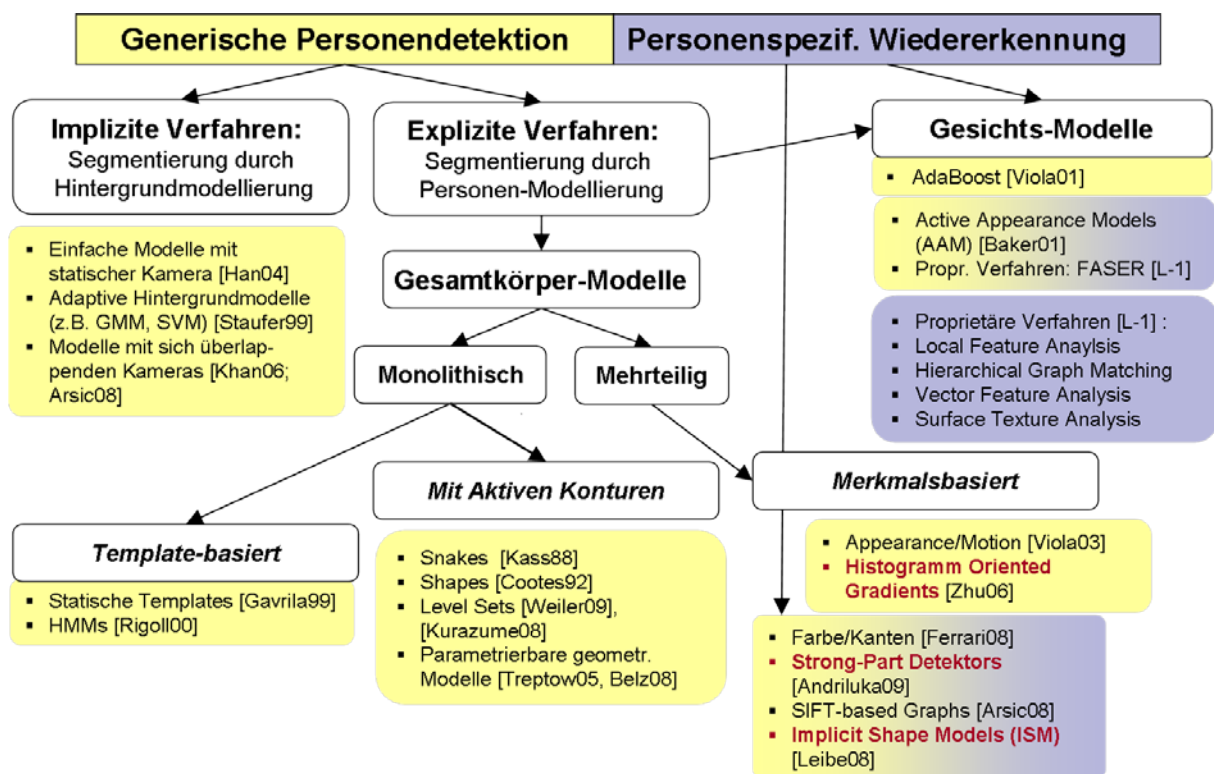


Abbildung 3: Im Ergebnis der Literaturrecherche erstellte systematische Übersicht zu den wichtigsten Verfahren der visuellen Personendetektion & Wiedererkennung. Die zu Beginn von APFEL durch den Antragsteller favorisierten Ansätze sind rot hervorgehoben.

Darin wurden die Verfahren zunächst grundsätzlich in solche unterschieden, die Personen allgemein aufgrund generischer Personenmerkmale detektieren und solche, die Personen aufgrund personenspezifischer Merkmale, wie Gesicht oder momentane Bekleidung, im Videodatenstrom wieder erkennen und damit auch tracken können. Die relevanten Arbeiten zur ersten Gruppe sind gelb hervorgehoben, während die in die zweite Gruppe einzuordnenden Verfahren blau hinterlegt sind. Besonders schwierig wird eine klare Zuordnung bei den Verfahren, die in Abb. 3 unten rechts aufgelistet sind. Sie gehören zum einen zu den expliziten Verfahren, die eine generische Personendetektion durch eine Vordergrundsegmentierung mittels eines Personenmodells vornehmen. Dabei handelt es sich um mehrteilige, das Erscheinungsbild des gesamten Körpers berücksichtigende Merkmale, die teilweise erst durch moderne Machine Learning Verfahren, wie AdaBoost, gewonnen werden. Zum anderen sind diese Verfahren bei entsprechender Erweiterung aber auch geeignet, Personen auf-

grund ihrer körperspezifischen Merkmale (Farbe und Textur der Bekleidung, auffällige Merkmale) wieder zu erkennen. Um diese Doppelzuordenbarkeit zu verdeutlichen, wurde diese Gruppe von Verfahren gelb und blau hinterlegt.

1.4.2 Laserbasiertes Personentracking

Im Forschungsfeld des laserbasierten Personentrackings existiert eine Vielzahl von unterschiedlichen Verfahren. Neben wenigen 3D- [Spinello 08a] oder Multi-Layer-Ansätzen [Gidel 08], verfolgt ein Großteil der Verfahren Personen in 2D Lasersensordaten. Häufige Gemeinsamkeit der Arbeiten ist die Extraktion von Features aus den Abstandsmessungen der Sensoren und die anschließende Klassifikation von Beinpaaren. Dies ermöglicht es Personen von anderen beweglichen Objekten zu unterscheiden.

Zu Projektbeginn adressierte Probleme im Themenfeld umfassten inter- und intraviduelle Verdeckungen, hohe Skalierbarkeit und das Verfolgen von großen Menschenmengen. Die damals aktuellsten Verfahren lernten daher besonders robuste, durch AdaBoost optimierte geometrische Features [Arras 07, Spinello 08b] oder beschäftigten sich mit der Entstehung, Auflösung und Verfolgung von Personengruppen [Lau 09].

Im Projekt APFeL wurde besonderer Wert auf die Trajektorie einer, durch den Operator ausgewählten Person bzw. der Person beinhaltenden Gruppe gelegt. Diese musste daher von den Bewegungstrajektorien anderer Personen (Gruppen) unterschieden werden. Durch das Verfahren von [Shao 08] können individuelle Trajektorien einer großen Anzahl von Personen auf Basis von spezifischen Bewegungsmustern und einem intelligenten Bewegungsmodell extrahiert werden. Die Verwendung eines Lasersensor-Netzwerks erhöht dabei die Robustheit gegenüber Verdeckungen. Wegen der hohen Skalierbarkeit und Passfähigkeit des Verfahrens zum Anwendungsfall wurde dieses Verfahren im Rahmen von AP 4 aufgegriffen und erweitert. Des Weiteren mussten in APFeL auch Gruppenentstehungs- und -Auflösungsprozesse erkannt werden, um eine mengentheoretische Zuordnung von Personen zu Gruppen zu ermöglichen. Hierfür wurde das Verfahren von [Lau 09] favorisiert.

1.4.3 Repräsentation und Analyse von Bewegungstrajektorien

Die Position der Personen, die von den unterschiedlichen Trackingmethoden erfasst werden, ergibt über die Zeit betrachtet eine Bewegungstrajektorie, welche auf verschiedene Weise repräsentiert werden kann. Die Wahl der Repräsentationsform ist dabei hochgradig von der Aufgabenstellung und den verwendeten Methoden für die Weiterverarbeitung abhängig [Müller 08]. Eine erste Unterscheidung findet bei der Wahl des Bezugssystems statt. Auch abhängig von den Trackingmethoden kann einerseits die 2D Koordinate im Kamerabild und andererseits ein 2D oder 3D Weltkoordinatensystem Anwendung finden [Junejo 07], [Murakami 00]. Letzteres bietet entscheidende Vorteile, insbesondere beim kameraübergreifenden Tracking.

Ein weiterer Aspekt bei der Modellierung von Bewegungsmustern ist der zeitliche Horizont, der Einfluss auf die Zukunft der Trajektorie hat, sowie die räumliche Auflösung. Eine mikroskopische Sicht, die sehr genau die Position beschreibt, bietet eher weniger Möglichkeiten, bedingte Abfolgen von besuchten Plätzen zu kodieren als eine makroskopische Sicht [Zajdel 05]. Mit einer Menge von diskreten Plätzen, welche auch unüberwacht durch Clustering gefunden werden können, ist eine solche Modellierung z.B. mit probabilistischen Zustandsübergangsmodellen, wie Hidden Markov Modellen (HMM), möglich.

Das Problem der Clustering von Trajektorien auf zum Teil multiplen räumlichen und zeitlichen Ebenen [Suzuki 07], [Weser 06], [Vasquez 04] ist mehrfach anzutreffen, um die bei der mikroskopischen Sicht anfallende hohe Datenmenge zu komprimieren und vergleichbar zu machen. Aber auch lokale Formmerkmale, die aus den mikroskopischen Spuren extrahiert werden, können der Klassifikation von Personeneigenschaften, wie Eile oder Aggressivität dienen [Müller 08], [Weser 06], [Suzuki 07], [Lee 03]. Hierbei wurden die Koordinaten über der Zeit mittels Fourierdeskriptoren, Polynomen oder Splines beschrieben und eventuell durch Anwendung von PCA komprimiert, bevor HMMs oder Neu-

ronale Klassifikatoren zur Klassifikation der Merkmalssequenzen genutzt wurden [Müller 08], [Rhodes 07].

1.4.4 Auffälligkeitsanalyse, Rück- und Vorwärtsprädiktion

Für die Auffälligkeitsanalyse bieten sich statistische Repräsentationen an. [Suzuki 07] clustert beispielsweise häufig auftretende Bewegungsmuster. Ebenfalls verbreitet ist der Aufbau von Aufenthaltswahrscheinlichkeitskarten, die für jede Raumposition neben der Häufigkeit einer Überquerung auch die Vorzugsrichtung und Geschwindigkeit akkumulieren. Neben der Detektion von weniger häufigen (potenziell auffälligen Trajektorien) ist damit auch die Prädiktion der Bewegungsrichtung (oder bei Umkehrung der Zeitrichtung auch die Rückverfolgung) möglich. Die Prädiktion von abstrakten Zeitreihen, auf welche sich die Bewegungsprädiktion zurückführen lässt, lässt einen vielfältigen Fundus an Methoden zu [Hahn 08], [Hellbach 08]. Beispielsweise erlauben parametrische Beschreibungen des Funktionsverlaufs die Extrapolation oder es werden lokale Modelle vom Verlauf der Funktionen gebildet, die jeweils den bedingten Verlauf zu einem Folgezeitpunkt in Abhängigkeit von der lokalen Vorgeschichte abbilden und anschließend stückweise aneinander gesetzt werden. Ebenfalls möglich ist die Vorhersage basierend auf einer neuronalen SOM Repräsentation [Weser 06]. Probabilistische Modelle auf makroskopischer Ebene (HMMs) können ebenfalls zur Vorhersage der nächsten Aufenthaltsorte eingesetzt werden.

1.4.5 Detektion, Tracking und Analyse von Personengruppen

Das Behandeln von Gruppen in Videodaten ist ein neues Forschungsfeld und baut im Wesentlichen auf die Detektion [Ge 12], [Zanotto 12] und das Tracken [Ge 12], [Bazzani 12a] von Einzelpersonen auf. Dabei werden die Personen einzeln detektiert und eine enge räumliche Nähe der Personen sowie erkannte Interaktionen zwischen den Personen als Hinweise auf das Vorhandensein einer Gruppe verwendet. Die so detektierten Gruppen werden dann als Ganzes mit probabilistischen Trackern verfolgt. Anhand der Dynamiken, die diese Gruppen aufweisen, wird in [Chang 11] versucht die Gruppenbewegungen zu klassifizieren. In [Bazzani 12b] wird in Videodaten berechnet, welche Regionen das hauptsächliche Interesse von Personengruppen auf sich ziehen.

1.5 Zusammenarbeit mit anderen Stellen

Die Zusammenarbeit des Fachgebiets Neuroinformatik und Kognitive Robotik der TU Ilmenau mit anderen Stellen erfolgte im Projekt APFeL an mehreren Stellen: In enger Zusammenarbeit wurde mit L-1 Identity Solutions und der Hochschule Ruhr West ein Konzept zur Detektion, Tracking und Wiedererkennung von Personen mit entsprechenden Komponenten zur Datenaufnahme, -Speicherung und -Kommunikation erstellt und implementiert. Um die Ergebnisse der einzelnen Verfahren der Projektparten zusammenzuführen wurde außerdem ein gemeinsames Konzept zur Fusion erarbeitet und umgesetzt. Eine Zusammenarbeit mit AVISTRA erfolgte bei der Prognose von Personenbewegungsspuren und bei Konzepten zur Verwendung von Gruppeninformationen. Um realistische Daten des Einsatzfeldes zu bekommen, wurden in Zusammenarbeit mit dem EASC und dem Flughafen Erfurt-Weimar Daten am Flugplatz Schönhagen und dem Flughafen Erfurt-Weimar aufgezeichnet. Zu Projektbeginn wurden in Zusammenarbeit mit dem Flughafen Hannover realistische Daten von Passagierströmen mittels Lasern aufgezeichnet. Um die rechtlichen Rahmenbedingungen festzulegen erfolgte die Abstimmung mit dem Thüringer Datenschutz in Zusammenarbeit mit der Ruhr Universität Bochum. Der Austausch der erzielten Ergebnisse erfolgte mit allen Projektpartnern.

2 Eingehende Darstellung

2.1 Erzielte Ergebnisse

Nachfolgend werden die in APFeL erzielten Ergebnisse gruppiert nach Arbeitspaketen aufgelistet.

2.1.1 Arbeitspaket 1: Spezifikation

In Zusammenarbeit mit den technischen Projektpartnern wurde ein Spezifikationsdokument erstellt. Weiterhin wurden im Rahmen des AP1 zu mehreren Zeitpunkten Testaufnahmen zur Evaluation der Algorithmen erstellt. Dabei wurde darauf geachtet, den Schwierigkeitsgrad für Detektion, Tracking und Wiedererkennung von Personen zu vorherigen Aufnahmen jeweils zu steigern. Zu den Aufnahmen konnte mittels Laser-Range-Scannern halbautomatisch eine Ground-Truth-Datenbasis erstellt werden (siehe AP4). Dadurch konnten erzielte Fortschritte bei den Algorithmen zur Detektion und dem Tracking von Personen (siehe AP3), sowie zu deren Wiedererkennung (siehe AP2) dokumentiert werden. Außerdem wurden die Aufnahmen für den Vergleich mit komplementären Ansätzen der Projektpartner genutzt.

Ein weiterer Schwerpunkt von AP1 lag in der Erstellung von Konzepten für die Datenablage in einer Datenbank und für die Kommunikation von Modulen zur Videodatenstromanalyse in Zusammenarbeit mit den technischen Projektpartnern. Die Tragbarkeit der Konzepte wurde anhand mehrerer Demonstratorstufen nachgewiesen (siehe AP7). Die Umsetzung der Datenbank und der Kommunikations-Komponente im Zusammenwirken mit allen APFel-Komponenten ist in [Kolarow 13] publiziert.

2.1.2 Arbeitspaket 2: Kamerabezogene Videodatenstromanalyse

Im Rahmen des AP2 wurden Verfahren zur (1) Vordergrund-Hintergrund-Segmentierung, zur (2) Detektion und zum (3) Tracking von Personen und zur (4) Fusion von Tracksegmenten entwickelt.

1) Um neben den schnellen Verfahren zur Vorverarbeitung des Projektpartners HRW auch ein sehr genaues und robustes Verfahren zur Vordergrund-Hintergrund-Segmentierung zu erhalten, wurde im Rahmen des AP2.2 eine Verfahren mittels Mixture of Gaussians umgesetzt. Dieses wird in Kombination mit dem Personendetektor genutzt, um die Robustheit der Detektionen zu steigern (AP2.2, 2.3).

2) Zur Detektion von Personen wurde zunächst ein Histogram of Oriented Gradients (HOG) [Dalal 05] eingesetzt. Dabei wurde das Verfahren mit der Vordergrund-Hintergrund-Segmentierung kombiniert. Außerdem wurden geeignete Bildbereiche pro Auflösungsstufe ausgewählt. Dies wurde erreicht, indem die extrinsischen Parameter der Kamera durch eine Kamerakalibrierung ermittelt wurden (siehe AP3). Durch diese Anpassungen und Optimierungen konnte das HOG-Verfahren nahezu echtzeitfähig gemacht werden. Um eine echtzeitfähige Detektion von Personen bei 10 Bildern pro Sekunde zu ermöglichen, wurde das HOG-Verfahren durch das 2011 veröffentlichte Contour-Cues-Verfahren [Wu 11] ersetzt. In Kombination mit der Vordergrund-Hintergrund-Segmentierung und der Kamerakalibrierung kann die Personendetektion auf 10 HD-Bildern (1600x1200 Pixel) pro Sekunde durchgeführt werden. Dies entspricht einer Beschleunigung des Original-HOG-Verfahrens um den Faktor 112 (Kamerakalibrierung Faktor 14, Vordergrund-Hintergrund-Segmentierung Faktor 8). Die Ergebnisse sind in [Kolarow 13] publiziert.

3) Basierend auf den gewonnenen Regionen der Detektoren wurde in Kooperation mit dem Projektpartner L-1 ein Verfahren zur Gewinnung eines Personenmodells und des darauf basierenden kamerabezogenen Trackings erstellt (AP 2.3, 2.4). Das Verfahren basiert auf einer logarithmischen Suche und benutzt spezifisch ausgewählte Merkmale aus homogenen Regionen. Das Ergebnis ist ein Verfahren, welches bei gleicher Qualität zehn Mal schneller ist als aktuelle, aus der Literatur bekannte, Verfahren. Durch die zusätzliche Annahme der Bewegung auf einer ebenen Fläche kann die Performance weiter gesteigert werden. Das Ziel eines hyper-echtzeitfähigen Trackings konnte damit vollständig erfüllt werden: Vier Personen können gleichzeitig auf HD-Bildern in mindestens 30-facher Videoechtzeit getrackt werden. Damit ist es auch möglich Personen in Bildern zu tracken, die vom Personendetektor nicht rechtzeitig bearbeitet werden konnten. Die Ergebnisse sind in [Kolarow 12] publiziert.

4) Zur Fusion der Trackingergebnisse aller Partner (ganze Person, Oberkörper, Gesicht) wurde ein Verfahren entwickelt, dass zunächst alle Teilergebnisse mittels bekannter extrinsischer Kameraparameter (siehe AP3.1 Kamerakalibrierung) in ein globales Koordinatensystem transformiert und ähnliche Personenhypothesen auf Grund der räumlichen Nähe mittels Kalman-Filtern fusioniert (AP 2.4).

Die Fusion erfolgt in Echtzeit. Es können auch Hypothesen verschiedener Kameras fusioniert werden, wenn sich die entsprechenden Kameras überlappen. Das Verfahren ist in [Kolarow 13] publiziert. Die Teilergebnisse aus den APs 2.2–2.4 wurden zu einem robusten funktionsfähigen Demonstrator zum kamerabezogenen Tracken zusammengeführt (AP2.6). Auch das erfolgreiche Zusammenspiel mit den Verfahren der Projektpartner konnte gezeigt werden (AP2.6).

2.1.3 Arbeitspaket 3: Kameraübergreifende Personenverfolgung

Im Rahmen des AP3 wurden Verfahren zur (1) Kamerakalibrierung, zur (2) kleidungsbasierten Wiedererkennung von Personen und zur (3) Fusion von Wiedererkennungsverfahren entwickelt.

1) Die Kalibrierung der Kamera (AP3.1) zur Ermittlung der intrinsischen Parameter erfolgt mittels Schachbrett-Testmuster. Die extrinsischen Kameraparameter werden mit Hilfe eines Referenzobjektes ermittelt. Das Objekt kann sowohl im Bild einer Kamera als auch in der globalen Karte mittels Laser erkannt werden. Durch ein Optimierungsverfahren werden die wahrscheinlichsten extrinsischen Parameter ermittelt. Um alle Kameras auch bei nicht überlappenden Ansichten in ein gemeinsames Koordinatensystem zu integrieren erfolgt eine Kopplung mit der Laser-Kalibrierung (siehe AP4).

2) Zum kameraübergreifenden Tracken (AP 3.2) wurde ein Verfahren zur kleidungsbasierten Wiedererkennung von Personen implementiert. Für jede Person werden einfache Textur- und Farbmerkmale ermittelt und zu einem Modell zusammengefasst. Um möglichst wenige Merkmale pro Person ermitteln zu müssen – um damit die Echtzeitanforderungen zu erfüllen – wird eine automatische Merkmalsauswahl eingesetzt. Das Verfahren ist in [Eisenbach 12] publiziert. Ergänzend zu diesem Verfahren werden Distance-Metric-Learning-Techniken eingesetzt um eine geeignete Ähnlichkeitsfunktion für den Vergleich von Personenhypothesen zu ermitteln. Hierfür kommt ein 2013 veröffentlichter Algorithmus [An 13] zum Einsatz, der eine Canonical Correlation Analysis (CCA) in Kombination mit einem Referenz-Deskriptor verwendet. Des Weiteren wurden Arbeiten zur Normierung von Farben der Kleidung von Personen bei unterschiedlichen Beleuchtungsbedingungen durchgeführt. Das Verfahren basiert auf der Beobachtung unterschiedlicher Personen während sie sich durch die Szene bewegen. Bei Beobachtung mehrerer Personen an der gleichen Stelle kann ein Optimierungsproblem gelöst werden um auf die tatsächliche Farbe der Kleidung zu schließen. Das Verfahren ist in [Eisenbach 13] publiziert.

3) Die Kopplungen mit den Verfahren der technischen Projektpartner zur Wiedererkennung von Personen konnte erfolgreich über eine Datenbank in Kombination mit einer Kommunikationskomponente realisiert werden (AP3.5, Konzept siehe AP1). Die Ergebnisse der einzelnen Wiedererkennungsverfahren werden anhand der False Acceptance Rate (FAR) normiert und mittels Score-Level-Fusion kombiniert. Für die Zuordnung von Personenhypothesen aus Detektionen des Kopfes, des Oberkörpers und ganzer Personen wird auf Ergebnisse aus AP2.4 (Fusion der Trajektoriensegmente) zurückgegriffen. Das Verfahren zur Score-Level-Fusion ist in [Kolarow 13] publiziert.

Das Ergebnis von AP3 ist ein funktionsfähiger Demonstrator zur kameraübergreifenden Wiedererkennung von Personen, der im Kontext der Vorwärts- und Rückwärtsanalyse fertig gestellt wurde (AP3.5). Die Vorwärts- bzw. Rückwärtsanalyse kann aufgrund zahlreicher Optimierungen Szenarioabhängig in etwa 600-facher Videoechtzeit erfolgen, wenn die dafür benötigten Merkmale bereits während der Live-Analyse berechnet werden.

2.1.4 Arbeitspaket 4: Erfassung von Passagierströmen

Im Rahmen von AP4 wurde ein Verfahren zum Aufzeichnen von Bewegungsspuren mittels Lasertracking erstellt. Es wurden zwei Ansätze verfolgt: Das Tracking auf Höhe des Oberkörpers und auf Höhe der Füße. Zusätzlich wurde ein transportables und autarkes (kabelloses) Lasermesssystem entwickelt, welches eine schnelle Aufzeichnung der Trajektorien ohne eine Behinderung der Passagiere ermöglicht (AP4.3). Dadurch können über 100 Personen in Echtzeit getrackt und lückenlose Trajektorien aufgezeichnet werden. Nur in seltenen Fällen treten vereinzelt Fehler auf, die das Resultat von

ungünstigen, längeren Verdeckungen sind. Diese können aber durch gut geplante Aufstellpositionen der Laser-Range-Scanner vermieden werden. Das Verfahren ist in [Schenk 11b] publiziert. Eingesetzt wurde dieses Verfahren zur Erstellung einer Ground-Truth-Datenbasis zu den Video-Aufnahmen (siehe AP1).

Des Weiteren wurde eine halb-automatische Kalibrierung des Multi-LRS-Systems implementiert (AP4.3). Die Ergebnisse sind in [Schenk 12a] und [Schenk 12b] publiziert. Das kalibrierte Multi-LRS-System wird eingesetzt um eine Kalibrierung mehrerer nicht überlappender Kameras in ein gemeinsames globales Koordinatensystem zu ermöglichen (AP3.1). Außerdem werden damit aufgezeichnete Bewegungsspuren als Datengrundlage für die Prädiktion von Bewegungsspuren (AP5) genutzt.

2.1.5 Arbeitspaket 5: Analyse von Personentrajektorien

Die Arbeiten zu AP5 umfassten (1) Untersuchungen zur Repräsentation von Personentrajektorien, (2) die Entwicklung eines Verfahrens zur Ermittlung und Repräsentation typischer Bewegungsspuren und (3) die Implementierung von Algorithmen zur Prädiktion von Personenbewegungen.

1) Um Personentrajektorien zu typischen Bewegungsspuren gruppieren zu können ist eine geeignete Repräsentation notwendig. Hierzu wurden vergleichende Untersuchungen durchgeführt (AP5.1). Es wurden unter anderem die Unterschiede zwischen aus Laserdaten und Kameradaten gewonnenen Trajektorien berücksichtigt. Als am geeignetsten erwies sich eine äquidistante Repräsentation der Bewegungsspuren, bei der Übergangszeiten zwischen Knoten auf der Trajektorie gespeichert werden.

2) Zur Ermittlung typischer Bewegungsspuren (AP5.2) wurden die Bewegungsspuren mehrerer Personen im Erfassungsbereich mittels Mean-Shift-Verfahren geclustert. Die Transitionen zwischen Clustern ergeben sich aus der Statistik von Personenbewegungen zwischen den entsprechenden Clustern. Das Ergebnis ist ein Graph entlang dessen Kanten sich Personen durch die Szene bewegen können (siehe Abb. 4). Übergangswahrscheinlichkeiten und -zeiten zwischen Knoten werden berücksichtigt. Der Graph ist gut geeignet um damit die wahrscheinliche Fortsetzung von Personenbewegungsspuren zu präzisieren.

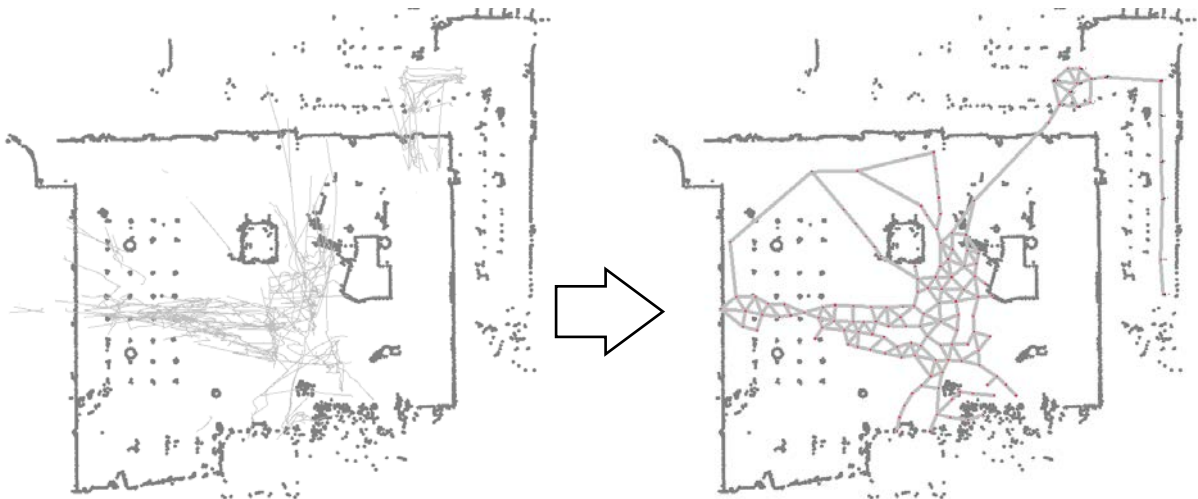


Abbildung 4: Repräsentation typischer Bewegungsspuren in Form eines Graphs. Bild links: Beobachtete Trajektorien (graue Linien) des visuellen Trackers (siehe AP2.3); Bild rechts: Durch Clustering erzeugter Graph typischer Bewegungsspuren (Knoten rot, Kanten grau).

3) Die ursprünglich für AP5.3 geplanten Arbeiten zum Vergleich von Personentrajektorien mit typischen Bewegungsspuren wurden nach Abstimmung mit den Projektpartnern und dem Projektträger durch Arbeiten zur Prädiktion der Fortsetzung von Personentrajektorien ersetzt. Dies war erforderlich, weil aus datenschutzrechtlichen Gründen geeignete Realweltdaten nicht aufgezeichnet werden konnten. Für die Vorhersage wahrscheinlicher Fortsetzungen der Trajektorien wurden zwei Ansätze

implementiert: Eine Monte-Carlo-Simulation und eine mathematische Faltung der Übergangswahrscheinlichkeiten, die beide auf dem Graph zur Repräsentation typischer Bewegungsspuren aufsetzen. Letztere Methode erzielt die genaueren Ergebnisse. Sie ließ sich außerdem durch eine Übertragung der Berechnungen in den Frequenzraum deutlich beschleunigen. Das Ergebnis ist eine Prädiktion, die Szenario-abhängig in etwa 2700-facher Echtzeit berechnet werden kann. Durch die Verwendung von Markov-Prozessen 5. Ordnung (d.h. die letzten 5 Knoten in dem Graph typischer Bewegungsspuren, die die Person passiert hat, werden für die Prädiktion verwendet) konnten außerdem sehr präzise Vorhersagen über die Fortsetzung von Personentrajektorien gemacht werden (siehe Abb. 5).

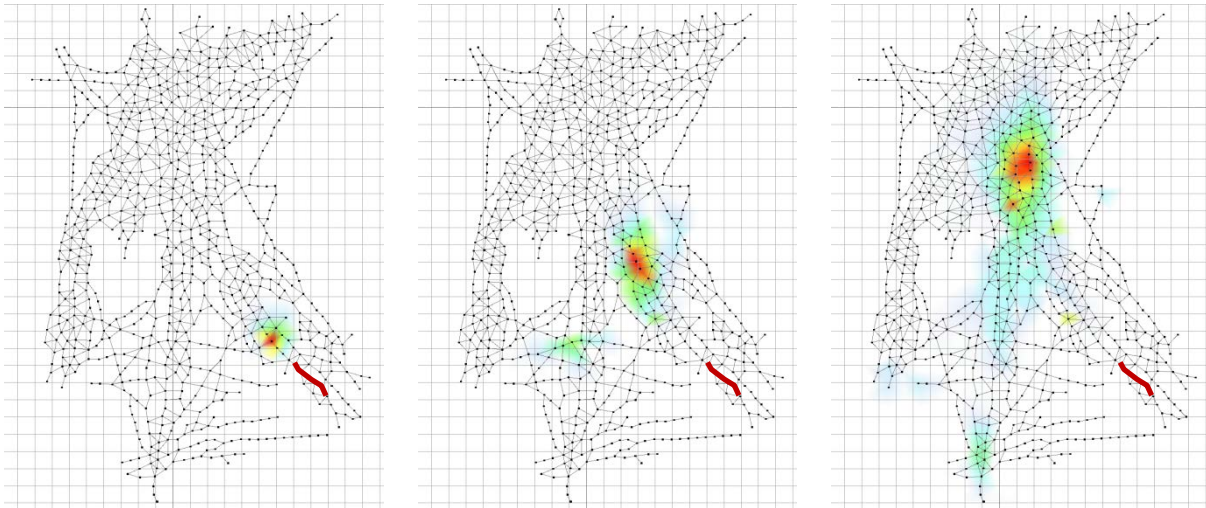


Abbildung 5: Prädiktion von Bewegungsspuren. Prädizierter Aufenthaltsort einer Person nach 1, 3 und 6 Sekunden (rot = hohe Wahrscheinlichkeit, blau = niedrige Wahrscheinlichkeit, ohne Einfärbung = unwahrscheinlich) bei Beobachtung der rot markierten Trajektorie (5 Knoten).

Die robuste Prädiktion der Fortsetzung von Bewegungsspuren auf Grundlagen von Statistiken zuvor beobachteter Trajektorien wurde anhand eines Demonstrators nachgewiesen (AP5.4). Die Arbeiten zur Prädiktion konnten großen Nutzen für die Einschränkung des Suchraums bei der Wiedererkennung von Personen (siehe AP3) erzielen. Ergebnisse des Projektpartners AVISTRA, bei denen der Passagiertyp anhand der Bewegungsspur festgestellt wurde, konnten ebenfalls integriert werden.

2.1.6 Arbeitspaket 8: Videoanalyse im Kontext von Gruppen

In AP8 wurden Arbeiten zur Detektion, dem Tracking, der Wiedererkennung und der Prädiktion von Bewegungsspuren von Einzelpersonen in Gruppen und Gruppen als Ganzes durchgeführt. Schwerpunkte dieses APs bildeten (1) die Verbesserung von Detektoren für die Detektion von Personen in Gruppen, (2) die Zuordnung von Einzelpersonen zu Gruppen, (3) die Wiedererkennung ganzer Gruppen und (4) die Prädiktion von Gruppenbewegungen.

1) Um Personen in Gruppen detektieren zu können, wurden zwei Ansätze verfolgt: Training von Detektoren mit künstlichen Daten und Anpassung der Parameter abhängig von der Gruppengröße. Das Ziel des Trainings mit künstlich erzeugten Daten ist, die Unzulänglichkeiten normaler Trainingsdatensätze für die Personendetektion zu beheben. Das Problem bei öffentlichen Datensätzen ist, dass häufig nur einzeln stehende Personen enthalten sind. Zusätzlich nehmen diese Personen viele Posen ein, die im Einsatzszenario eher ungewöhnlich sind. So sind an einem Flughafen Personen nicht auf Skiern oder Fahrrädern zu erwarten. Um diese Problem zu umgehen, wurde ein Szenario-spezifischer Datensatz erzeugt (AP8.1). Dafür wurden zunächst 3D-Modelle von Personen erstellt, mit verschiedener Kleidung ausgestattet und dann mit typischen Bewegungen animiert und aus verschiedenen Perspektiven gerendert. Damit lassen sich die 3D-Modelle der Personen mittels bekannten Kameraparametern (siehe AP3.1) in die leere Szene einfügen (siehe Abb. 6). Damit kann ein reichhaltiger Datensatz

erstellt werden, der das Einsatzszenario deutlich besser repräsentiert und dadurch zu besseren Personendetektionen führt.



Abbildung 6: Training von Detektoren mit künstlichen Daten. Links: Bilder aus INRIA-Personendatensatz, die nicht für ein Flughafenszenario geeignet sind. Mitte: 3D-modellierte Person mit Ground-Truth. Rechts: Person in die Szene eingefügt.

Die zweite Verbesserung zur Detektion von Personen in Gruppen wurde durch eine dynamische Anpassung der Parameter anhand der Gruppengröße erzielt (AP8.1). Die Gruppengröße wird dabei durch die Größe der segmentierte Vordergrundregion (siehe AP2.2) abgeschätzt (siehe Gleichung 1 und 2, sowie Abb. 7).

$$\text{Personenfläche} = \pi * \frac{\text{Personenhöhe}}{2} * \frac{\text{Personbreite}}{2} \quad (1)$$

$$\text{Anzahl der Personen} \approx \frac{\text{Fläche der Binärmaske}}{\text{Personenfläche}} \quad (2)$$

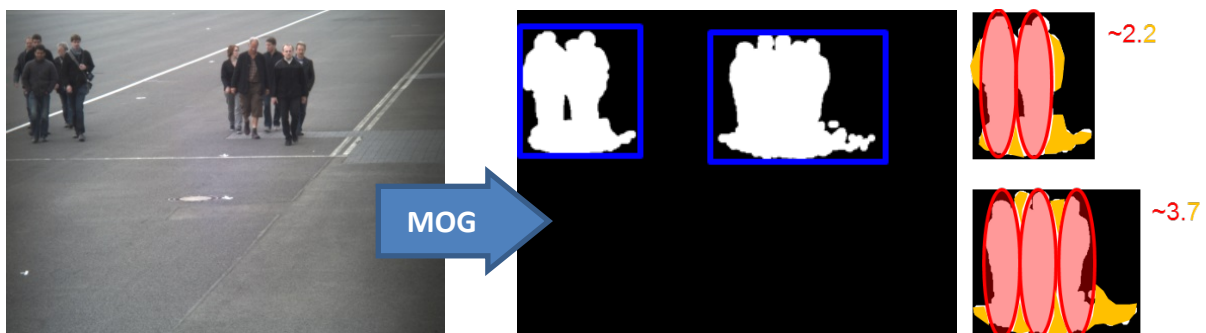


Abbildung 7: Abschätzung der Gruppengröße anhand der segmentierten Vordergrundregion.

Der Parameter zur Gruppierung von Personenhypothesen wird abhängig von der Größe der Gruppe gewählt. Hierbei konnten deutliche Verbesserungen erzielt werden. Experimente zur dynamischen Anpassung anderer Parameter brachten keine signifikanten Verbesserungen. Durch die beschriebenen Anpassungen konnten deutliche Verbesserungen bei der Detektion von Einzelpersonen in Gruppen erzielt werden.

2) Um Einzelpersonen Gruppen zuzuordnen wurden zwei komplementäre Ansätze verfolgt: Im ersten Ansatz wird die Zuordnung bildbasiert anhand des Bewegungsflusses vorgenommen (AP8.1). Im zweiten Ansatz werden Personen aufgrund der räumlichen Nähe im globalen Koordinatensystem gruppiert (AP8.1). Beim bildbasierten Ansatz werden die Bewegungen aller Personen im Bild mittels Optischen Fluss ermittelt. Anschließend werden Personen mit gleicher Bewegungsrichtung gruppiert. Hierzu kommt ein mehrstufiges Verfahren des Mean-Shift-Clusterings zum Einsatz. Das Verfahren ist sehr gut geeignet um zwei sich durchdringende Gruppen zu separieren und die entsprechenden Personendetektionen der korrekten Gruppe zuzuordnen (siehe Abb. 8). Das zweite Verfahren operiert nicht auf Bildern sondern auf Personenhypothesen, die durch bekannte Kameraparameter in ein globales Koordinatensystem projiziert werden. Für die Gruppierung wird das Mean-Shift-Clustering-



Abbildung 8: Segmentierung von Gruppen anhand der Bewegungsrichtung. Links: Segmentierung der Gruppen (vereinfachte Darstellung: Gruppierung wird eigentlich über Detektionen von Einzelpersonen anhand ähnlicher Bewegungs-Cluster vorgenommen). Rechts: Zu Grunde liegendes Clustering der Bewegungen (Gruppe 1: vorwiegend blau, grün dargestellte Bewegung nach links, Gruppe 2: cyan dargestellte Bewegung nach rechts, Gruppe 3: stehende Person; Bewegung orange dargestellt)

Verfahren mit unterschiedlichen Kernel-Breiten ausgeführt. Es wird weiterhin dem Operator überlassen, ab welcher Nähe zueinander mehrere Personen als Gruppe aufgefasst werden (siehe Abb. 9). Diese Design-Entscheidung wurde getroffen, da Voruntersuchungen des Unterauftragnehmers AVISTRA (siehe Anlage) ergeben hatten, dass es unterschiedliche Arten von Gruppen gibt (z.B. temporäre Gruppe, echte Gruppe, u.ä.), die jeweils über die Nähe von Personen zueinander definiert sind und sich die Nähe der Personen zueinander außerdem für schmale Gänge und große freie Flächen unterscheidet. Die Zuordnung von Personen zu Gruppen findet Anwendung in der Wiedererkennung von Gruppen (AP8.2), bei der Prognose von Gruppenbewegungen (AP8.3), bei der Detektion von Gruppenbildungs- und Auflösungsprozessen (AP9.1), sowie bei den Mengentheoretischen Analysen zu Gruppen.

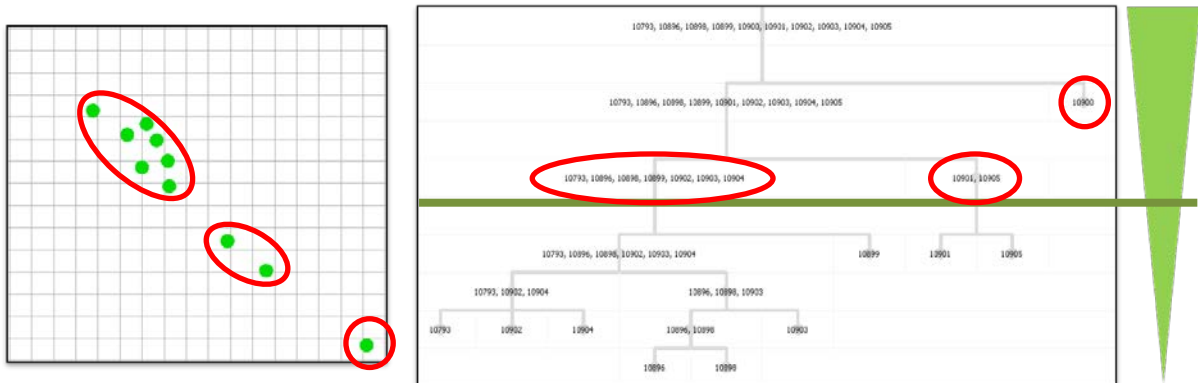


Abbildung 9: Gruppierung von Einzelpersonen anhand räumlicher Nähe. Der Gruppierungsparameter kann vom Operator frei gewählt werden. Die Gruppen ergeben sich entsprechend des Schnitts im Clustering-Baum.

3) Die Wiedererkennung von Gruppen (AP8.2) basiert auf der Wiedererkennung der in der Gruppe enthaltenen Personen. Beim Vergleich zweier Gruppen werden alle Personen aus einer Gruppe A mit allen Personen aus einer ggf. anderen Gruppe B verglichen. Dies entspricht dem menschlichen Vorgehen,



bei dem man versucht einzelne Personen der einen Gruppe in der anderen wiederzufinden (siehe Abb. 10).

Abbildung 10: Menschliches Vorgehen beim Wiedererkennen von Gruppen.

Für den Vergleich der Gruppen wird eine Matrix aufgestellt, welche die Scores (siehe AP3.2, AP3.5) aus den Vergleichen aller Personen untereinander enthält. Anschließend wird mittels Ungarischem Algorithmus die bestmögliche Zuordnung von Personen aus Gruppe A zu Personen aus Gruppe B berechnet. Die Ähnlichkeit zwischen den beiden Gruppen ergibt sich dann aus der Summe der Scores der berechneten Zuordnung (siehe Abb. 11). Sie ist höher, je mehr Personen der einen Gruppe in der anderen Gruppe wiedererkannt werden.

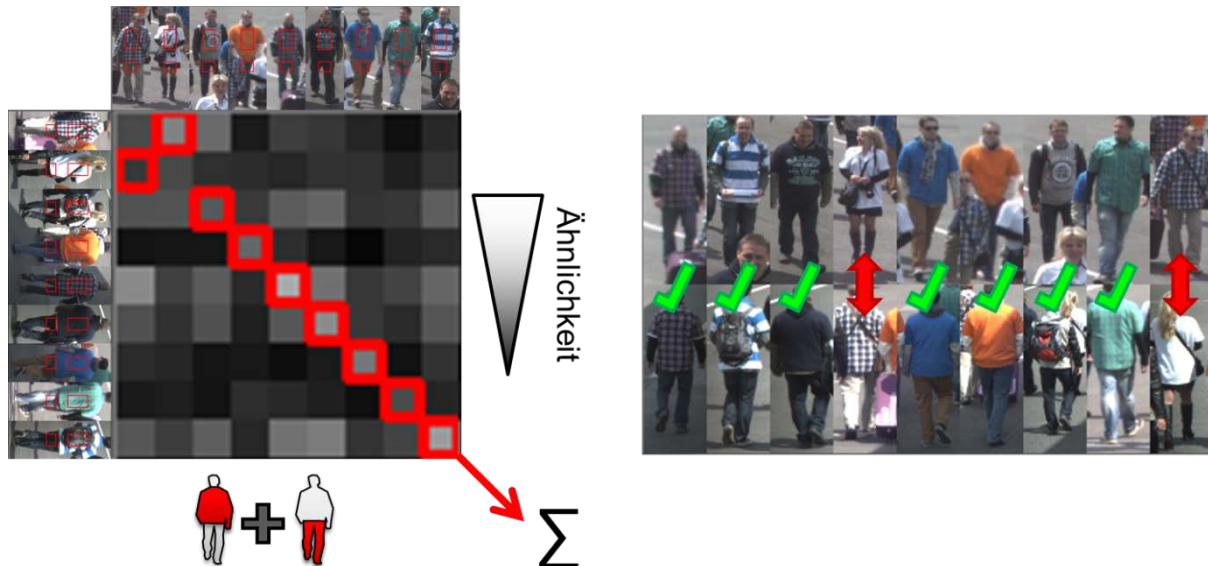


Abbildung 11: Prinzip der Wiedererkennung von Gruppen anhand der Ähnlichkeit der in der Gruppe enthaltenen Personen. Die Vorgehensweise orientiert sich am menschlichen Vorgehen (siehe Abb. 10).

4) Für die Prognose von Gruppenbewegungen (AP8.3) ließen sich die Verfahren zur Prädiktion der wahrscheinlichsten Fortsetzung von Bewegungsspuren einzelner Personen auf Gruppen übertragen. Es wurden jedoch zusätzlich Aspekte zur Laufgeschwindigkeit in Abhängigkeit der Gruppengröße berücksichtigt. Hierzu hatten Voruntersuchungen des Unterauftragnehmers AVISTRA (siehe Anlage) ergeben, dass sich die Laufgeschwindigkeit von Personen mit steigender Gruppengröße verringert. Durch Arbeiten aus AP8 konnten die Verfahren zur Detektion (AP2) und Wiedererkennung (AP3) von Einzelpersonen deutlich verbessert werden, indem der Kontext der Gruppe berücksichtigt wurde.

2.1.7 Arbeitspaket 9: Mengentheoretische Gruppenanalyse

Im Rahmen von AP9 wurde eine Wiedererkennung von Gruppen durch mengentheoretische Analyse realisiert. Dazu müssen (1) Gruppenbildungs- und Auflösungsprozesse detektiert und (2) Personen durch den Kontext benachbarter Personen, auch bei vollständiger Verdeckung, zu jeweils einer von vielen Gruppen zugeordnet werden.

1) Für die Detektion von Gruppenbildungs- und Auflösungsprozessen (AP9.1) werden die Ergebnisse aus der Gruppierung von Einzelpersonen auf Grund der räumlichen Nähe analysiert (Abb. 9). Sollten sich die in einem Cluster enthaltenen Personen in aufeinander folgenden Zeitschritten unterscheiden, so deuten dies auf einen Gruppenbildungs- oder Auflösungsprozess hin. Bei einem Gruppenbildungsprozess werden dabei zwei Cluster zu einem größeren Cluster vereinigt, dass anschließend alle Personen der beiden vorherigen Cluster enthält. Bei einem Gruppenauflösungsprozess zerfällt ein großes Cluster in mehrere kleinere Cluster.

2) Für die mengentheoretische Gruppenanalyse wurde zunächst in AP9.2 zusammen mit dem Projektpartner L-1 ein Konzept erstellt und anschließend in AP9.3 implementiert. Das Ziel der mengentheoretischen Betrachtungen lag in der Zuordnung von Gruppen einer Menge M_1 zum Zeitpunktes t_1 zu (neuen) Gruppen einer Menge M_2 zum Zeitpunktes t_2 . Dies sollte ermöglichen, Personen zu verfolgen, die zeitweise innerhalb einer Gruppe vollständig verdeckt waren, sich jedoch anhand der Grup-

pe verfolgen ließen (siehe Abb. 12). Die Zuordnung der Gruppen in Menge M_1 zu Gruppen der Menge M_2 wurde ähnlich wie in AP8.2 durch den ungarischen Algorithmus realisiert. Als Scores kamen die berechneten Ähnlichkeiten von Gruppen aus AP8.2 zum Einsatz.

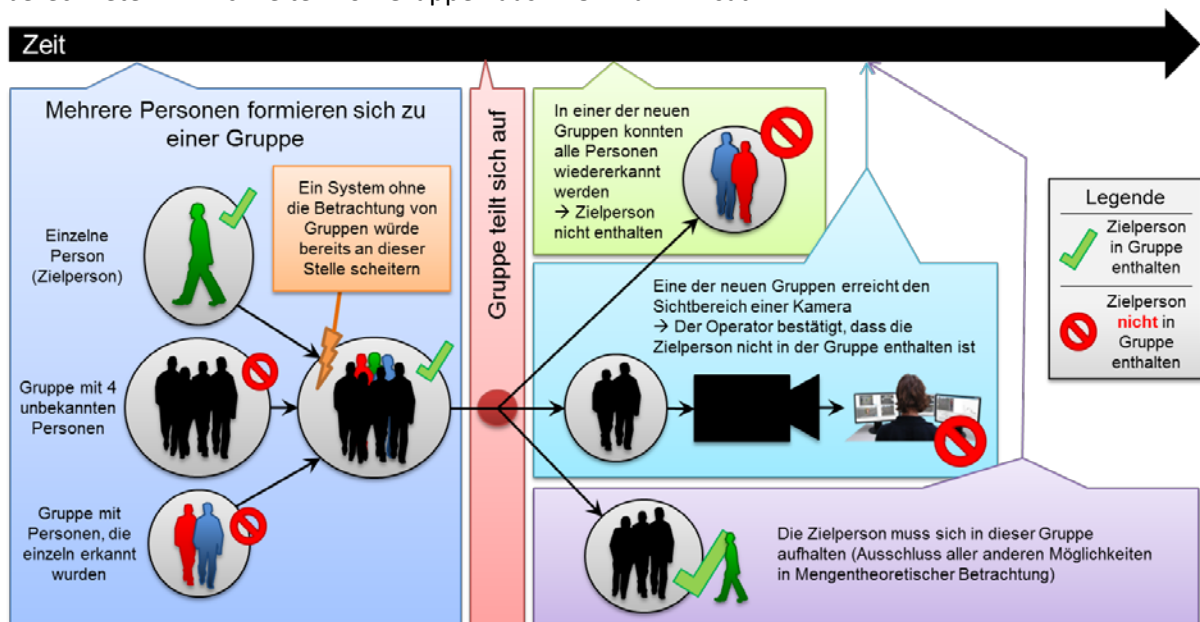


Abbildung 12: Beispiel zur mengentheoretischen Gruppenanalyse bei der eine Person trotz vollständiger Verdeckung anhand der Gruppe verfolgt werden kann.

Durch Arbeiten aus AP9 ist es möglich eine Rückwärts- oder Vorwärtsanalyse bezüglich Gruppen durchzuführen, wenn die Rückwärts- bzw. Vorwärtsanalyse für einzelne Personen der Gruppe scheitern würde.

2.1.8 Arbeitspaket 7: Implementierung, Test der Demonstratorstufen

Im Rahmen von AP7 wurden mehrere Demonstratorstufen implementiert: Zu M14 wurden Verfahren zur Live-Analyse (Personendetektion) auf aufgezeichneten Videodaten demonstriert. Dies lieferte den Nachweis, dass generelle Konzepte zur Datenablage in der Datenbank und der Kommunikation von Modulen untereinander tragen (siehe AP1, AP2).

Für den Meilenstein (M19) wurde eine Basisversion für das Gesamtsystem zur Rückwärtsanalyse erstellt. Die Funktionsfähigkeit aller Komponenten zur Live-Analyse wurde live am Flughafen Schönhagen demonstriert und in einem Mensch-Maschine-Interface (MMI) visualisiert. Die robuste Verfolgung von Einzelpersonen und Personen in losen Gruppen konnte gezeigt werden (siehe AP2). Jedoch traten Probleme beim Tracking von Personen in dichten Gruppen auf. Diese stellen ein gesondertes Problem dar und mussten daher später im Projekt separat betrachtet werden. Hierzu wurden die neuen Arbeitspakete 8 und 9 definiert. Außerdem konnten die Fortschritte bei Algorithmen zum Tracking und zur Wiedererkennung von Personen anhand von Offline-Demos aufgezeigt werden (siehe AP2, AP3).

Zu M26 wurden die fertigen Komponenten zur Live-Analyse demonstriert. Dabei konnten die robuste Funktionsweise des Trackers und der Fusion von Tracks zu Meta-Tracks gezeigt werden (siehe AP2). Des Weiteren wurden erste funktionierende Basisversionen zur Prädiktion (siehe AP5) und zur Wiedererkennung (siehe AP3) offline demonstriert. Zusätzlich wurde das fertiggestellte Multi-LRS-System mit automatischer Kalibrierung live demonstriert (siehe AP4).

Zu M33 wurde erstmalig das Gesamtsystem demonstriert, dass auf vorberechneten Live-Analyse-Ergebnissen auf Aufnahmen von Flughafen Schönhagen live die Rückwärtsanalyse berechnete (siehe AP2, AP3).

Zu M39 wurde das Gesamtsystem live anhand vier am Flughafen Erfurt-Weimar installierten Kameras demonstriert. Die robuste Funktionsfähigkeit und das gute Zusammenspiel aller Komponenten aller

Projektpartner bei Einzelpersonen und Personen in losen Gruppen konnte belegt werden (siehe APs 2-5). Schwierigkeiten gab es bei der Rückwärts- und Vorwärtsanalyse von Personen in dichten Gruppen. Lösungen dieser Probleme wurden zu M47 demonstriert. Die Strategie lag in der Berücksichtigung des Kontextes der Gruppe statt in der alleinigen Betrachtung von Einzelpersonen (siehe AP8, AP9).

In der Abschlussdemonstration zu M51 wurde die robuste Funktionsfähigkeit des Gesamtsystems bezüglich Einzelpersonen und Gruppen anhand von fünf praxisrelevanten Szenarien, die am Flughafen Erfurt-Weimar aufgenommen wurden, gezeigt (siehe APs 2-5, 8, 9). Das System ist frei bedienbar und belegt die Vorteile des Einsatzes von Algorithmen für die halbautomatische Auswertung von Videodaten für einen menschlichen Operator an einem Flughafen (Screenshot siehe Abb. 13). Der Abschlussdemonstrator wurde allen Projektpartnern zur Verfügung gestellt. Er wird auch dem Projektträger zur Verfügung gestellt (siehe beiliegender USB-Stick).

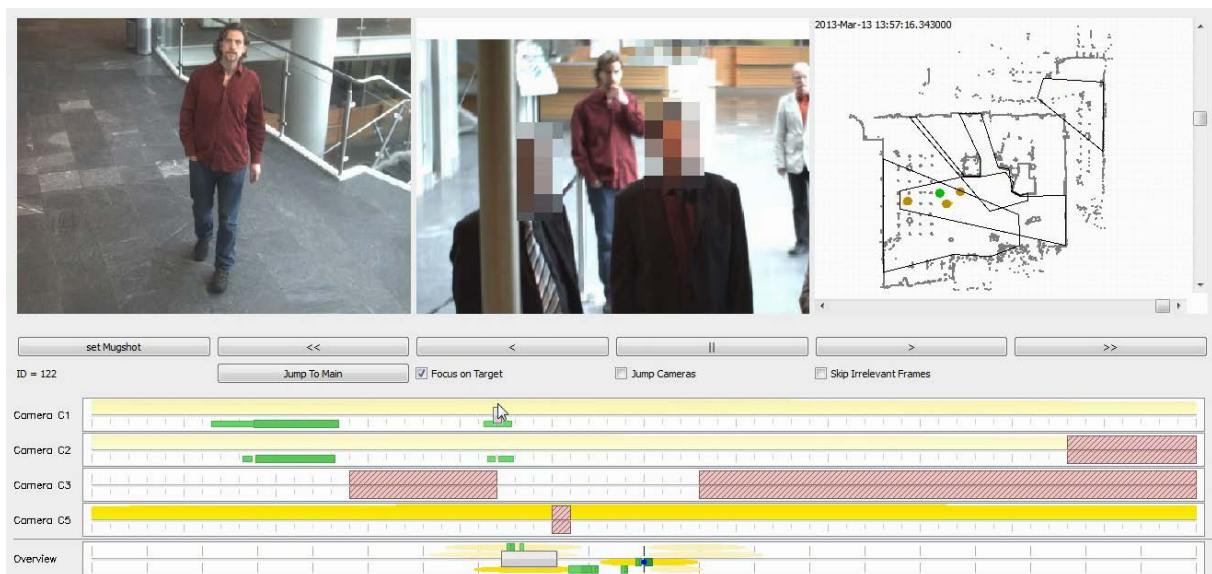


Abbildung 13: Screenshot des Abschlussdemonstrators. Oben links ist die ausgewählte Zielperson zu sehen. Die vom APFEL-System erstellten Analyse-Ergebnisse werden in einem Zeitstrahl visualisiert (unten). Gekennzeichnet sind Stellen im Video in denen das Auftreten der Person ausgeschlossen werden kann (rot), Prädiktionsergebnisse zu wahrscheinlichsten Aufenthaltsorten (gelb) und Zeitpunkte an denen die Person im Video wiedererkannt wurde (grün). Zusätzlich ist die Position der Zielperson zum ausgewählten Zeitpunkt (Videobild oben Mitte) in die Übersichtskarte des Flughafens eingetragen (grüner Punkt im Bild oben rechts).

2.2 Voraussichtlicher Nutzen und Verwertbarkeit der Ergebnisse

Die Ergebnisse aus APFEL konnten wie folgt wissenschaftlich verwertet werden: Zu den Themenschwerpunkten Lasertracking, visuelles Personentracking und Wiedererkennung von Personen lieferten sieben Bachelorarbeiten und zwei Masterarbeiten Beiträge. Die erzielten Ergebnisse in den Bereichen Lasertracking, visuelle Personendetektion, -Tracking und -Wiedererkennung wurden in sieben Konferenzbeiträgen publiziert und im Rahmen der Konferenzen (KI, AVSS, IROS, FWS) durch Poster oder Vorträge vorgestellt. Außerdem sind aktuell drei Dissertationen zu den Themen Prädiktion von Personentrajektorien, Personentracking und Wiedererkennung von Personen in Bearbeitung. Nutzen bringen die erzielten Ergebnisse in mehreren Bereichen: (1) Forschung, (2) Lehre und (3) Erwerb von Drittmitteln.

1) Die Verfahren zum schnellen Personentracking finden Einsatz im Projekt SERROGA (Service-Robotik für die Gesundheitsassistenten, 01.05.2012 - 30.04.2015, gefördert vom Thüringer Ministerium für Wirtschaft, Arbeit und Technologie). Die Verfahren zur Wiedererkennung von Personen werden im Projekt ROREAS eingesetzt (Interaktiver robotischer Reha-Assistent für das Lauf- und Orientierungstraining von Patienten nach Schlaganfällen, 01.07.2013 - 31.12.2015, gefördert vom Bundesmi-

nisterium für Bildung und Forschung). Außerdem wird das Muti-Laser-Range-Finder-System für das Benchmarking der beiden zuvor genannten Roboter-Systeme genutzt.

2) Die neu gewonnenen Erkenntnisse in den Bereichen Tracking und Wiedererkennung werden in die Lehrveranstaltungen des Fachgebiets Neuroinformatik und Kognitive Robotik einfließen. Durch die Einbindung aktueller Forschungsergebnisse kann die Qualität und Attraktivität der Lehre an der TU Ilmenau in den Bachelor- und Masterstudiengängen Informatik und Ingenieurinformatik verbessert werden. Außerdem konnte durch die gezielte Ausgabe von Bachelor- und Masterarbeiten im anspruchsvollen Themenfeld des Projekts eine deutliche Verbreiterung der Systemlösungskompetenz für die Studierenden erreicht werden.

3) Durch die Forschungsarbeiten im Rahmen des Projekts APFeI konnte das Fachgebiets Neuroinformatik und Kognitive Robotik seine Expertise in den Forschungsfeldern Personenerkennung und Personenverfolgung sichern und weiter ausbauen und kann damit weiterhin in dem sich verschärfenden nationalen und internationalen Wettbewerb um Forschungsgelder erfolgreich bestehen. Die im Vorhaben entwickelten informationstechnischen Verfahren der Mustererkennung sind aufgrund der Tatsache, dass diese auf einem Demonstrator in einem Einsatzfeld mit sehr anspruchsvollen Realweltbedingungen angewandt wurden, für eine Vielzahl weiterer Projekte verallgemeinerbar. Durch die sehr guten Ergebnisse steigert das Fachgebiets Neuroinformatik und Kognitive Robotik der TU Ilmenau seine Chancen beim Erwerb von Drittmitteln in den bearbeiteten und verwandten Themenbereichen.

2.3 Fortschritt auf dem Gebiet des Vorhabens bei anderen Stellen

Während des Vorhabens gab es einige Fortschritte bei anderen Stellen im Bereich der (1) Personendetektion und der (2) Wiedererkennung von Personen.

1) Während der Projektlaufzeit (01/2010 – 03/2014) gab es zwei Veröffentlichungen im Themenbereich Personendetektion, die eine deutliche Verbesserung gegenüber dem bisherigen Stand der Technik bezüglich Laufzeiteffizienz bei gleicher Qualität erzielen konnten [Wu 11], [Dollar 10]. Das Verfahren nach [Wu 11] wurde in das Projekt integriert. Zum Verfahren nach [Dollar 10] gab es Untersuchungen und erste Arbeiten zur Re-Implementierung. Ein guter Überblick zum aktuellen State-of-the-Art im Themenbereich Personendetektion ist in [Dollar 12] zu finden.

2) Im Bereich der Wiedererkennung von Personen gab es während der Projektlaufzeit zahlreiche Veröffentlichungen zu robusten Merkmalen zur Beschreibung von Personen und zum Lernen einer geeigneten Distanzmetrik für den Vergleich von Personen. Während des Projekts wurden Merkmale aus [Farenzena 10] im Rahmen einer Masterarbeit näher untersucht. Die Merkmale erwiesen sich, wie in der Literatur beschrieben, als sehr gut geeignet für die Wiedererkennung, jedoch verhinderte die schlechte zeitliche Performanz beim Matching (trotz deutlicher Optimierungen) eine Einbindung in das Projekt. Bezüglich des Erlernens einer geeigneten Distanzmetrik für den Vergleich von Personen erwies sich das Verfahren von An et al. [An 13] als nützlich. Das Verfahren wurde daher in das Projekt integriert. Ein guter Überblick zum aktuellen State-of-the-Art im Themenbereich Personen-Wiedererkennung ist in [Saghafi 14] und [Gong 14] zu finden.

In den Bereichen Lasertracking, Kalibrierung von Multi-LRS-Systemen, Personentracking und Score-Level-Fusion wurden während der Projektlaufzeit keine Verfahren in der Literatur gefunden, die unter Berücksichtigung der Anforderungen an das Einsatzszenario die Leistung der implementierten (und publizierten) Verfahren übertreffen konnten.

2.4 Erfolgte oder geplante Veröffentlichungen

Neben den Fortschritten bei anderen Stellen gab es auch zahlreiche eigene innovative Forschungsergebnisse in den Bereichen (1) Personendetektion, (2) Personentracking, (3) Personen-Wiedererkennung, (4) Lasertracking, (5) Trajektorien-Prädiktion und der (6) Zusammenführung zu einem integrierten unterstützenden System für die teilautomatisierte Videoüberwachung an Flughäfen. Nachfolgend werden Publikationen zu den genannten Bereichen aufgezählt.

- 1) Für die Verbesserung der Detektion von Personen wurden zwei Bachelorarbeiten [Sieder 10], [Meder 11] angefertigt, die sich mit der Bestimmung von Vordergrundbereichen beschäftigten um Falschdetektionen vermeiden zu können. In einer weiteren Bachelorarbeit [Morgenstern 12] wurden Optimierungen der Berechnungen für die Personendetektion beschrieben, wodurch das Verfahren echtzeitfähig gemacht werden konnte. Weitere Publikationen zum Personendetektion sind geplant.
- 2) Zum Thema des schnellen Personentrackings wurde eine Bachelorarbeit [Stolberg 11] angefertigt, die unterschiedliche Merkmale evaluierte. Das im Projekt eingesetzte Verfahren wurde in [Kolarow 12] publiziert und im Rahmen eines Vortrages auf der IROS 2012 vorgestellt. Es beschreibt das Tracking von Personen oder anderen Objekten mittels eines Sets ausgewählter Merkmale homogener Regionen, die sich besonders gut für eine logarithmische Suche eignen. Durch die eingesetzte Suchstrategie der logarithmischen Suche und einem spärlichen Merkmalsset kann ein Tracking von vier Personen gleichzeitig auf mindestens 300 Bildern pro Sekunde unabhängig von der Bildgröße durchgeführt werden. Das Verfahren erreicht auf öffentlichen Benchmarkdatensätzen etwa die gleiche Qualität wie die besten in der Literatur beschriebenen echtzeitfähigen Verfahren (teilweise leicht besser) und benötigt dafür nur ein Zehntel der Rechenzeit.
- 3) Im Bereich der Wiedererkennung von Personen wurden drei Bachelorarbeiten und eine Masterarbeit angefertigt. Zwei Bachelorarbeiten beschäftigen sich mit der Normierung von Farben um Beleuchtungsunterschiede in der Szene und zwischen Kameras auszugleichen [Scheiner 12], [Krause 13]. Die dritte Bachelorarbeit und die Masterarbeit behandeln geeignete Merkmale zur Beschreibung von Personen [Trinh 11], [Sorge 13]. Das im Projekt verwendete Verfahren zum Beleuchtungsausgleich ist in [Eisenbach 13] publiziert. Die verschiedenen Beleuchtungsbedingungen innerhalb einer Kamera werden anhand von beobachteten Personen, die sich durch die Szene bewegen, abgeschätzt. Durch das Lösen eines Optimierungsproblems kann auf die wahre Farbe der Kleidung von Personen geschlossen werden. Das im Projekt verwendete Verfahren für die Wiedererkennung von Personen ist in [Eisenbach 12] publiziert. Bei diesem Ansatz wird eine Personen-spezifische Auswahl geeigneter Merkmale getroffen. Für den Vergleich von zwei Personen wird die Mahalanobisdistanz verwendet. Dazu werden Merkmale des Ober- und Unterkörpers verwendet und die jeweiligen Ähnlichkeits-Scores werden mittels Score-Level-Fusion fusioniert. Weitere Publikationen im Bereich der Personen-Wiedererkennung sind geplant.
- 4) Zum Thema Personentracking mittels Lasern wurde eine Masterarbeit angefertigt [Schenk 11a]. Das resultierende robuste Verfahren, das auch im Projekt Verwendung findet, wurde in [Schenk 11b] publiziert. Dabei werden zwei Varianten vorgestellt: Das Tracking mit Lasern auf Fußhöhe und das Tracking mit Lasern auf Oberkörperhöhe. Letztere Variante erzielt die besseren Ergebnisse. Das im Projekt eingesetzte Verfahren zur automatischen Kalibrierung eines Multi-Laser-Range-Finder-Systems wurde in [Schenk 12a] und [Schenk 12b] publiziert. Der entwickelte Algorithmus verwendet Trajektorien von Personen, die sich durch die Szene bewegen und benutzt gemeinsame Beobachtungen einer Trajektorie verschiedener Laser um diese relativ zueinander auszurichten.
- 5) Im Bereich der Prädiktion der Fortsetzung von Personenbewegungsspuren wurden bisher noch keine Ergebnisse publiziert. Eine Publikation wird derzeit vorbereitet.
- 6) Das während des Projekts APFEL implementierte Gesamtsystem zur vorwärts und rückwärts gerichteten Personensuche in Videodatenströmen wurde in [Kolarow 13] publiziert. Es werden alle Komponenten zur Detektion, dem Tracking und der Gesicht- sowie Kleidungs-basierten Wiedererkennung von Personen beschrieben. Des Weiteren werden die umgesetzten Verfahren zur Fusion beschrieben. Abschließend werden Untersuchungen beschrieben, die den Nutzen des Systems für einen menschlichen Operator im Bereich der Videoüberwachung an Flughäfen belegen.

3 Literatur

Eigene Publikationen im Rahmen des Projektes APFEL

- [Eisenbach 12] Eisenbach, M., Kolarow, A., Schenk, K., Debes, K., Gross, H.-M.: View Invariant Appearance-based Person Reidentification Using Fast Online Feature Selection and Score Level Fusion. In: Proc. 9th IEEE Int. Conf. on Advanced Video and Signal-Based Surveillance (AVSS 2012), Beijing, China, pp. 184-190, IEEE 2012.
- [Eisenbach 13] Eisenbach, M., Scheiner, P., Kolarow, A., Schenk, K., Gross, H.-M., Weinreich, I.: Learning Illumination Maps for Color Constancy in Person Reidentification. In: Proc. 19. Workshop Farbbildverarbeitung (FWS13), Berlin, Germany, pp. 103-114, GFAI 2013.
- [Kolarow 12] Kolarow, A., Brauckmann, M., Eisenbach, M., Schenk, K., Einhorn, E., Debes, K., Gross, H.-M.: Vision-based Hyper-Real-Time Object Tracker for Human-Robot Applications. In: Proc. IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS 2012), Vilamoura, Portugal, pp. 2108-2115, IEEE 2012.
- [Kolarow 13] Kolarow, A., Schenk, K., Eisenbach, M., Dose, M., Brauckmann, M., Debes, K., Gross, H.-M.: APFEL: The Intelligent Video Analysis and Surveillance System for Assisting Human Operators. In: Proc. 10th IEEE Int. Conf. on Advanced Video and Signal-Based Surveillance (AVSS 2013), Krakow, Poland, pp. 195-201, IEEE 2013
- [Krause 13] Krause, K.: Normalisierung von Farben in einem Multi-Kamera-System zur Wiedererkennung von Personen, Bachelorarbeit, TU Ilmenau, 2013.
- [Meder 11] Meder, J.: Neuronale RBF-Netze zur Vordergrund-Hintergrund-Segmentierung, Bachelorarbeit, TU Ilmenau, 2011.
- [Morgenstern 12] Morgenstern, W.: Berechnung von HOG-Features auf der GPU zum echtzeitfähigen Personentracking mit Kameras, Bachelorarbeit, TU Ilmenau, 2012.
- [Scheiner 12] Scheiner, P.: Berechnung von Beleuchtungskarten für Statische Kamerakonfigurationen, Bachelorarbeit, TU Ilmenau, Hochschule Koblenz, 2012.
- [Schenk 11a] Schenk, K.: Laserbasiertes Verfahren zur vollautomatischen Erfassung und Protokollierung von Personenbewegungstrajektorien, Masterarbeit, TU Ilmenau, 2011.
- [Schenk 11b] Schenk, K., Eisenbach, M., Kolarow, A., Gross, H.-M.: Comparison of Laser-based Person Tracking at Feet and Upper-Body Height. In: Proc. 34th Annual Conference on Artificial Intelligence (KI 2011), Berlin, Germany, LNAI 7006, pp. 277-288, Springer 2011.
- [Schenk 12a] Schenk, K., Kolarow, A., Eisenbach, M., Debes, K., Gross, H.-M.: Automatic Calibration of Multiple Stationary Laser Range Finders using Trajectories. In: Proc. 9th IEEE Int. Conf. on Advanced Video and Signal-Based Surveillance (AVSS 2012), Beijing, China, pp. 306-312, IEEE 2012.
- [Schenk 12b] Schenk, K., Kolarow, A., Eisenbach, M., Debes, K., Gross, H.-M.: Automatic Calibration of a Stationary Network of Laser Range Finders by Matching Movement Trajectories. In: Proc. IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS 2012), Vilamoura, Portugal, pp. 431-437, IEEE 2012.
- [Sieder 10] Sieder, R.: Berechnung der Vordergrund-Hintergrund-Segmentierung auf der GPU zum echtzeitfähigen Personentracking mit Kamera, Bachelorarbeit, TU Ilmenau, 2010.
- [Sorge 13] Sorge, S.: Wiedererkennung von Personen durch symmetrietriebene Extraktion von Merkmalen, Masterarbeit, TU Ilmenau, 2013.
- [Stolberg 11] Stolberg, S.: Farb-, form- und texturbasierte Features zum Tracking von Personen, Bachelorarbeit, TU Ilmenau, 2011.
- [Trinh 11] Trinh, T.Q.: Berechnung von SURF-Features auf der GPU zum echtzeitfähigen Personentracking mit Kameras, Bachelorarbeit, TU Ilmenau, 2011.

Publikationen des Fachgebietes

- [Corradini 00] Corradini, A., Boehme, H.-J., Gross, H.-M. A Hybrid Stochastic-Connectionist Approach to Gesture Recognition. *International Journal on Artificial Intelligence Tools*, 9 (2000), no. 2, pp. 177-204, 2000.
- [Gross 06] Gross, H.-M., Richarz, J., Müller, St., Scheidig, A., Martin, Chr. Probabilistic Multi-modal People Tracker and Monocular Pointing Pose Estimator for Visual Instruction of Mobile Robot Assistants. in: Proc. 2006 IEEE World Congr. on Computational Intelligence (WCCI), Intern. Joint Conf. on Neural Networks, Vancouver, Canada, pp. 8325-8333, IEEE Omnipress, 2006.
- [Gross 08] Gross, H.-M., Böhme, H.-J., Schröter, Ch., Müller, St., König, A., Martin, Ch., Merten, M., Bley, A. ShopBot: Progress in Developing an Interactive Mobile Shopping Assistant for Every-day Use. in: Proc. IEEE Int. Conf. on Systems, Man and Cybernetics (SMC), Singapore, pp. 3471-3478, IEEE 2008.
- [Hellbach 08] Hellbach, S., Strauss, S., Eggert, J., Körner, E., Gross, H.-M. Echo State Networks for Online Prediction of Movement Data-Comparing Investigations. in: Proc. Int. Conf. on Artificial Neural Networks (ICANN), Prague, LNCS 5163, pp. 710-719, Springer Verlag, 2008.
- [Hellbach 09] Hellbach, S., Eggert, J., Koerner, E., Gross, H.-M. Prediction and Classification of Motion Trajectories using Spatio-Temporal NMF. to appear: Proc. 32nd Annual Conference on Artificial Intelligence, Paderborn, LNCS, Springer Verlag, 2009.
- [Martin 06] Martin, Chr., Schaffernicht, E., Scheidig, A., Gross, H.-M. Sensor Fusion using a Probabilistic Aggregation Scheme for People Detection and People Tracking. *Robotics and Autonomous Systems*, vol. 54, (2006)9, pp. 721-728, Elsevier ScienceDirect, 2006.
- [Martin 08] Martin, Ch., Gross, H.-M. A Real-time Facial Expression Recognition System based on Active Appearance Models using Gray Images and Edge Images. in: Proc. of the 8th IEEE Int. Conf. on Face and Gesture Recognition (FG'08), paper no. 299, 6 pages, IEEE, 2008.
- [Martin 09] Martin, Ch., Steege, F.-F., Gross, H.-M. Estimation of Pointing Poses for Visual Instructing Mobile Robots under Real-World Conditions. in: *Robotics and Autonomous Systems*, 2009.
- [Müller 07] Müller, St., Schaffernicht, E., Scheidig, A., Böhme, H.-J., Gross, H.-M. Are you still following me? in: Proc. 3rd European Conference on Mobile Robots (ECMR), Freiburg, pp. 211-216, Albert-Ludwigs-Universität Freiburg - Universitätsverlag 2007.
- [Müller 08] Müller, St., Hellbach, S., Schaffernicht, E., Ober, A., Scheidig, A., Gross, H.-M. Whom to talk to? Estimating user interest from movement trajectories. in: Proc. 17th IEEE Int. Symposium on Robot and Human Interactive Communication, (ROMAN), Munich, pp. 532-538, IEEE Omnipress, 2008.

- [Richarz 06] Richarz, J., Martin, Chr., Scheidig, A., Gross, H.-M. There You Go! - Estimation Point-ing Gestures in Monocular Images for Mobile Robot Instruction. in: Proc. ROMAN - 15th IEEE Int. Symp. on Robot and Human Interactive Communication, 2006, pp. 546-551, IEEE, 2006
- [Richarz 07] Richarz, J., Scheidig, A., Martin, Chr., Müller, St., Gross, H.-M. A Monocular Pointing Pose Estimator for Gestural Instruction of a Mobile Robot. International Journal of Advanced Ro-botic Systems, Vol. 4, No. 1, pp. 139-150, 2007.
- [Schaffernicht 05] Schaffernicht, E., Martin, Ch., Scheidig, A., Gross, H.-M. A Probabilistic Multi-modal Sensor Aggregation Scheme Applied for a Mobile Robot. in: Proc. 28th German Conference on Artificial Intelligence, Koblenz, 2005, Lecture Notes in Artificial Intelligence (LNAI) 3698, pp. 320-334, Springer Verlag, 2005.
- [Scheidig 06] Scheidig, A., Müller, St., Martin, Chr., Gross, H.-M. Generating Person's Movement Trajectories on a Mobile Robot. in: Proc. ROMAN - 15th IEEE Int. Symposium on Robot and Human Interactive Communication, Hatfield (UK), pp. 747-752, IEEE, 2006.
- [Stricker 09] Stricker, R., Martin, Ch., Gross, H.-M. Increasing the Robustness of 2D Active Appearance Models for Real-World Applications. to appear Proc. Int. Conf. on Computer Vision Systems (ICVS), Liege, 2009.
- [Wilhelm 04] Wilhelm, T., Boehme, H.-J., Gross, H.-M.: A Multi-modal System for Tracking and Analyzing Faces on a Mobile Robot. Robotics and Autonomous Systems, 48, pp. 31-40, 2004.
- [Wilhelm 05] Wilhelm, T., Boehme, H.-J., Gross, H.-M.: Classification of Face Images for Gender, Age, Facial Expression, and Identity. Proc. Int. Conf. on Artificial Neural Networks (ICANN), LNCS 3696, pp. 569-574, Springer Verlag, 2005.

Fremde Publikationen

- [An 13] L. An, M. Kafai, S. Yang, B. Bhanu: Reference-Based Person Re-Identification. In: AVSS, pp. 244–249, 2013.
- [Andriluka 09] Andriluka, M., Roth, S., Schiele, B. Pictorial Structures Revisited: People Detection and Articulated Pose Estimation. In: CVPR, pp. 1014-1021, 2009.
- [Arras 07] Arras, K.O., Mozos, O.M., Burgard, W. Using Boosted Features for the Detection of People in 2D Range Data. In: IROS, pp. 3402 – 3407, 2007.
- [Arsic 08] Arsic, D., Hristov, E., Lehment, N., Hornler, B., Schuller, B., Rigoll, G. Applying multi layer homography for multi camera person tracking. In: ICSDC, pp. 1-9, 2008.
- [Baker 01] Baker, S., Matthews, I. Equivalence and efficiency of image alignment algorithms. In: CVPR, pp. 1090-1097, 2001.
- [Bazzani 12a] Bazzani, L., Cristani, M., Murino, V. Decentralized Particle Filter for Joint Individual-Group Tracking. In: CVPR, pp. 1886-1893, 2012.
- [Bazzani 12b] Bazzani, L., Cristani, M., Paggetti, G., Tosato, D., Menegaz, G., Murino, V. Analyzing groups: a social signaling perspective. In: VABI, pp. 271-305, 2012.
- [Belz 08] Belz, S. Optimierung eines Personentrackers auf Basis von Konturen zum Einsatz auf einem mobilen Robotersystem. Diplomarbeit, TU Ilmenau, 2008
- [Chang 11] Chang, M.C., Krahnstoever, N., Ge, W. Probabilistic group-level motion analysis and scenario recognition. In: ICCV, pp. 747-754, 2011
- [Cootes 92] Cootes, T. F., Taylor, C. J., Cooper, D. H., Graham, J. Training models of shape from sets of examples. In: BMVC, pp. 9-18, 1992.
- [Dalal 05] N. Dalal, B. Triggs: Histograms of oriented gradients for human detection. In: CVPR, pp. 886–893, 2005.
- [Dollar 10] P. Dollar, S. Belongie, P. Perona: The Fastest Pedestrian Detector in the West. In: BMVC, 2010.
- [Dollar 12] P. Dollar, C. Wojek, B. Schiele, P. Perona: Pedestrian Detection: An Evaluation of the State of the Art. TPAMI, Vol. 34, Is. 4, pp. 743–761, 2012.
- [Farenzena 10] M. Farenzena, L. Bazzani, A. Perina, V. Murino, M. Cristani: Person reidentification by symmetry-driven accumulation of local features. In: CVPR, pp. 2360–2367, 2010.
- [Ferrari 08] Ferrari, V., Marin-Jimenez, M., Zisserman, A. Progressive search space reduction for human pose estimation. In: CVPR, pp. 1-8, 2008.
- [Gavrila 99] Gavrila, D.M. Real-time object detection for “smart” vehicles. In: ICCV, pp. 87-93, 1999.
- [Ge 12] Ge, W., Collins, R.T., Ruback, R.B. Vision-based analysis of small groups in pedestrian crowds. TPAMI, pp. 1003-1016, 2012.
- [Gidel 08] Gidel, S., Checchin, P., Blanc, C., Chateau, T., Trassoudaine, L. Pedestrian detection method using a multilayer laserscanner: Application in urban environment. In: IROS, pp. 173 – 178, 2008.
- [Gong 14] S. Gong, M. Cristani, S. Yan, C.C. Loy: Person Re-Identification. Advances in Computer Vision and Pattern Recognition, Springer, 2014.
- [Hahn 08] Hahn, M., Krüger, L., Wöhler, C. 3D Action Recognition and Long-Term Prediction of Human Motion. In: ICVS, 2008.
- [Han 04] Han, B., Comaniciu, D., Davis, L. Sequential kernel density approximation through mode propagation: applications to background modelling. In: ACCV, 2004.
- [Junejo 07] Junejo, I.N., Foroosh, H.: Trajectory Rectification and Path Modeling for Video Surveillance. In: ICCV, 2007.
- [Kass 88] Kass, M., Witkin, A. P., Terzopoulos, D. Snakes: Active contour models. IJCV, Vol. 1, (1988) 4, pp. 321-331, 1988.
- [Khan 06] Khan, S.-M., Shah, M. A Multiview Approach to Tracking People In Crowded Scenes Using A Planar Homography Constraint. In: ECCV, pp. 133-146, 2006.
- [Kurazume 08] Kurazume, R., Yamada, H., Murakami, K., Iwashita, Y., Hasegawa, T. Target track-ing using SIR and MCMC particle filters by multiple cameras and laser range finders. In: IROS, pp. 3838-3844, 2008.
- [Lau 09] Lau, B., Arras, K. O., Burgard, W. Tracking Groups of People with a Multi-Model Hypothesis Tracker. In: ICRA, 2009.
- [Lee 03] Lee, K.K., Yu, M., Xu, Y. Modeling of human walking trajectories for surveillance. In: IROS, 2003.
- [Leibe 08] Leibe, B., Leonardis, A., Schiele, B. Robust Object Detection with Interleaved Categorization and Segmentation. IJCV, Vol. 77, 1-3, pp. 259-289, 2008.
- [Murakami 00] Murakami, S. Wada, A. An automatic extraction and display method of walking persons' trajectory. In: ICPR, 2000.
- [Rhodes07] Rhodes, B. J., Bomberger, N. A., Zandipour, M., Waxman, A. M., Seibert, M. Cognitively-Inspired Motion Pattern Learning & Analysis Algorithms for Higher-Level Fusion and Auto-mated Scene Understanding. In: MCC, 2007.
- [Rigoll 00] Rigoll, G., Eickeler, S., Muller, S. Person tracking in real-world scenarios using statistical methods. In: ICAFG, pp. 342-347, 2000.
- [Saghafi 14] M.A. Saghafi, A. Hussain, H.B. Zaman, M.H.M. Saad: Review of person re-identification techniques. IETCV, 2014.
- [Shao 08] Shao, X., Katabira, K., Shibasaki, R., Zhao H., Nakagawa, Y. Tracking a variable number of pedestrians in crowded scenes by using laser range scanners. In: SMC, pp. 1545 – 1551, 2008.
- [Spinello 08a] Spinello, L., Triebel, R., Siegwart, R. Multimodal detection and tracking of pedestrians in urban environments with explicit ground plane extraction. In: IROS, pp. 1823 – 1829, 2008.

- [Spinello 08b] Spinello, L., Siegwart, R. Human Detection using Multimodal and Multidimensional Features. In: ICRA, pp. 3264 – 3269, 2008.
- [Stauffer 99] Stauffer, C., Grimson, W.E.L. Adaptive Background Mixture Models for Real-Time Tracking, In: CVPR, pp.2246, 1999.
- [Suzuki 07] Suzuki, N., Hirasawa, K., Tanaka, K., Kobayashi. Y. Learning motion patterns and anomaly detection by Human trajectory analysis. In: SMC, 2007.
- [Treptow 05] Treptow, A., Cielniak, G., Duckett, T. Comparing Measurement Models for Tracking People in Thermal Images on a Mobile Robot. In: ECOMR, 2005.
- [Vasquez 04] Vasquez, D. Fraichard, T. Motion prediction for moving objects: a statistical approach. In: ICRA, 2004.
- [Viola 01] Viola, P., Jones, M. Robust Real-time Object Detection. IJCV, 2001.
- [Viola 03] Viola, P., Jones, M., Snow, D. Detecting Pedestrians Using Patterns of Motion and Appearance. In: ICCV, pp. 734-741, 2003.
- [Weiler 09] Weiler, D., Eggert, J. Level-Set Segmentation with Contour based Object Representation. In: IJCNN, 2009.
- [Weser 06] Weser, M. Westhoff, D., Huser, M., Zhang, J. Multimodal People Tracking and Trajectory Prediction based on Learned Generalized Motion Patterns. In: ICMSFIS, pp. 541 – 546, 2006.
- [Wu 11] J. Wu, C. Geyer, J. Rehe: Real-time human detection using contour cues. In: ICRA, pp. 860–867, 2011.
- [Zajdel 05] Zajdel, W., Krose, BJA. A sequential Bayesian algorithm for surveillance with nonover-lapping cameras, IJPRAI, vol.19, nr. 8, pp.977–996, 2005
- [Zanotto 12] Zanotto, M., Bazzani, L., Cristani, M., Murino, V. Online Bayesian Non-parametrics for Social Group Detection. In: BMVC, pp. 111.1-111.12, 2012.
- [Zhu 06] Zhu, Q.; Avidan, S.; Yeh, M. C. & Cheng, K. T. Fast Human Detection Using a Cascade of Histograms of Oriented Gradients. In: CVPR, pp. 1491–1498, 2006.

Anhang:

Beratung zur Analyse von Gruppenverhalten an Flughäfen.
Bericht zum Unterauftrag der TU Ilmenau. Avistra GmbH, Berlin, März 2014

Abschlussbericht

Beratung zur Analyse von Gruppenverhalten an Flughäfen

Im Rahmen des
Aufstockungsantrags zum
Verbundprojekt APFeI+

20. März 2014



Inhalt

Inhalt	1
Abbildungsverzeichnis.....	2
KAPITEL 1 Hintergrund	3
KAPITEL 2 Beschreibung von Gruppenmerkmalen	4
2.1 Zeitliche Stabilität von Gruppen	4
2.2 Räumliche Distanz innerhalb von Gruppen.....	8
KAPITEL 3 Nutzung von Gruppenmerkmalen.....	10
3.1 Nutzung der Gruppengröße für die Abschätzung der Laufgeschwindigkeit	10
3.2 Nutzung der Gruppenzugehörigkeit für die Wiedererkennung.....	12

Abbildungsverzeichnis

Abbildung 2-1:	Menschenansammlung im Flughafenterminal	4
Abbildung 2-2:	Bahnen- (a) und Clusterbildung (b) sowie "Stau" vor Durchgängen	5
Abbildung 2-3:	Warteschlange am Flughafen	6
Abbildung 2-4:	Gruppengröße am Flughafen Dresden.....	7
Abbildung 2-5:	Gruppengröße am Flughafen Stuttgart.....	7
Abbildung 2-6:	Zeitliche Stabilität verschiedener Gruppen	8
Abbildung 2-7:	Distanz-Zonen.....	9
Abbildung 3-1:	Laufgeschwindigkeiten in Abhängigkeit der Gruppengröße am Flughafen Dresden.....	10
Abbildung 3-2:	Verteilungsfunktion der Laufgeschwindigkeit in Abhängigkeit der Gruppengröße am Flughafen Dresden	11
Abbildung 3-3:	Laufgeschwindigkeit in Abhängigkeit der Gruppengröße auf der EXPRO-Hannover.....	11
Abbildung 3-4:	Ablaufschema der Wiedererkennung anhand von Gruppenzugehörigkeiten.....	12

KAPITEL 1 Hintergrund

In Anbetracht steigender organisierter Kriminalität und anhaltender Terrorismusgefahr stellen kritische Infrastrukturen sowie öffentlich zugängliche Bereiche mit hohem Personenaufkommen, wie zum Beispiel Flughäfen, Bahnhöfe oder Stadien, besonders schützenswerte Bereiche mit einem hohem Sicherheitsbedarf dar. Die verstärkte Überwachung dieser entsprechenden Infrastrukturen mit Hilfe von Videokameras zur frühzeitigen Erkennung möglicher Gefahrenpotenziale erfolgt dabei typischerweise zentral in einem Leitstand von Fachpersonal, welches zumeist ohne weitere technische Unterstützung das Bildmaterial auf Auffälligkeiten analysiert.

Im Verbundprojekt APFeI (1/2010-3/2013) wurden u.a. von Avistra grundlegende konzeptionelle, methodisch-technische und experimentelle Arbeiten zur kameraübergreifenden Personensuche und zur Ermittlung des zurückgelegten und des intendierten Laufweges in realen Flughafenszenarien in Hyper-Echtzeit durchgeführt. Im Rahmen dieses Vorhabens konnte erfolgreich ein funktionsfähiger Demonstrator entwickelt werden, der in der Lage ist, eine durch einen Operator als „relevant“ markierte Personen mit bis zu 100-facher Beschleunigung im Videodatenstrom zu verfolgen. Vor dem Hintergrund, dass in APFeI zunächst die prinzipielle Machbarkeit einer solchen schnellen Suche im Fokus stand, erfolgte die Entwicklung des Demonstrators unter erleichternden Randbedingungen (z.B. aufrecht stehende oder laufende Personen, Bewegung höchstens in kleinen, aufgelockerten Gruppen). Für einen Einsatz unter realen Bedingungen eines Flughafens wäre dieser Demonstrator allerdings noch nicht geeignet, da er durch die notwendige Fokussierung auf Einzelpersonen oder durchlässigen bzw. losen Personengruppen im realen Betrieb immer dann an Grenzen stößt, wenn die zu verfolgende Person in einer dichtgedrängten Personengruppe verschwindet.

Auf Basis der gesammelten Erfahrungen aus dem Projekt APFeI beschäftigte sich das Folgevorhaben APFeI+ mit der Verbesserung der Leistungsfähigkeit des Demonstrators, um einen praxistauglichen Betrieb gewährleisten zu können. Speziell das Verhalten und die Dynamik einer Personengruppe lagen dabei im Fokus der Arbeiten von Avistra, um die hohen Personenzahlen an Flughäfen effektiv erfassen und verarbeiten zu können. Dadurch sollen auch Bewegungen einzelner Personen bspw. durch eine Gruppe hindurch oder mit einer Gruppe erfasst werden können, um so eine durchgängige Personenspur zu erhalten. Hierbei spielt insbesondere die Thematik des Verhaltens einer Gruppe in einer spezifischen Umgebung, wie z.B. einem Flughafen, eine besondere Rolle.

KAPITEL 2 Beschreibung von Gruppenmerkmalen

2.1 Zeitliche Stabilität von Gruppen

Um im Rahmen der Videoüberwachung von öffentlichen Räumen gruppenspezifische Phänomene für die Personenverfolgung nutzbar machen zu können, gilt es zunächst, eine Gruppe anhand bestimmter Merkmale als solche zu identifizieren.

Gerade in Situationen, in denen sich viele Menschen gleichzeitig im zu überwachenden Raum befinden, ist die Zuordnung einzelner Personen zu Gruppen schwierig.



Abbildung 2-1: Menschenansammlung im Flughafenterminal¹

Um dennoch Aussagen über eine Gruppenzugehörigkeit einzelner Personen machen zu können, ist zunächst grundsätzlich zwischen Personen zu unterscheiden, die ruhig stehen und solchen, die sich in Bewegung befinden. Daher wird für die weiteren Betrachtungen davon ausgegangen, dass die Erkennung nicht anhand einzelner Bilder sondern auf der Grundlage einer längeren Videosequenz erfolgen muss.

Da das bisherige APFEL-System bereits in der Lage ist, die Spuren einzelner Personen aufzuzeichnen, kann man diese Informationen nutzen, um gemeinsame Bewegungen mehrerer Personen zu identifizieren, oder um die räumliche Nähe mehrerer Personen über einen längeren Zeitraum zu erkennen, auch ohne dass eine wesentliche Bewegung stattfindet.

Anhand von Beobachtungen an Flughäfen konnte festgestellt werden, dass sich grundsätzlich drei verschiedene Arten von Gruppen unterscheiden lassen:

1. „Menschenansammlung“
2. "Temporäre Gruppe"
3. "Echte Gruppe"

Die erste Kategorie **Menschenansammlung** beschreibt jegliches gemeinsames Auftreten von Personen in einem definierten Raum, ohne dass diese Personen in

¹ Quelle: <http://www.fr-online.de/image/view/2012/7/31/17016784,14326860,highRes,14823E00337B922C.jpg>

einem Zusammenhang stehen. Typische Beispiele hierfür sind die Bahnen- und Clusterbildung, bei denen sich Personen über einen relativ kurzen Zeitraum gruppieren. Die Abgrenzung des untersuchten Raumes ist dabei beliebig, beschränkt sich aber typischerweise auf wenige Quadratmeter.

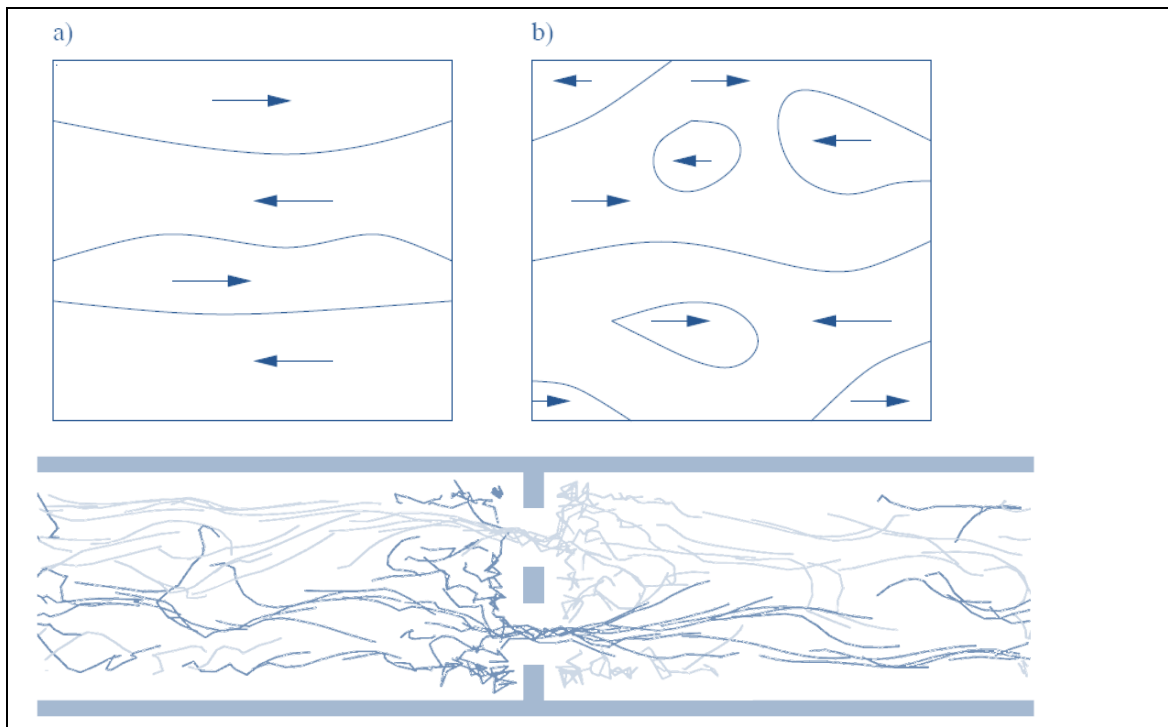


Abbildung 2-2: Bahnen- (a) und Clusterbildung (b) sowie "Stau" vor Durchgängen²

Bahnenbildung beobachtet man häufig in längeren Gängen, wenn sich für entgegengesetzte Laufrichtungen Bahnen ausbilden, ohne dass hierfür externe Vorgaben gemacht werden. Dies gilt insbesondere, wenn in beide Laufrichtungen in etwa dieselbe Geschwindigkeit und Personendichte herrscht. Unterscheiden sich die Geschwindigkeiten jedoch, so kommt es (zusätzlich) zu einer sogenannten Clusterbildung, bei der sich langsamere Gruppen zu Clustern zusammenfassen lassen, um die der schnellere Strom von Einzelpersonen herumläuft.

Eine Besonderheit bei der Gruppenbildung ist vor künstlich erzeugten Engpässen wie Türen etc. zu beobachten. Auch hier kommt es ohne äußere Vorgaben zu einer sich selbst organisierenden "Vereinbarung", dass bei mehreren Durchgängen je einer pro Laufrichtung genutzt wird, weil so der Durchsatz insgesamt maximiert wird. Gibt es hingegen nur einen Durchgang, so beobachtet man eine zeitlich alternierende Nutzung der Tür o.ä. in jeweils entgegengesetzter Richtung.

Die Kategorie "Menschenansammlung" ist für die weiteren Untersuchungen dahingehend von Nutzen, um bereits anhand anderer Merkmale identifizierte Gruppenzugehörigkeiten auch bei Durchmischungen von Gruppen aufrecht zu erhalten. Hierfür erfolgt eine Auswertung der Laufrichtungen, die Aufschluss darüber gibt, welche Personen zu welcher (Teil-)gruppe zuzuordnen sind.

Insgesamt sind Menschenansammlungen aber aufgrund der geringen zeitlichen Stabilität sowie der generell geringen Aussagekraft für die weiteren Betrachtungen eher uninteressant.

² Quelle: Hubert Klüpfel, Michael Schreckenberg, Tim Meyer-König, *Models for Crowd Movement and Egress Simulation*, Duisburg 2004

Temporären Gruppen ist demgegenüber zu eigen, dass Personen aufgrund eines gemeinsamen "Zwecks" einander zugeordnet werden. Ein typisches Beispiel hierfür sind Warteschlangen vor einem Schalter, in denen sich Personen über einen längeren Zeitraum auf engem Raum versammeln.

Praktisch nutzbar wird diese Information beispielsweise dann, wenn man aufgrund der "Schlangenzugehörigkeit" beim Check-in Rückschlüsse auf den Zielflughafen eines Passagiers ziehen kann.



Abbildung 2-3: Warteschlange am Flughafen³

Wenn man sich die Personen in Abbildung 2-3 näher betrachtet, so fällt auf, dass es innerhalb der temporären Gruppe "Warteschlange" weitere, kleinere Untergruppen gibt, in denen die Personen miteinander interagieren.

Diese **echten Gruppen** zeichnen sich in erster Linie dadurch aus, dass die Personen miteinander in Kontakt stehen, sich unterhalten und sich offenbar kennen. Hierbei handelt es sich typischerweise um Familien, Freunde, Bekannte oder Kollegen, die gemeinsam eine Flugreise antreten oder von einer solchen zurückkommen. Die typische Größe dieser echten Gruppen beträgt dabei im Schnitt lediglich 2-3 Personen, wobei die Gruppengröße auch vom Reisezweck abhängig ist. Während Geschäftsreisende mehrheitlich (73 Prozent) allein unterwegs sind, liegt die Gruppengröße bei Touristen im Schnitt bei 2, wie die Erhebungen an den Flughäfen Dresden und Stuttgart zeigen (vgl. Abbildung 2-4 f.).

³ Quelle: http://www.dlr.de/dlr/presse/de/Portaldata/1/Resources/bilder/portal/portal_fue_2011_1/scaled/Warteschlange_16_9_1.jpg

Gruppengröße n (Personen)	Häufigkeiten (absolut und prozentual)					
	Geschäftsreisende		Touristen		Gesamt	
1	183	(73 %)	23	(19 %)	206	(56 %)
2	57	(23 %)	66	(55 %)	123	(33 %)
3	11	(4 %)	31	(26 %)	42	(11 %)

Abbildung 2-4: Gruppengröße am Flughafen Dresden⁴

Gruppengröße n (Personen)	Häufigkeiten (%)		
	LVG A	LVG B	Gesamt
1	24	21	22
2	64	65	65
3+	12	14	13

Abbildung 2-5: Gruppengröße am Flughafen Stuttgart⁵

Ein wesentliches Merkmal, das echten Gruppen im Sinne dieser Definition zu eigen ist, ist deren zeitliche Stabilität. Während Menschenansammlungen und temporäre Gruppen lediglich für einen kurzen Zeitraum von höchstens wenigen Minuten stabil sind, kann in den meisten Fällen davon ausgegangen werden, dass echte Gruppen während ihres gesamten Aufenthalts am Flughafen bestehen bleiben. Dabei ist es unerheblich, ob sie sich zwischenzeitlich für eine gewisse Zeit auflösen, zum Beispiel beim Aufsuchen der Waschräume, da davon ausgegangen werden kann, dass sie sich danach wieder vereinigen.

Im Hinblick auf die Nutzung dieser Erkenntnisse im Rahmen von APFeI+ ist der Aspekt der zeitlichen Stabilität von großer Bedeutung, da es in erster Linie um die Verbesserung der Leistungsfähigkeit der Wiedererkennung von Personen geht (vgl. Kapitel 3.2). Eine Person, die einer echten Gruppen zugeordnet werden kann und deren Gruppenzugehörigkeit über den gesamten Aufenthalt am Flughafen stabil ist, lässt sich anhand von Gruppenmerkmalen gegebenenfalls leichter wiedererkennen als anhand von Einzelmerkmalen der Person selbst.

Insgesamt lässt sich festhalten, dass die zeitliche Stabilität einer Gruppe mit der Verbindlichkeit der Zusammengehörigkeit zunimmt. Während Menschenansammlungen im Sinne von "Zufallsbegegnungen lediglich wenige Sekunden stabil sind, liegt die zeitliche Stabilität von temporären Gruppen üblicherweise im Bereich mehrerer Minuten. Eine Besonderheit bilden hierbei Gruppenzugehörigkeiten, die sich zwischenzeitlich lösen, um sich zu einem späteren Zeitpunkt wieder zu vereinigen. Ein typisches Beispiel hierfür sind Warteschlangen vor einem dezidierten Check-in-Schalter, bei dem sich die Passagiere zunächst einem Flug zuordnen lassen, sich dann jedoch in der Zeit bis zum Abflug wieder verstreuen. Erst kurz vorm Boarding finden sich dann alle Passagiere des betreffenden Fluges wieder im Abflugwarteraum ein.

Erst echte Gruppen weisen jedoch eine zeitliche Stabilität auf, die im Rahmen des Vorhabens zu Wiedererkennung von Personen im Videomaterial sinnvoll nutzbar ist. Sie dauert im Extremfall vom Betreten des Flughafens bis zum Abflug bzw. von der Ankunft bis zum Verlassen des Flughafens.

⁴ Quelle: Michael Schultz, Entwicklung eines individuenbasierten Modells zur Abbildung des Bewegungsverhaltens von Passagieren im Flughafenterminal, Dresden 2010

⁵ Quelle: ebenda

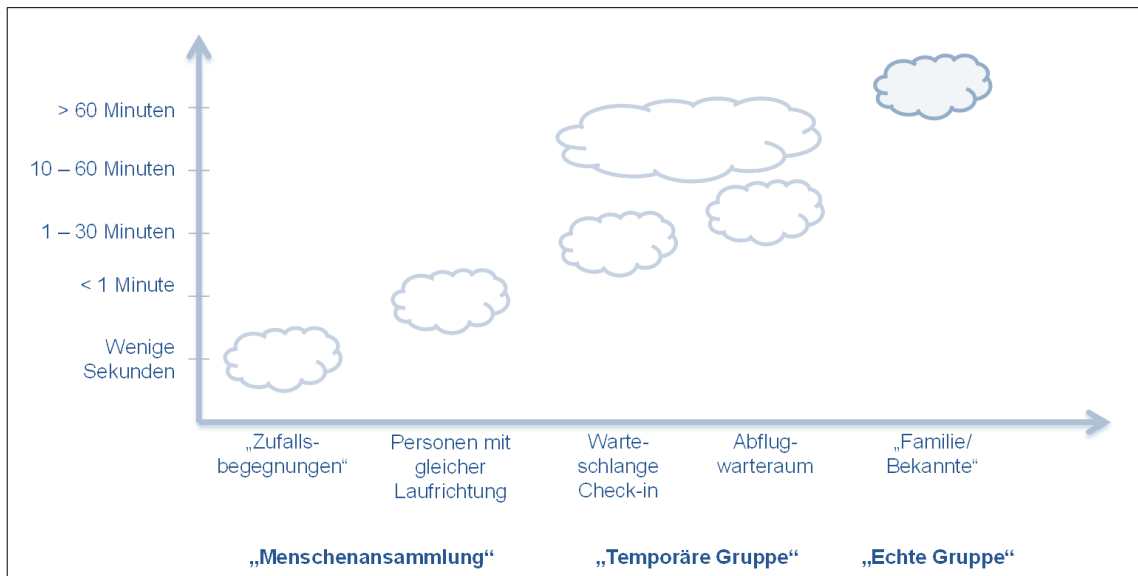


Abbildung 2-6: Zeitliche Stabilität verschiedener Gruppen

2.2 Räumliche Distanz innerhalb von Gruppen

Um in der Lage zu sein, Gruppenzugehörigkeiten einzelner Personen zu erkennen und zu beschreiben, ist es sinnvoll, sich auf messbare Größen aus dem Videodatenstrom zu konzentrieren, da eine automatisierte Erkennung echter sozialer Interaktionen mit heutigen technischen Hilfsmitteln noch nicht möglich ist.

Die einfachste Größe zur Beschreibung, ob zwei oder mehr Personen miteinander in einer Verbindung stehen, ist daher der räumliche Abstand zwischen diesen Personen. Da die Positionen aus der Personendetektion von APFEL bekannt sind, kann anhand einer Auswertung der Abstände zwischen den Personenspuren auf eine Gruppenzugehörigkeit geschlossen werden.

Hierfür ist zu untersuchen, welche Distanzen zwischen Personen für die Entscheidung relevant sind, um auf eine Gruppenzugehörigkeit schließen zu können. Hierbei unterscheidet man in der Proxemik, als einem Teilgebiet der Psychologie und der Kommunikationswissenschaften, zwischen vier unterschiedlichen Distanzklassen, die nachfolgend näher beschrieben werden.

Bei Begegnungen auf der Straße oder in der Öffentlichkeit versuchen die meisten Menschen, automatisch einen Abstand von mehr als dreieinhalb Metern zwischen sich und anderen Passanten einzuhalten. Dieser Abstand wird daher auch **öffentliche Distanz** (> 3,50m) genannt. Wenn man einem fremden Menschen auf der Straße näher kommt als die öffentliche Distanz es zulässt, weckt das automatisch seine Aufmerksamkeit. Der Betroffene versucht auszuweichen oder analysiert die Körpersprache um herauszufinden, ob er die Person kennt und ob sie mit ihm in Kontakt treten möchte, um zum Beispiel nach dem Weg zu fragen.

Die sogenannte **gesellschaftliche oder soziale Distanz** entspricht einer Entfernung von 1,50 Meter bis 3,50 Metern. Bei öffentlichen Reden, formellen Anlässen oder offiziellen Gesprächen gilt diese Entfernung zueinander als angemessen. Verletzt jemand diese Distanz und kommt in die direkte Nähe von anderen Personen, können diese sich bedroht fühlen.

Die **persönliche Distanz** von zirka einem halben Meter bis hin zu 1,50 Metern ist Menschen vorbehalten, mit denen man vertraut ist. In diesem Interaktionsraum werden auch persönliche Themen besprochen, die eine nähere Beziehung zueinander voraussetzen. Die Akteure sind hier miteinander vertraut, jedoch nicht so sehr, dass sie automatisch Körperkontakt zulassen würden.

Die **intime Distanz** beträgt weniger als einen halben Meter. Diese Nähe eignet sich, um Themen zu besprechen, bei denen man sehr vertraut miteinander umgeht. Menschen, die in die unmittelbare Nähe einer Person treten dürfen, kennen diese sehr gut und sind in der Regel berechtigt, sie auch zu berühren oder zu umarmen. Ein unberechtigtes Eindringen in die intime Distanz ist hingegen problematisch und führt sehr schnell zu Abwehrreaktionen.

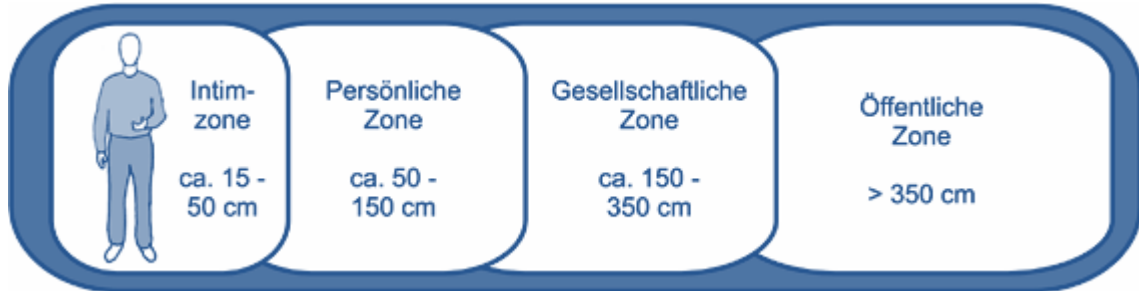


Abbildung 2-7: Distanz-Zonen⁶

Etwas anders verhält es sich allerdings, wenn begrenzte Platzverhältnisse herrschen und somit unabhängig von der Beziehung der Personen zueinander diese definierten Abstände im Gedränge häufig unterschritten werden. Daher ist neben dem Abstand selbst auch die zeitliche Dauer bzw. die Häufigkeit der Unterschreitung von Bedeutung.

Während in reinen Menschenansammlungen im Gedränge eines Flughafens die intime Distanz regelmäßig unterschritten wird, so erfolgt dies jeweils nur für einen kurzen Moment, bevor sich die zwischenzeitlich in Beziehung stehenden Personen wieder verlieren.

Bei temporären Gruppen kann es durchaus vorkommen, dass die intime Distanz für eine längere Zeit unterschritten wird, wenngleich die betreffenden Personen dies typischerweise zu vermeiden suchen. Für die Auswertung ist daher ebenfalls von Bedeutung, wie lange die Unterschreitung dauert und wie große die Personendichte um die betreffenden Personen herum ist. Kann nachgewiesen werden, dass dichtes Gedränge herrschte, so ist entsprechend eine höhere Toleranz angebracht, während eine Unterschreitung der intimen Distanz ohne Gedränge einen deutlichen Hinweis liefert, dass es sich bei den betreffenden Personen tatsächlich um eine echte Gruppe handelt.

Nachdem nun für quasi jede Person, die im Fokus eines Operator steht, anhand der räumlichen Abschätzung eine (wahrscheinliche) Gruppenzugehörigkeit angegeben werden kann, soll nachfolgend darauf eingegangen werden, wie diese Erkenntnisse im APFeL-Verbund weiter genutzt werden können.

⁶ <http://www.kalkus.at/wp-content/uploads/2011/07/distanzzonen.gif>

KAPITEL 3 Nutzung von Gruppenmerkmalen

3.1 Nutzung der Gruppengröße für die Abschätzung der Laufgeschwindigkeit

Eines der Ziele von APFEL ist es, den wahrscheinlichen Aufenthaltsort einer vom Operator markierten Person zu finden, nachdem diese zuletzt in einem Videobild erkannt wurde. Für die Abschätzung, welchen Weg bzw. welche Entfernung diese Person innerhalb eines bestimmten Zeitraumes zurückgelegt haben kann, ist unter anderem festzulegen, wie schnell sich die Person bewegt hat.

Auf der Grundlage eines Lageplans und der mittlerweile verstrichenen Zeit lässt sich so abschätzen, wie weit entfernt diese Person maximal gelaufen sein kann. Diese Analyse kann noch weiter verfeinert werden, indem bestimmte Wege mit einer höheren Wahrscheinlichkeit belegt werden als andere. Wenn beispielsweise bekannt ist, dass die Person bereits für einen bestimmten Flug eingecheckt hat, so besteht eine höhere Wahrscheinlichkeit, dass sich diese Person in Richtung ihres Abfluggates begeben hat. Eine weitere Detaillierung ergibt sich aus der Kenntnis der Prozesszeiten dazwischen, zum Beispiel der Wartezeiten an der Sicherheitskontrolle. Bestimmte Wege können aufgrund von Zugangsbeschränkungen hingegen komplett ausgeschlossen werden, die sich beispielsweise aus der Trennung ankommender und abfliegender Passagierströme oder aus der Unterscheidung Schengen/Non-Schengen ergeben.

Ein weiteres Unterscheidungsmerkmal, das sich aus den Erkenntnissen der Untersuchungen zum Gruppenverhalten ergibt, ist die Laufgeschwindigkeit selbst. Anhand von Beobachtungen konnte festgestellt werden, dass sich die Laufgeschwindigkeit mit zunehmender Gruppengröße verringert. Kann man also einer gesuchten Person eine (echte) Gruppe zuordnen, so kann man dieser Person eine andere (maximale) Laufgeschwindigkeit zuordnen als einer Einzelperson und somit den Suchraum weiter präzisieren.

Die Untersuchungen haben weiter gezeigt, dass auch der Reisezweck (Tourist/Geschäftsreisende) einen Einfluss auf die Laufgeschwindigkeit hat, wie nachfolgende Darstellungen für den exemplarisch untersuchten Flughafen Dresden zeigen. Während Einzelpersonen je nach Reisezwecke zwischen 1,19 m/s (Touristen) und 1,37 m/s (Geschäftsreisende) zurücklegen, sinkt diese Geschwindigkeit für eine Gruppengröße von 3 Personen auf 0,96 m/s bzw. 1,04 m/s.

Gruppengröße n (Personen)	Geschäftsreisende		Tourist		Gesamt	
	(mittlere Geschwindigkeit, Standardabweichung in m/s)					
	v_G	σ_G	v_T	σ_T	v_n	σ_n
1	1,38	0,21	1,19	0,25	1,36	0,23
2	1,17	0,17	0,97	0,20	1,06	0,21
3	1,04	0,23	0,93	0,17	0,96	0,19

Abbildung 3-1: Laufgeschwindigkeiten in Abhängigkeit der Gruppengröße am Flughafen Dresden⁷

Insgesamt betrachtet lässt sich die Laufgeschwindigkeit auch als Normalverteilung darstellen, wie nachfolgende Abbildung zeigt. Beispielhaft wurden hier die Kurven für einen bzw. drei Passagiere dargestellt.

⁷ Quelle: Michael Schultz, 2010

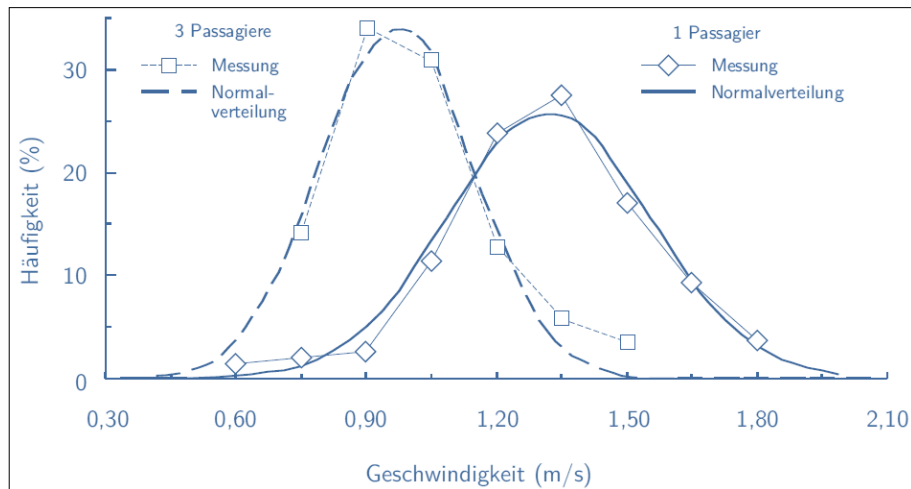


Abbildung 3-2: Verteilungsfunktion der Laufgeschwindigkeit in Abhängigkeit der Gruppengröße am Flughafen Dresden⁸

Neben der Untersuchung zum Laufverhalten in Gruppen am Flughafen Dresden wurde eine solche Untersuchung auch während der EXPO in Hannover durchgeführt, bei der das Laufverhalten beim Überqueren einer Brücke beobachtet wurde. Auch hierbei zeigte sich ein ähnliches Verhalten: Einzelpersonen bewegten sich mit einer durchschnittlichen Geschwindigkeit von 1,38 m/s, während Personen in Gruppen langsamer liefen.

Group size	Number of groups	Mean Velocity
1	95	1.38
2	149	1.28
3	59	1.24
4	17	1.24
5	10	1.22
6	2	1.10
	700	1.30

Abbildung 3-3: Laufgeschwindigkeit in Abhängigkeit der Gruppengröße auf der EXPRO-Hannover⁹

Durch eine Einbeziehung dieser Erkenntnisse in die Prognosefunktion von APFEL lässt sich der Suchraum entsprechend weiter verfeinern.

⁸ Quelle: Michael Schultz, 2010

⁹ Quelle: Hubert Ludwig Klüpfel, *A Cellular Automaton Model for Crowd Movement and Egress Simulation*, Würzburg 2003

3.2 Nutzung der Gruppenzugehörigkeit für die Wiedererkennung

Abschließend soll dargestellt werden, wie die zuvor beschriebenen Erkenntnisse zur Gruppenzugehörigkeit Eingang in die APFEl-Umgebung finden und so für einen Operator im Leitstand nutzbar gemacht werden können.

Nachfolgende Abbildung gibt zunächst einen Überblick über die mögliche Einbeziehung von Gruppenmerkmalen in den Prozess der Wiedererkennung von markierten Personen im Videodatenstrom.

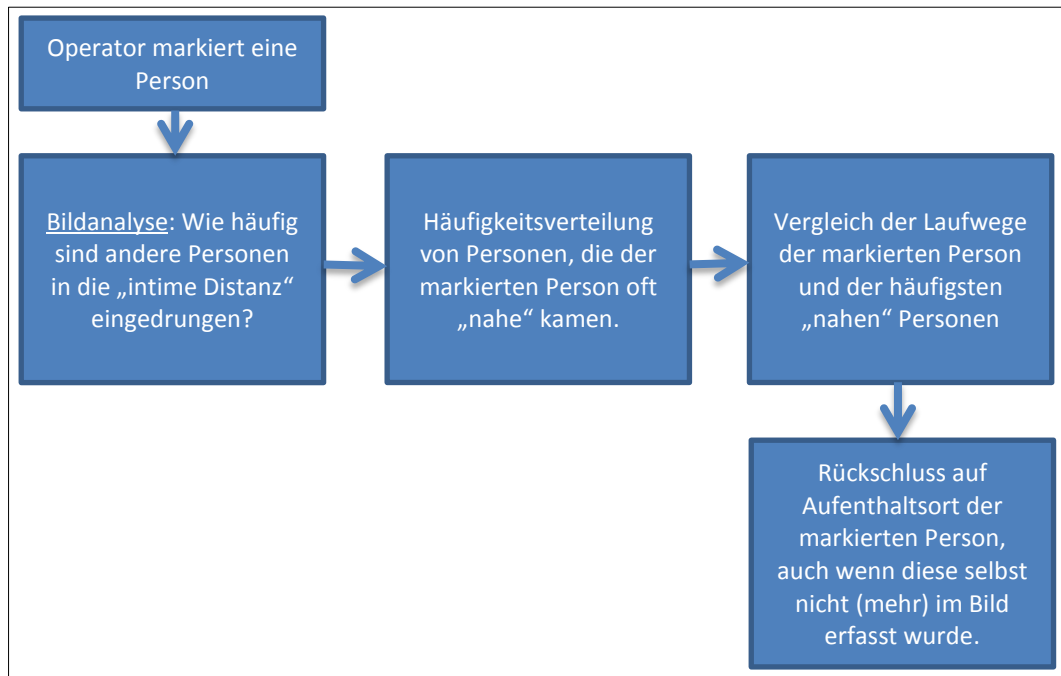


Abbildung 3-4: Ablaufschema der Wiedererkennung anhand von Gruppenzugehörigkeiten

Ausgangspunkt in der APFEl-Philosophie ist in jedem Fall die Markierung einer Person im Videodatenstrom durch den Operator. Anhand der Mustererkennung sowie weiterer Algorithmen wird danach zunächst analysiert, in welchen Videosequenzen die markierte Person ebenfalls zu erkennen war. Auf der Grundlage dieser Daten lässt sich eine erste Abschätzung treffen, woher diese Person kam.

Untersucht man darüber hinaus, in welchen Sequenzen die Person auch noch nach der Markierung erkannt wurde, so lässt sich daraus rückschließen, wohin die Person (nach der Markierung im Bild) gegangen sein könnte. Grundlage für diese Wiedererkennung sind allein personenbezogene Bild-Merkmale, die als Metadaten gespeichert wurden und anhand derer die markierte Person identifiziert bzw. wiedererkannt werden kann. Das funktioniert grundsätzlich auch über mehrere Kameras hinweg, wobei in Abhängigkeit der Perspektive sowie weiterer Randbedingungen, wie zum Beispiel der Beleuchtung, eine Erkennung der Person selbst erschwert bis unmöglich wird.

An dieser Stelle setzen die Algorithmen zur Gruppenanalyse an. Die grundlegende Idee dabei ist, die gesuchte (=markierte) Person nicht selbst im Videodatenstrom zu finden, sondern eine andere Person, die jedoch zur selben Gruppe gehört. Die Wahrscheinlichkeit, dass bei einer "stabilen" Gruppe, zum Beispiel einer Familie mit kleinen Kindern, zumindest eine Person hieraus wiedergefunden wird, ist signifikant höher, als würde man ausschließlich nach einer Einzelperson suchen. Über die Gruppenzugehörigkeit lässt sich dennoch so auch auf den Aufenthaltsort der eigentlich gesuchten Person rückschließen.

Hierfür ist in einem ersten Schritt nach der Markierung zu analysieren, welche weiteren Personen der eigentlich gesuchten Person häufig nahe gekommen, also in deren

intime Distanz eingedrungen sind. Darüber lassen sich Rückschlüsse ziehen, welche weiteren Personen möglicherweise zur Gruppe der eigentlich gesuchten Person gehören.

Dies geschieht anhand einer getakteten Zählung der Unterschreitung der intimen Distanz bezogen auf die gesuchte Person. Gezählt werden die Zeiteinheiten, zum Beispiel Sekunden, in denen eine andere Person die intime Distanz unterschritten hat. Somit lässt sich nicht nur erkennen, welche Personen in die intime Distanz eingedrungen sind, sondern auch wie lange dies jeweils dauerte.

Aus der Liste der Personen, die überhaupt die vorab definierte Distanz unterschritten haben, lassen sich somit in einem zweiten Schritt durch eine Sortierung genau diejenigen herausfiltern, die dies besonders häufig bzw. über einen langen Zeitraum getan haben. Zufallsbegegnungen, bei denen die intime Distanz nur kurz unterschritten wurde, z.B. im Gedränge, werden damit ebenso herausgefiltert wie die Zugehörigkeit zu temporären Gruppen, wenn der Schwellenwert der Distanz ausreichend klein gewählt wurde.

In einem dritten Schritt wird die Suche im Videodatenstrom schließlich auf diejenigen Personen ausgedehnt, die besonders häufig in die intime Distanz eingedrungen sind und somit mit einer hohen Wahrscheinlichkeit derselben Gruppe angehören wie die eigentlich gesuchte Person. Wird dabei eine andere Person aus derselben Gruppe im Videodatenstrom wiedererkannt, so kann der Operator anhand einer Betrachtung des Bildmaterials der betreffenden Sequenz erkennen, ob auch die gesuchte Person selbst in dieser Szene vorhanden ist, auch wenn im Rahmen der automatisierten Erkennung kein Ergebnis geliefert wurde.

Damit ist durch die Einbeziehung von Gruppenmerkmalen in die Suche einer Person, die zuvor vom Operator markiert wurde, eine höhere Wiedererkennungsrates zu erwarten, als würde man im Videomaterial ausschließlich nach Merkmalen einer einzelnen Person suchen.

Liste der Veröffentlichungen im Rahmen des F+E-Vorhabens

FKZ: 13 N 10 797

Außerhalb des Abschlussberichtes sind die Ergebnisse des o.g. F+E-Vorhabens in folgenden Beiträgen publiziert worden:

Eisenbach, M., Kolarow, A., Schenk, K., Debes, K., Gross, H.-M.: View Invariant Appearance-based Person Reidentification Using Fast Online Feature Selection and Score Level Fusion. In: Proc. 9th IEEE Int. Conf. on Advanced Video and Signal-Based Surveillance (AVSS 2012), Beijing, China , pp. 184-190, IEEE 2012.

Eisenbach, M., Scheiner, P., Kolarow, A., Schenk, K., Gross, H.-M., Weinreich, I.: Learning Illumination Maps for Color Constancy in Person Reidentification. In: Proc. 19. Workshop Farbbildverarbeitung (FWS13), Berlin, Germany, pp. 103-114, GFAI 2013.

Kolarow, A., Brauckmann, M., Eisenbach, M., Schenk, K., Einhorn, E., Debes, K., Gross, H.-M.: Vision-based Hyper-Real-Time Object Tracker for Human-Robot Applications. In: Proc. IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS 2012), Vilamoura, Portugal, pp. 2108-2115, IEEE 2012.

Kolarow, A., Schenk, K., Eisenbach, M., Dose, M., Brauckmann, M., Debes, K., Gross, H.-M.: APFeI: The Intelligent Video Analysis and Surveillance System for Assisting Human Operators. In: Proc. 10th IEEE Int. Conf. on Advanced Video and Signal-Based Surveillance (AVSS 2013), Krakow, Poland, pp. 195-201, IEEE 2013

Schenk, K., Eisenbach, M., Kolarow, A., Gross, H.-M.: Comparison of Laser-based Person Tracking at Feet and Upper-Body Height. In: Proc. 34th Annual Conference on Artificial Intelligence (KI 2011), Berlin, Germany, LNAI 7006, pp. 277-288, Springer 2011.

Schenk, K., Kolarow, A., Eisenbach, M., Debes, K., Gross, H.-M.: Automatic Calibration of Multiple Stationary Laser Range Finders using Trajectories. In: Proc. 9th IEEE Int. Conf. on Advanced Video and Signal-Based Surveillance (AVSS 2012), Beijing, China , pp. 306-312, IEEE 2012.

Schenk, K., Kolarow, A., Eisenbach, M., Debes, K., Gross, H.-M.: Automatic Calibration of a Stationary Network of Laser Range Finders by Matching Movement Trajectories. In: Proc. IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS 2012), Vilamoura, Portugal , pp. 431-437, IEEE 2012.

Ein Freixemplar der jeweiligen Veröffentlichungen haben wir als Anlage beigefügt.

Ilmenau, den 22.9.2014

Ort und Datum

Unterschrift