

# Visuelle Sensorik für AAL-Anwendungen: Ansätze und Herausforderungen

Jens Spehr, Simon Winkelbach und Friedrich M. Wahl  
Institut für Robotik und Prozessinformatik  
Technische Universität Braunschweig  
Mühlenpfordtstraße 23  
38106 Braunschweig  
{j.spehr, s.winkelbach, f.wahl}@tu-bs.de

**Abstract:** Ambient Assisted Living (AAL) ermöglicht eine situationsabhängige und unaufdringliche Unterstützung des Menschen im alltäglichen Leben. Es wird daher besonders aus der Sicht des demographischen Wandels in Zukunft an gesellschaftlicher Relevanz gewinnen. Entscheidend für die erfolgreiche Umsetzung von AAL-Anwendungen und für die Praxistauglichkeit sind robuste, benutzerfreundliche und unauffällige Sensoren, die in der Lage sind, anwendungsspezifische Messgrößen zu bestimmen. Visuelle Sensoren, wie z.B. Kameras, sind als universelle Sensoren für die Erfassung von unterschiedlichen Messgrößen und sehr großen Datenmengen geeignet. Diese Arbeit fasst einige aus der Literatur bekannte Ansätze, die im AAL-Kontext stehen, in verschiedene Anwendungsgebiete zusammen und ordnet diese in ein einheitliches Bildverarbeitungsframework ein. Um die Praxistauglichkeit vorhandener Verfahren zu evaluieren, wurde eine Langzeitinstitution über einen Zeitraum von 14 Monaten in einem realen Einpersonenhaushalt vorgenommen. Erste Auswertungsergebnisse zeigen, dass die in der Literatur vorgestellten Ansätze für den Einsatz in einer realen Wohnung nicht geeignet sind. Offene Probleme und Herausforderungen werden in dieser Arbeit herausgearbeitet und diskutiert. Im Forschungsverbund 'Gestaltung altersgerechter Lebenswelten' stellen diese Herausforderungen u.a. eine Motivation dar, um die vorhandenen Ansätze weiterzuentwickeln bzw. neue Ansätze zu finden.

## 1 Einleitung

Die situationsabhängige und unaufdringliche Unterstützung des Menschen im alltäglichen Leben, auch Ambient Assisted Living (AAL) genannt, kann je nach Nutzergruppe sehr unterschiedlich aussehen. Während bei jüngeren, gesunden Menschen hauptsächlich die Unterhaltung und Lifestyle-Funktionen zur Steigerung der Lebensqualität im Vordergrund stehen, geht es bei älteren Menschen darum, ein sicheres, selbstständiges Leben im häuslichen Umfeld überhaupt erst möglich zu machen bzw. zu verlängern. Diese Ziele ergeben sich aus Motivationen wie demographischer Wandel, Trend zum Alleinleben, steigende Ansprüche an die Lebensqualität sowie steigende Bedürfnisse an Komfort und Sicherheit. Hierbei stellt besonders der Aspekt der unaufdringlichen Unterstützung hohe Ansprüche an die nutzerzentrierte Technik. Die Technik muss möglichst unauffällig durch passende Sensorik den Zustand der Person und den der Wohnung erfassen, interpretieren

und kontextabhängig Aktionen durchführen können. Visuelle Sensoren, wie z.B. Kameras, ermöglichen eine unauffällige Erfassung des Zustandes von Personen und Raum. Seit der Einführung von günstigen Kameras und leistungsstarken Rechnern werden sie daher immer häufiger im Bereich AAL eingesetzt.

Diese Arbeit ist wie folgt aufgebaut. In Kapitel 2 werden zunächst die wichtigsten AAL-relevanten Anwendungsgebiete mit exemplarischen Arbeiten verdeutlicht. Diese Anwendungsgebiete werden anschließend in Kapitel 3 in ein einheitliches Framework eingebettet. Basierend auf einer neu erstellten Langzeitdatenbank, die über einen Zeitraum von über einem Jahr in einer realen Wohnung aufgenommen wurde, werden in Kapitel 4 Herausforderungen von 'State-of-the-art'-Ansätzen aus der Literatur deutlich gemacht und diskutiert. Die Arbeit endet mit einer kurzen Zusammenfassung und einem Ausblick auf die Weiterentwicklung im Kontext des Forschungsverbundes zur 'Gestaltung altersgerechter Lebenswelten'[EHH<sup>+</sup>09].

## **2 Anwendungsgebiete**

Im Folgenden werden die wichtigsten AAL-relevanten Arbeiten auf dem Gebiet der visuellen Sensoren und digitalen Bildverarbeitung kurz erläutert.

### **2.1 Telemonitoring / Ganganalyse**

Das Telemonitoring hat zum Ziel, eine Fernuntersuchung eines Patienten und die Ferndiagnose des behandelnden Arztes zu ermöglichen. Im Bereich AAL versucht das Telemonitoring im häuslichen Umfeld Vitaldaten zu sammeln, die dann dem Haus- oder Facharzt übermittelt werden. Visuelle Sensorik versucht allein auf Basis visueller Bildinformationen Rückschlüsse auf den Gesundheitszustand des Patienten zu ziehen. Im Wesentlichen beruhen diese Rückschlüsse auf Ganganalysen, die berührungslos durchgeführt werden.

Hier kann zwischen zwei Ansätzen zur Ganganalyse unterschieden werden. Zum einen sind Verfahren bekannt, bei denen die Person mit Markern ausgestattet werden muss, und zum anderen sind Arbeiten zur markerlosen Ganganalyse bekannt. Die markerbasierten Ansätze haben den Vorteil, dass die Berechnungen zur 3d Rekonstruktion einfacher sind und die Ergebnisse im Allgemeinen auch genauer sind als diejenigen der markerlosen Ganganalyse. Allerdings müssen die Marker möglichst körpernah angebracht werden, was in der Regel sehr aufwendig ist und auch einen nicht zu vernachlässigenden Einfluss auf den Gang hat. Markerlose System stellen hohe Ansprüche an die Auswertung, da die Analyse der Bilddaten sehr aufwendig ist. Im Folgenden werden einige Arbeiten auf dem Gebiet der markerlosen Personenerkennung kurz vorgestellt. In [RDWE91] wird ein Verfahren zur Vermessung von Kniewinkeln beschrieben. Das Verfahren basiert auf der Verwendung von Infrarotdiodenarrays, die an definierten Körperstellen angebracht werden. Mithilfe eines Infrarotlokalisierungssystems werden die zu vermessenden Winkel bestimmt. Aktuelle kommerzielle Ganganalysen basieren auf aktiven Mehrkamerasystemen [MS09]. Jede Kamera ist mit einer Infrarotlichtquelle ausgestattet. Das Infrarotlicht wird

von den Markern, die am Körper des Probanden angebracht sind, reflektiert und das reflektierte Signal wird anschließend von den Kameras verwendet, um die 3d Position des Markers zu triangulieren. Markerlose Systeme sind hauptsächlich noch Gegenstand der Forschung, was auch an der hohen Anzahl an aktuellen Veröffentlichungen zu erkennen ist. Einen guten Überblick geben die Arbeiten von Moeslund [MG01, MHK06]. Exemplarisch sei hier die Arbeit von Rosenhahn et al. [RBK<sup>+</sup>06] hervorgehoben. Die Pose des menschlichen Körpers im 3d Raum wird mithilfe der Silhouette, die im Bild erkannt wurde, bestimmt. Ein kalibriertes Mehrkammersystem erlaubt anschließend die 3d Rekonstruktion. In [SSR<sup>+</sup>04] werden zunächst einzelne Körperteile, wie Kopf, Arme und Beine im Bild gesucht. Die erkannten Glieder werden anschließend zu einer optimalen globalen Gesamtkörperposition fusioniert.

## 2.2 Aktivitätenmonitoring

Beim Aktivitätenmonitoring werden die Aktivitäten des Bewohners kontinuierlich erfasst. Typische Aktivitäten, die mit visueller Sensorik erfasst werden können, sind: Gehen, Stehen, Laufen, Liegen und Essen. Eine Herausforderung beim Erkennen von Aktivitäten ist das Finden einer geeigneten Repräsentation. Meist kommen hierarchische Modelle zum Einsatz, die aus folgenden Ebenen bestehen: Aktionsprimitive, Aktionen und Aktivitäten. Aktionsprimitive sind atomare Bewegungseinheiten. Mehrere Aktionsprimitive bilden eine Aktion und mehrere Aktionen wiederum eine Aktivität [MHK06]. Exemplarisch seien hier zwei Arbeiten hervorgehoben [AS98, BD96]. In [AS98] werden zunächst Low-Level Verfahren wie Hautfarbenerkennung und Vordergrund-Hintergrund-Segmentierung verwendet, um die Person zu erkennen. Aktivitäten werden mithilfe eines Zustandsmodells und Vorwissen über den Grundriss der Wohnung klassifiziert. Bobick et al. verwenden in [BD96] 'Motion History Images', um Aktivitäten zu beschreiben. Aktivitäten werden anhand eines Abgleichs mit einer Datenbank klassifiziert. Diese erwähnten Arbeiten beschäftigen sich im Wesentlichen mit der Erkennung der Aktionsprimitive. Arbeiten zum Thema High-Level Aktivitätenerkennung finden sich in [Fin98, RR05].

## 2.3 Lokalisierung

Die Lokalisierung ermöglicht die Bestimmung des Aufenthaltsortes der Person im Raum. Hauptherausforderung der Lokalisierung ist die Rekonstruktion der dritten Dimension, die bei der Aufnahme mit einer einzelnen Kamera verloren geht. Um eine Zuordnung der 2d Bilddaten und des 3d Raumes zu ermöglichen, muss das Kamerasystem kalibriert werden. Die dritte Dimension kann mithilfe von aktiven oder passiven Ansätzen rekonstruiert werden. Nähere Informationen hierzu sind in Kapitel 3 zu finden.

Beispiele für kamerabasierte Lokalisierungssysteme finden sich in [GNTH00, LWN08]. In [GNTH00] wird ein Personenlokalisierungssystem vorgestellt, das die Person zunächst anhand des Kopfes erkennt und anschließend mithilfe eines Depth-From-Focus-Ansatzes im Raum lokalisiert. Andere Ansätze, wie z.B. [LWN08], verwenden ein kalibriertes Ka-

merasystem. Zunächst wird die Person im 2d Bild erkannt. Die erkannte Person wird dann anhand der Annahme, dass die Person Kontakt zur Bodenebene hat, im Raum lokalisiert.

## 2.4 Verhaltensmuster

Verhaltensmuster stellen Bewegungsabfolgen dar, die vom Menschen kontextabhängig immer in einer bestimmten, vorhersagbaren Weise und Reihenfolge durchgeführt werden. Arbeiten auf diesem Gebiet versuchen in den erfassten Sensorsinformationen, sich wiederholende Muster zu erkennen und diesen einen Kontext zuzuordnen. Stauffer und Grimson analysieren in [SG00] Verhaltensmuster von Personen und Fahrzeugen im 'Outdoor' Bereich. Hierbei werden sich wiederholende Muster durch eine einfache Akkumulation bestimmt und ein hierarchischer Binärbaum wird zur Klassifikation genutzt. In [LCCT07] werden Verhaltensmuster in einer Pflegeeinrichtung mithilfe visueller Sensorik bestimmt. Zusätzlich wird in dieser Arbeit Vorwissen verwendet, um das menschliche Verhalten zu verstehen.

## 2.5 Erkennen von Anomalien und Stürzen

Die Anomalieerkennung versucht auffällige Verhaltensmuster in Videosequenzen zu erkennen. Hauptmotivation liegt hierbei in der Überwachung von öffentlichen Plätzen, wo speziell kriminelle Handlungen erkannt werden sollen. In [SSSar] wird das Szenenverhalten mithilfe nicht parametrischer Modelle repräsentiert. Eine Markovkette wird zur Erkennung von Normalverhalten und Anomalien verwendet. Au et al. stellen in [ASC06] ein Verfahren vor, welches die Anomalien über ein kompressionsbasiertes Ähnlichkeitsmaß erkennt. Anomalien stellen in dieser Arbeit Bilder dar, die einer Datenbank noch nicht bekannt sind. Die Arbeit von Xiang und Gong [XG08] versucht Verhaltensmuster mithilfe von dynamischen Bayesschen Netzen zu modellieren. Ein Ähnlichkeitsmaß zwischen diesen Netzen ermöglicht das Erkennen von Anomalien. In [BNS08] wird zusätzlich Vorwissen zur Anomalieerkennung genutzt.

Eine spezielle Form der Anomalie ist der Sturz einer Person. Aufgrund der Relevanz seien im Folgenden die wichtigsten Arbeiten auf diesem Gebiet kurz erläutert. Tao verwendet in [TTW<sup>+</sup>05] ein Kamerasystem zur Erkennung von Stürzen. Zunächst wird die Person mithilfe einer Standard-Vordergrund-Hintergrund-Segmentierung im Bild detektiert. Anschließend wird ein umschließendes Rechteck um die Person gelegt. Das Verhältnis von Höhe und Breite dieses Rechtecks wird als Merkmal zur Sturzerkennung verwendet. Abrupte Änderungen dieses Verhältnisses werden als Sturz interpretiert. In 50 Testdurchläufen (Entfernung 4 bis 7m zur Kamera, einfache Testumgebung) wurde die Sturzerkennung evaluiert. Die durchschnittliche Erkennungsrate lag bei 72,2%. Ähnliche Merkmale zur Sturzerkennung werden in [MSH06] verwendet. Allerdings basieren die Messungen auf einer Fischaugenkamera. Die Erkennungsrate lag bei 79.8% in Tests mit 20 Probanden. Die Sturzerkennung in [SGW<sup>+</sup>08] basiert auch auf Fischaugenkamerabildern. Die Standard-Vordergrund-Hintergrund-Segmentierung wurde in dieser Arbeit auf die An-

wendung im häuslichen Umfeld angepasst. Im ersten Schritt wird hier die Orientierung der Körperachse der Person bestimmt; anschließend wird die gemessene Orientierung mit der einer stehenden Person verglichen und daraus die Sturzwahrscheinlichkeit berechnet. Das Verfahren erreichte eine Falsch-Positiv Rate von 0,31 und eine Falsch-Negativ Rate von 0,22. In [RAR07] wird ein Mehrkamerasystem zur Erkennung verwendet. Es werden gefährliche Situationen (wie Stürze) erkannt, indem die Veränderung des Erscheinungsbildes der Person mithilfe eines Hidden-Markov-Modells analysiert wird. Eine Evaluation der Sturzerkennung wurde nicht durchgeführt. Die Sturzerkennung in [RM06] konzentriert sich auf die Erkennung des Kopfes. Schnelle Bewegungen des erkannten Kopfes in Richtung des Bodens werden als Sturz interpretiert. Eine Evaluierung fehlt auch hier. Eine Fischaugenkamera, die an der Decke befestigt ist, wird in [NCM04] zur Sturzerkennung verwendet. Die Person wird mithilfe eines Partikelfilters zeitlich verfolgt. Stürze werden erkannt als 'Nichtaktivitäten' in einem 'Aktivitätenbereich'. 'Aktivitätenbereich' bezeichnet hierbei ein Bereich des Raumes, in dem die Person für gewöhnlich nicht stehen bleibt oder sich hinlegt.

## **2.6 Personenidentifikation**

Die Personenidentifikation versucht aus den Bilddaten biometrische Merkmale abzuleiten und diese zur Identifikation zu nutzen. Entscheidend hierfür ist eine Datenbank, die den biometrischen Daten einen entsprechenden Datenbankeintrag zuordnet. Solche Identifikationsmechanismen sind im Bereich AAL hauptsächlich für Mehrpersonenhaushalte interessant. Erst die Identifikation ermöglicht es z.B. die Verhaltensmuster und Ganganalysen von mehreren Personen gleichzeitig zu verwalten. Ein weiteres Anwendungsfeld ist die Zugangskontrolle.

Die berührungslose Identifikation beruhend auf visueller Sensorik basiert im Wesentlichen auf zwei verschiedenen Ansätzen: Ganganalyse z.B. [SKR<sup>+</sup>04, WTM<sup>+</sup>03, YNC02] oder Gesichtserkennung z.B. [SK87, LVB<sup>+</sup>93, VJ01]. Auf die Ganganalyse wurde bereits im Kapitel 2.1 genauer eingegangen. Die Gesichtserkennung identifiziert die Person aufgrund der Gesichtsgeometrie. Hierbei werden meist Bilder des frontalen Kopfes auf sichtbare ausgeprägte Merkmale und geometrische Relationen hin untersucht.

## **2.7 Gestererkennung**

Das automatische Erkennen von Gesten ist besonders für die Mensch-Maschine Interaktion interessant (z.B. [ESYD03, PSH97, Koh98]). Im Bereich AAL können hierdurch z.B. Haushaltsgeräte eingeschaltet oder Fernsehgeräte bedient werden. Hauptsächlich sind in der Literatur bisher Handgesten untersucht worden. In [ESYD03] werden Handgesten als Abfolge von eingelernten Handposen repräsentiert und mithilfe eines nicht parametrischen Hidden-Markov-Modells klassifiziert. Ein Überblick über prinzipielle Ansätze der Gestererkennung sind in [PSH97, Koh98] zu finden.

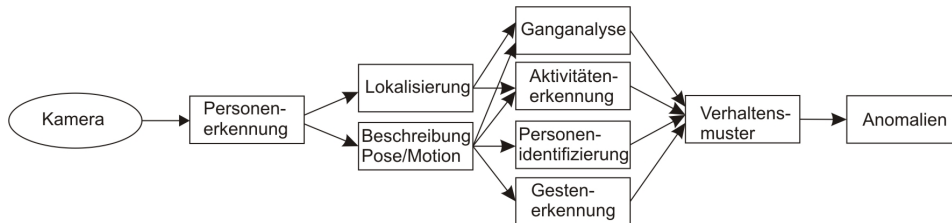


Abbildung 1: Allgemeines Bildverarbeitungsframework für AAL-Anwendungen

### 3 Bildverarbeitungsframework für AAL-Anwendungen

Die visuelle Sensorik umfasst im Allgemeinen alle Sensoren, die elektromagnetische Wellen im sichtbaren Spektralbereich (auch angrenzende Bereiche: Infrarot (IR) und Ultraviolett) in weiterverarbeitbare Größen umformen. Besonders Bildsensoren sind für den Einsatz im Bereich AAL gut geeignet, da sie als universelle Sensoren für die Erfassung von unterschiedlichen Messgrößen und sehr großen Datenmengen geeignet sind. Prinzipiell wird bei bildgebenden Verfahren zwischen zwei Ansätzen unterschieden. Bei passiven Verfahren wird lediglich ein Bildaufnahmesystem verwendet, das die bereits vorhandene elektromagnetische Strahlung der Szene misst. Passive Verfahren sind z.B. Monokular-, Wärmebild- oder Stereokamerasysteme. Aktive Verfahren verwenden zusätzliche Beleuchtungen wie z.B. IR-Strahler zur Steigerung der Robustheit oder Beseitigung von Mehrdeutigkeiten.

Die große Herausforderung der visuellen Bildsensoren ist die digitale Verarbeitung der Daten. Aufgrund der hohen Datenmenge und dem Verlust der dritten Dimension ist dies ein sehr aufwendiger Schritt, für den es noch keine einheitliche Herangehensweise gibt. Wie bereits in Kapitel 2 deutlich wurde, sind in der Literatur viele Ansätze vorhanden, die meist für einen Spezialfall entwickelt wurden. Grundsätzlich lässt sich ein einheitliches Bildverarbeitungsframework nach Abbildung 1 ableiten, in dem alle AAL-relevanten Anwendungsfälle als Verarbeitungseinheiten integriert sind. Fast alle Ansätze in der Literatur beginnen zunächst mit der Erkennung der Person im Bild. Da im AAL-Bereich meist stationäre Kameras, d.h. Kameras, die fest im Raum montiert sind, zum Einsatz kommen, wird meistens die Vordergrund-Hintergrund-Segmentierung zur Personenerkennung genutzt. Diese Segmentierung versucht den statischen Hintergrund durch ein passendes Modell zu beschreiben. Einfache Ansätze nutzen als Hintergrundmodell ein Bild vom personenleeren Raum. Meist wird hier der Mittelwert oder Median über die letzten Zeitschritte verwendet. Erweiterungen modellieren jeden Bildpunkt als normalverteilte Zufallsvariable, so dass neben dem Mittelwert auch die Varianz für Grauwertbilder bzw. die Kovarianz für Farbbilder bestimmt wird. Als besonders robust hat sich das Verfahren von Stauffer und Grimson [SG98] herausgestellt. Hier wird jeder Bildpunkt mit mehreren Gaußfunktionen modelliert. Dadurch ist das Hintergrundmodell in der Lage verschiedene Zustände zu repräsentieren, wie sie in einer Wohnung z.B. im Türbereich bei einer geschlossenen und einer geöffneten Tür auftreten.

Nachdem die möglichen Personenbereiche erkannt wurden, werden die Bereiche mit pas-

senden Personenklassifikatoren klassifiziert (z.B. [VJ01]). Dieser Schritt ist besonders in 'Outdoor'-Szenarien wichtig, um z.B. Fahrzeuge von Fußgängern zu unterscheiden. Im häuslichen Umfeld wäre der Klassifikationsschritt für die Unterscheidung von Personen und Haustieren sinnvoll.

Da bisher das Bild im Schritt 'Personenerkennung' lediglich in verschiedene Bereiche segmentiert wurde, müssen in den Verarbeitungsschritten 'Lokalisierung' und 'Beschreibung Pose/Motion' diese Bereiche genauer analysiert werden. Wie bereits in Kapitel 2.3 beschrieben ist für den Lokalisierungsschritt ein kalibriertes Kamerasystem entscheidend. Unter der Annahme, dass die Person Kontakt zur Bodenebene hat, lässt sich mit einem kalibrierten Kamerasystem die Position der Person im Raum bestimmen [LWN08]. Die Beschreibung der Pose und der Bewegung lässt sich im Wesentlichen unterteilen in modellbasierte und modellfreie Beschreibungen. Modellbasierte Beschreibungen nutzen Informationen über den menschlichen Körper und dessen kinematischen Aufbau. Hier werden meist Modelle basierend auf Linien-, 2d Kontur- oder Volumen-Repräsentationen verwendet. Modellfreie Ansätze nutzen hingegen keine zusätzlichen Modellinformationen und repräsentieren die Person z.B. durch Templates, Features oder Regionen. Genauere Details hierzu finden sich in [MG01, MHK06].

Während das Ergebnis der 'Beschreibung Pose/Motion' allen nachfolgenden Verarbeitungsschritten zur Verfügung gestellt werden muss, ist das Lokalisierungsergebnis nur für die 'Ganganalyse' und die 'Aktivitätenerkennung' von Interesse. Wie in Kapitel 2 deutlich wurde, sind die Ansätze für die Schritte 'Ganganalyse', 'Aktivitätenerkennung', 'Personenidentifikation' und 'Gestenerkennung' sehr unterschiedlich und sind daher in Abbildung 1 als parallele Verarbeitungsschritte dargestellt. Im Schritt 'Verhaltensmustererkennung' können nun alle vorangegangenen Ergebnisse verwendet werden. Dabei muss die Verhaltensmustererkennung mit entsprechenden Modellen zeitliche und räumliche Abhängigkeiten dieser Daten automatisch erlernen und darauf basierend Verhaltensmuster erkennen können. Aufgrund der Ähnlichkeit dieser Modelle zu den in Kapitel 2.2 vorgestellten Aktivitätsmodellen, ist eine Fusion der beiden Schritte und die Verwendung eines gemeinsamen Aktivitäts- und Verhaltensmustermodells denkbar.

Der Zusammenhang zwischen Lernen und Erkennen von Verhaltensmustern ist eng verknüpft mit dem Erkennen von Anomalien. Diese deuten auf unbekannte Verhaltensmuster hin, die auf der einen Seite eine Gefahrensituation (wie z.B. ein Sturz) darstellen können, auf der anderen Seite können solche Auffälligkeiten aber auch auf unkritische Änderungen des Verhaltens hindeuten (z.B. bedingt durch Änderungen des häuslichen Umfelds). In diesem Fall kann die erkannte Anomalie verwendet werden, um das Modell der Verhaltensmuster zu aktualisieren. Der Zusammenhang zwischen Lernen und Erkennen von Verhaltensmustern und von Anomalien wurde in der Literatur noch nicht ausführlich thematisiert.

Ein weiterer interessanter Aspekt ist die Gestenerkennung, die sich in der Forschung bisher hauptsächlich auf Handgesten konzentriert und sehr anwendungsspezifisch gelöst wurde. Der Schritt 'Gestenerkennung' ist daher einfach als zusätzliche parallele Verarbeitungseinheit in das Framework integriert. Im Hinblick auf das Verhaltensmustermodell ließen sich jedoch Gesten sehr viel allgemeiner interpretieren. Letztendlich kann eine Geste einfach eine Aktion, eine Aktivität oder sogar ein Verhaltensmuster sein. So lassen sich viele neue interessante Aspekte im Bereich der Mensch-Maschine Schnittstelle aufgreifen.



Abbildung 2: Diese Bilder zeigen typische Kamerabilder wie sie zur Personenerkennung im Labor (aus der HumanEva-Datenbank [SB06], links) und in der Realität (rechts) verwendet werden.

#### 4 Labor vs. Realität

Ausgangspunkte für die folgende Gegenüberstellung von einer Laborszene und einer realen Wohnung sind die gravierenden Unterschiede in der Art und Qualität der Bildinformationen, den Randbedingungen und den Kamerasetups (siehe Abbildung 2). Als ein Beispiel für eine typische Laborszene ist in Abbildung 2 (links) ein Bild aus dem HumanEva Datensatz zu sehen, welcher häufig zur Evaluierung von markerlosen Human Motion Capture Systemen verwendet wird. Um Aussagen über die AAL-Praxistauglichkeit, der in der Literatur bisher veröffentlichten Arbeiten, zu erlangen, wurde ein realer Einpersonenhaushalt über einen Zeitraum von 14 Monaten mit vier Kameras ausgestattet und die Videodaten zur Weiterverarbeitung gespeichert. In der Küche und im Wohnzimmer wurde je eine Kamera mit Fischaugenobjektiv an der Decke angebracht. Der große Öffnungswinkel der Kamera ermöglicht es, das gesamte Zimmer mit nur einem Kamerabild zu erfassen. Der längliche Flur wurde mit zwei Kameras ausgestattet, die an gegenüberliegenden Seiten montiert wurden (siehe Abbildung 3).

Nach eigenen Aussagen der 90jährigen Bewohnerin hatte sie sich bereits nach wenigen Tagen an die Kamerainstallation gewöhnt und ungestört gefüllt, so dass davon ausgegangen werden kann, dass die aufgenommenen Tätigkeiten weitestgehend den gewohnten alltäglichen Tätigkeiten entsprechen. Des Weiteren wurde die Bewohnerin nicht in kameraspezifische Probleme und Randbedingungen eingewiesen, um zu verhindern, dass die Bewohnerin Rücksicht auf das Kamerasystem nimmt.

Als 'State-of-the-art' Vordergrund-Hintergrund-Segmentierung wurde das Verfahren von Stauffer und Grimson [SG98] verwendet, um die Person im Bild zu erkennen. Die Schritte 'Lokalisierung' und 'Beschreibung der Pose/Bewegung' wurden ähnlich der Arbeiten [LWN08, HHD00] implementiert. In Abbildung 4 ist exemplarisch ein Beispielergebnis zu sehen. Abbildung 4 (links) zeigt das Kamerabild der Fischaugenkamera in der Küche. Die erkannte Person ist mit einer grünen Box hervorgehoben. Das Ergebnis der Vordergrund-Hintergrund-Segmentierung ist in Abbildung 4 (rechts) zu sehen. Die weißen Bildpunk-



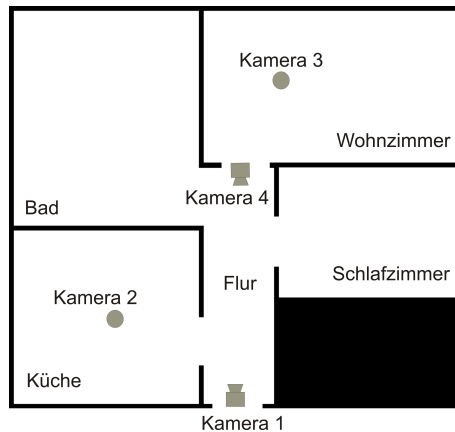


Abbildung 3: Wohnungsgrundriss der Langzeitdatenaquirierung

te repräsentieren den Vordergrund, die grauen Bildpunkte die Schattenflächen und die schwarzen Bildpunkte den Hintergrund. Während der Verarbeitung des Datensatzes haben sich die folgenden Schwierigkeiten ergeben:

- **Kleidung:** Gerade in der häuslichen Umgebung trägt der Bewohner häufig bequeme Kleidung. Diese ist im Allgemeinen recht weit geschnitten und erschwert daher die Lageschätzung der Körperteile. Bei Röcken und Umhängen ist die Posenschätzung der darunterliegenden Körperteile ohne Vorwissen nicht möglich. Hinzu kommt, dass die Kleidungsfarbe gerade bei älteren Menschen oft blas oder gräuliche ist (siehe Abbildung 5, oben links). Gerade die Unterscheidung von Person und Schatten ist hierdurch in vielen Situationen nicht möglich.
- **Gehhilfen:** Gehbehinderte oder körperlich schwache Personen verwenden auch im häuslichen Bereich oft Gehhilfen wie z.B. Rollatoren oder Gehstöcke. Gerade die Ganganalyse wird hierdurch erschwert, da es zu Verdeckungen der Beine kommen kann oder da die Gehhilfe selbst von der Bildverarbeitung falsch als Fuß klassifiziert wird.
- **Spiegelungen:** Im häuslichen Umfeld sind einige spiegelnde Flächen vorhanden, wie Spiegel selbst, Glastüren, glatte polierte Oberflächen (wie z.B. Küchenfronten) oder glatte Böden (Laminat, PVC). Sobald sich eine Person spiegelt, wird sie und ihr Spiegelbild von der Vordergrund-Hintergrund-Segmentierung als Vordergrund erkannt. Auch im Lokalisierungs- und Posen/Bewegungsbeschreibungsschritt kann es zu Mehrdeutigkeiten kommen.
- **Verdeckungen:** Im häuslichen Umfeld sind im Wesentlichen zwei Arten von Verdeckungen zu finden. Auf der einen Seite kommt es zu perspektivisch bedingten Selbstverdeckungen. Diese treten im Fall einer an der Decke angebrachten Kamera besonders dann auf, wenn sich die Person direkt unter der Kamera befindet. Auf

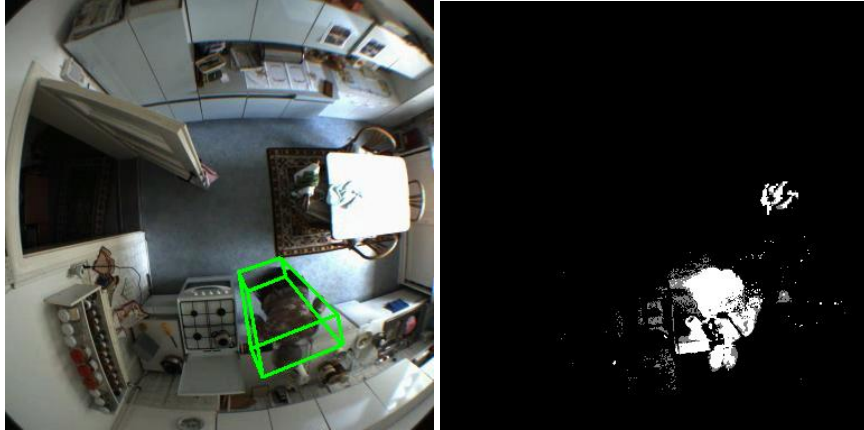


Abbildung 4: Diese Bilder zeigen exemplarisch das Ergebnis der Personenerkennung. Die erkannte Person ist mit einer umschließenden grünen Box hervorgehoben (links). Im Vordergrund-Hintergrund-Segmentierungsergebnis ist der Vordergrund weiß, Schatten grau und der Hintergrund schwarz gekennzeichnet (rechts).

der anderen Seite kann es zu Verdeckungen der Person durch das Mobiliar kommen (siehe Abbildung 5, oben ).

- **Hintergrundmodell:** Wie bereits in Kapitel 3 beschrieben, sind ausgereifte Hintergrundmodelle entwickelt worden. Dennoch haben sich einige Herausforderungen aus dem Langzeitdatensatz ergeben. Da sich die Person zu einem Zeitpunkt immer nur in einem Raum befinden kann, sind mehrere Räume zu diesem Zeitpunkt personenleer. Die Aktualisierung des Hintergrundmodells an Beleuchtungsänderungen ist in diesem Fall sehr einfach. Anders ist die Situation, wenn die Person einen Raum betritt. In diesem Fall muss das Hintergrundmodell sehr schnell an die verschiedenen Änderungen angepasst werden. So wird beim Betreten des Raumes die Tür geöffnet, meist auch das Licht eingeschaltet und evtl. noch etwas (z.B. der Einkauf) abgestellt. Prinzipiell ermöglichen Verfahren wie das Stauffer und Grimson Verfahren [SG98] bereits das Modellieren von mehreren Zuständen. Allerdings muss auch hier ständig das Hintergrundmodell an neue Änderungen angepasst werden. Hinzu kommt, dass mit der Verallgemeinerung des Hintergrundmodells die Erkennung der Person schwieriger wird. Je mehr Zustände und je allgemeiner das Hintergrundmodell, desto unempfindlicher wird die Vordergrunderkennung. Angewendet auf unsere Langzeitdatenbank, waren adaptive Verfahren bereits nach einigen Tagen so unempfindlich, dass sie nur noch kräftige Farbwerte als Vordergrund klassifizieren konnten.
- **Mehrere Benutzer:** Die Unterscheidung von mehreren Personen im Raum, ist besonders dann eine Herausforderung, wenn die Personen direkt nebeneinander stehen (siehe Abbildung 5, unten rechts). Im einfachsten Fall werden bei mehreren Personen benutzerspezifische Funktionen deaktiviert. Andernfalls muss eine Personen-

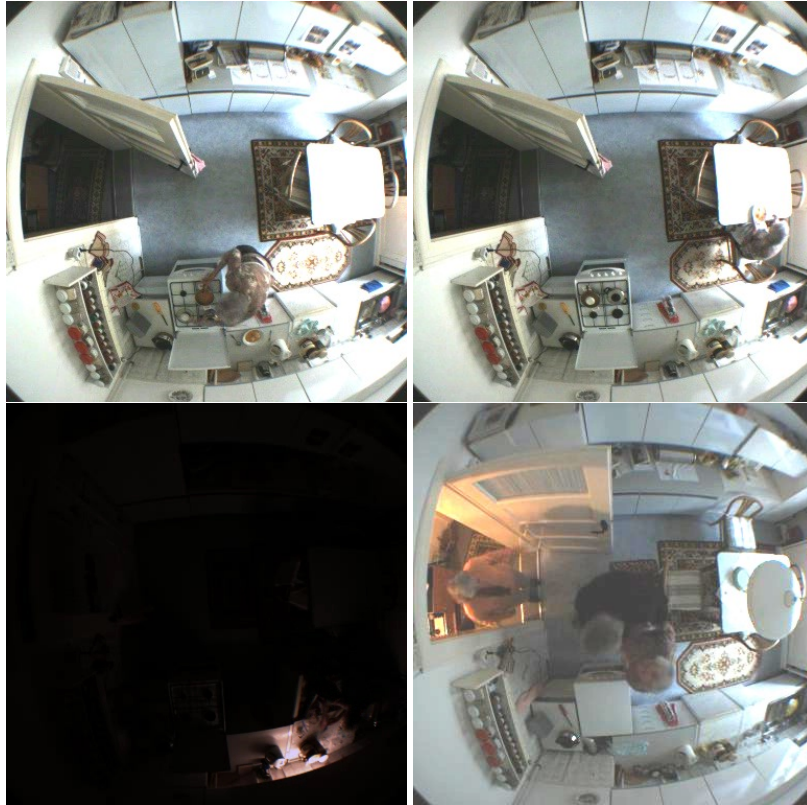


Abbildung 5: Herausforderungen der Personenerkennung im häuslichen Umfeld. Zu sehen sind typische Einzelbilder der Fischaugenkamera in der Küche. Deutlich werden in diesen Bildern Herausforderungen wie Verdeckungen und hintergrundähnliche Kleidung (oben), Beleuchtung (unten links) und mehrere Personen (unten rechts).

identifikation, wie in Kapitel 2.6 beschrieben, durchgeführt werden.

- **Haustiere:** Besonders evtl. vorhandene große Hunde erschweren die Erkennung der Person. Durch ihre waagrechte Körperhaltung sind Hunde sehr leicht mit liegenden Personen verwechselbar. Spezielle Klassifikatoren, die auf die Unterscheidung von Mensch und Tier trainiert sind, könnten hier eine Lösung sein.
- **Beleuchtung** Ein großes Problem stellen die verschiedenen Beleuchtungssituationen im häuslichen Umfeld dar (siehe Abbildung 5, unten links). Ein Problem ergibt sich durch die Anwendung von Fischaugenkameras. Durch den großen Öffnungswinkel hat die Kamera sowohl helle Bereiche (z.B. Fensterbereiche) im Bild als auch dunkle unbeleuchtete Bereiche. Standardkameras sind nicht in der Lage diesen großen Dynamikbereich entsprechend abzubilden. Zukünftig ist jedoch zu erwarten, dass 'High Dynamic Range'- Kameras zum Einsatz kommen werden, die momen-

tan jedoch noch sehr teuer sind. Ein weiteres Beleuchtungsproblem ergibt sich in Räumen ohne Fenster (siehe Flur in Abbildung 3 ). Im Fall unserer Langzeitaufnahmen hat sich ergeben, dass die Person durch den Flur geht, ohne das Licht einzuschalten. Da der Flur lediglich ein Durchgangszimmer ist, wird das Licht hier nicht von der Person beim Betreten ein- und beim Verlassen ausgeschaltet. Die Aufnahmen sind daher in der Regel etwas unterbelichtet. Die Erkennung der Person wird dadurch erschwert. Ein letztes Beleuchtungsproblem bezieht sich auf die verwendeten Leuchtmittel. Besonders problematisch sind hier Glühbirnen. Deren Strahlungsmaximum ist meist etwas in den Bereich der langwelligen Infrarotstrahlung (Wärmestrahlung) verschoben. Dadurch wirkt das Glühlampenlicht gelbrötlich. Im Vergleich zu weißem Licht wird durch Glühbirnen die Farbinformation zum Teil unbrauchbar, die Schattenerkennung wird zusätzlich erschwert.

## 5 Zusammenfassung und Ausblick

Ziel dieser Arbeit war es, grundsätzliche Anwendungsgebiete von visueller Sensorik im Bereich AAL herauszuarbeiten und relevante Ansätze aus der Literatur kurz mit repräsentativen Beispielen zu verdeutlichen. Es wurde ein einheitliches Bildverarbeitungsframework vorgestellt, das Synergieeffekte ausnutzt und so alle angesprochenen Anwendungsgebiete effizient vereint. Die Praxistauglichkeit von 'State-of-the-art'-Ansätzen wurde mit einem neu erstellten Langzeitdatensatz evaluiert und im vorliegenden Beitrag qualitativ diskutiert. Insgesamt zeigte sich, dass sich gerade beim Einsatz von visueller Sensorik in realen Umgebungen noch viele z.T. ungelöste Herausforderungen ergeben. Diese Herausforderungen sind u.a. Gegenstand des Forschungsverbundes zur 'Gestaltung altersgerechter Lebenswelten' [EHH<sup>+</sup>09]. Hier sollen die vorhandenen Ansätze weiterentwickelt bzw. neue Ansätze gefunden werden. Darüber hinaus werden hier mithilfe anderer Sensormodalitäten wie z.B. Beschleunigungssensoren Sensordatenfusionsaspekte aufgegriffen, um die situationsabhängige und unaufdringliche Unterstützung des Menschen im alltäglichen Leben noch zu verbessern.

## Literatur

- [AS98] D. Ayers und M. Shah. Recognizing Human Actions in a Static Room. In *WACV98*, Seiten 42–47, 1998.
- [ASC06] Carmen E. Au, Sandra Skaff und James J. Clark. Anomaly Detection for Video Surveillance Applications. *International Conference on Pattern Recognition*, 4:888–891, 2006.
- [BD96] A.F. Bobick und J.W. Davis. Real-time Recognition of Activity Using Temporal Templates. In *Proceedings 3rd IEEE Workshop on Applications of Computer Vision*, Seite Chapter 6, 1996.

- [BNS08] C. Brax, L. Niklasson und M. Smedberg. Finding behavioural anomalies in public areas using video surveillance data. In *11th International Conference on Information Fusion*, 2008.
- [EHH<sup>+</sup>09] M. Eichelberg, R. Haux, A. Hein, S. Winkelbach und F. Wahl. Der Niedersächsische Forschungsverbund Gestaltung altersgerechter Lebenswelten: Informations- und Kommunikationstechnik zur Gewinnung und Aufrechterhaltung von Lebensqualität, Gesundheit und Selbstbestimmung in der zweiten Lebenshälfte. In *2. Deutschen AAL-Kongress*, Berlin, Germany, Januar 2009.
- [ESYD03] Ahmed Elgammal, Vinay Shet, Yaser Yacoob und Larry S. Davis. Learning Dynamics for Exemplar-based Gesture Recognition. *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, 1:571, 2003.
- [Fin98] Shai Fine. The hierarchical hidden markov model: Analysis and applications. In *Machine Learning*, Seiten 41–62, 1998.
- [GNTH00] Saeed Shiry Ghidary, Yasushi Nakata, Toshi Takamori und Motofumi Hattori. Human Detection and Localization at Indoor Environment by Home Robot. In *in Proc. of IEEE Int. Conf. on Systems*, 2000.
- [HHD00] Ismail Haritaoglu, Davis Harwood und Larry S. David. W4: Real-Time Surveillance of People and Their Activities. *IEEE Trans. Pattern Anal. Mach. Intell.*, 22(8):809–830, August 2000.
- [Koh98] Markus Kohler. Special Topics of Gesture Recognition Applied in Intelligent Home Environments. In *In Proceedings of the Gesture Workshop*, Seiten 285–296, 1998.
- [LCCT07] Chin-De Liu, Pau-Choo Chung, Yi-Nung Chung und Monique Thonnat. Understanding of human behaviors from videos in nursing care monitoring systems. *J. High Speed Netw.*, 16(1):91–103, 2007.
- [LVB<sup>+</sup>93] Martin Lades, Jan C. Vorbrüggen, Joachim Buhmann, Jörg Lange, Christoph V. D. Malsburg, Rolf P. Würtz und Wolfgang Konen. Distortion Invariant Object Recognition in the Dynamic Link Architecture. *IEEE Trans. Computers*, 42:300–311, 1993.
- [LWN08] Yuan Li, Bo Wu und Ram Nevatia. Human detection by searching in 3d space using camera and scene knowledge. In *ICPR*, Seiten 1–5. IEEE, 2008.
- [MG01] Thomas B. Moeslund und Erik Granum. A Survey of Computer Vision-Based Human Motion Capture. *Computer Vision and Image Understanding*, 81(3):231–268, 2001.
- [MHK06] Thomas B. Moeslund, Adrian Hilton und Volker Krueger. A survey of advances in vision-based human motion capture and analysis. *Computer Vision and Image Understanding*, 104(2):90–126, 2006.
- [MS09] MoCap-Systems. Motion analysis, vicon, simi: Marker based, tracking systems. [www.motionanalysis.com](http://www.motionanalysis.com), [www.vicon.com](http://www.vicon.com), [www.simi.com/en/](http://www.simi.com/en/), 2009.
- [MSH06] Shaou-Gang Miaou, Pei-Hsu Sung und Chia-Yuan Huang. A Customized Human Fall Detection System Using Omni-Camera Images and Personal Information. In *Proceedings of the 1st Distributed Diagnosis and Home Healthcare (D2H2) Conference*, 2006.
- [NCM04] Hammadi Nait-Charif und Stephen J. Mckenna. Activity Summarisation and Fall Detection in a Supportive Home Environment. In *International Conference on Pattern Recognition*, 2004.

- [PSH97] Vladimir I. Pavlovic, Rajeev Sharma und Thomas S. Huang. Visual Interpretation of Hand Gestures for Human-Computer Interaction: A Review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19:677–695, 1997.
- [RAR07] Cucchiara Rita, Prati Andrea und Vezzani Roberto. A multi-camera vision system for fall detection and alarm generation. In *Expert Systems*, 2007.
- [RBK<sup>+</sup>06] B. Rosenhahn, T. Brox, U. G. Kersting, A. W. Smith, J. K. Gurney und R. Klette. A system for marker-less motion capture. *Künstliche Intelligenz (KI)*, 1:45–51, 2006.
- [RDWE91] B. Rogers, F. Dunn, G. Walden und E. Eidelberg. A low-cost, real-time Gait Analysis System, Engineering in Medicine and Biology Society. In *Proceedings of the Annual International Conference of the IEEE*, 1991.
- [RM06] Caroline Rougier und Jean Meunier. Demo : Fall Detection Using 3D Head Trajectory Extracted From a Single Camera Video Sequence. In *First International Workshop on Video Processing for Security (VP4S-06)*, June 7-9, Quebec City, Canada, 2006.
- [RR05] N.M. Robertson und I.D. Reid. Behaviour understanding in video: a combined method. In *Proceedings of the 10th International Conf. on Computer Vision*, Jgg. 1, Seiten 808–814, Beijing, Chine, 2005.
- [SB06] L. Sigal und M. J. Black. . HumanEva: Synchronized Video and Motion Capture Dataset for Evaluation of Articulated Human Motion. Bericht, CS-06-08, Brown University, 2006.
- [SG98] Chris Stauffer und W.E.L. Grimson. Adaptive background mixture models for real-time tracking. 1998.
- [SG00] C. Stauffer und W. E. L. Grimson. Learning patterns of activity using real-time tracking. *Transactions on Pattern Analysis and Machine Intelligence*, 22(8):747–757, 2000.
- [SGW<sup>+</sup>08] J. Spehr, M. Gövercin, S. Winkelbach, E. Steinhagen-Thiessen und F. Wahl. Visual Fall Detection in Home Environments. In *6th Int. Conference of the Int. Soc. for Gerontechnology*, Pisa, Italy, June 2008.
- [SK87] L. Sirovich und M. Kirby. Low-dimensional procedure for the characterization of human faces. *J. Opt. Soc. Am. A*, 4(3):519–524, 1987.
- [SKR<sup>+</sup>04] Kale Rajagopalan Sundaresan, A. Kale, A. N. Rajagopalan, A. Sundaresan, N. Cuntoor, A. Roychowdhury und V. Krueger. Identification of Humans Using Gait. *IEEE Transactions on Image Processing*, 13:1163–1173, 2004.
- [SSR<sup>+</sup>04] L. Sigal, B. Sidharth, S. Roth, M. J. Black und M. Isard. Tracking loose-limbed people. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2004.
- [SSSar] Imran Saleemi, Khurram Hassan Shafique und Mubarak Shah. Probabilistic Modeling of Scene Dynamics for Applications in Visual Surveillance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, To appear.
- [TTW<sup>+</sup>05] Ji Tao, Mukherjee Turjo, Mun-Fei Wong, Mengdi Wang und Yap-Peng Tan. Fall Incidents Detection for Intelligent Video Surveillance. In *Fifth International Conference on Communications and Signal Processing*, 2005.
- [VJ01] Paul Viola und Michael Jones. Robust Real-time Object Detection. In *International Journal of Computer Vision*, 2001.

- [WTM<sup>+</sup>03] Liang Wang, Tieniu Tan, Senior Member, Huazhong Ning und Weiming Hu. Silhouette Analysis-Based Gait Recognition for Human Identification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25:1505–1518, 2003.
- [XG08] Tao Xiang und Shaogang Gong. Video Behavior Profiling for Anomaly Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(5):893–908, 2008.
- [YNC02] ChewYean Yam, Mark S. Nixon und John N. Carter. On the Relationship of Human Walking and Running: Automatic Person Identification by Gait. In *International Conference on Pattern Recognition (ICPR 2002)*, Seiten 287–290, August 2002.