

# FiFi – Steuerung eines FTF durch Gesten- und Personenerkennung

FiFi – Controlling an AGV by detection of gestures and persons

**Andreas Trenkle**  
**Zäzilia Seibold**  
**Thomas Stoll**  
**Kai Furmans**

*Institut für Fördertechnik und Logistiksysteme (IFL)  
Karlsruher Institut für Technologie (KIT)*

**U**m den manuellen Transport in der Intralogistik zu erleichtern, wurde ein Fahrerloses Transportfahrzeug (FTF) entwickelt, das berührungslos vom Bediener gesteuert wird. Die Steuerung erfolgt durch Gesten- und Personenerkennung basierend auf 3D-Daten der Umgebung. Das Paper beschreibt sowohl Zielsetzung, Betriebsarten und Anwendungsmöglichkeiten als auch das Steuerungskonzept der berührungslosen Steuerung und die technische Umsetzung der Plattform. Erste Experimente bestätigen, dass ein Roboter basierend auf 3D-Daten gesteuert werden kann. Verbesserungsmöglichkeiten in der Robustheit werden aufgezeigt.

*[Schlüsselwörter: Gestensteuerung, FTF, Personenverfolgung, Kinect]*

**A**bstract: To facilitate manual transport in intralogistics, an Automated Guided Vehicle (AGV) with a contact-free control has been developed. The vehicle is controlled by recognition of gestures and persons based on 3D data of the environment. The paper describes both objectives, modes and applications as well as the concept of contactless control and the technical implementation of the platform. Initial experiments confirm that a robot can be controlled based on 3D data. Possibilities of improvement in robustness are described.

*[Keywords: gesture control, AGV, Person Following-Robot (PFR), Kinect]*

## 1 MOTIVATION

Der manuelle Transport von Gütern in der Intralogistik mittels Handhubwagen, elektrischem Hubwagen oder Kommissionierwagen ist oft mit körperlichen Anstrengungen verbunden. Insbesondere in Verbindung mit anderen Tätigkeiten – wie dem Be-/Entladen oder dem Kommissionieren – erfordert die manuelle Bewegung des Fahrzeuges die Aufmerksamkeit des Bedieners und kostet

Zeit. Ausgehend von der Fragestellung, wie die körperlichen Anstrengungen verringert und die Ergonomie verbessert werden können, wurden im Rahmen des Forschungsprojekts „FiFi“ zwei Versuchsfahrzeuge entwickelt, die dem Benutzer selbstständig folgen und sich mittels Gesten steuern lassen.

FiFi wurde nach Plug&Work-Prinzipien [Fur10] gestaltet. Daraus folgt, dass FiFi ohne Konfiguration direkt einsetzbar ist. Trotz komplexer Technik zeichnet sich die Interaktion für den Benutzer durch Einfachheit und intuitive Bedienbarkeit aus. Durch die Nutzung mehrerer FiFi-Elemente durch eine Person, erfüllt FiFi die Forderungen nach Skalierbarkeit und Effizienz. Da Performanz nicht auf Kosten der Personensicherheit ausgetragen werden darf, ist FiFi mit Sicherheitsmechanismen ausgestattet um durch Mensch oder Maschine entstandene Gefährdungen abzuwenden.

Nach Darlegung des aktuellen Stands der Forschung im Bereich der Gesten-, Spracherkennung und Personenverfolgung, werden in Abschnitt 3 die Zielsetzung, Betriebsarten und die Anwendungsmöglichkeiten beschrieben. Darauf folgend werden die entwickelten Interaktionskonzepte zwischen Mensch und FiFi dargestellt. Hier wird erläutert, welche Merkmale aus Sprache, Gesten und Bewegungen genutzt werden um Befehle für FiFi zu generieren und wie die Interaktionsmechanismen von FiFi zum Menschen gestaltet sind. In Abschnitt 5 werden die Anforderungen und die Umsetzung von FiFi anhand von zwei realisierten Versuchsfahrzeugen erläutert, bevor mit einem Ausblick auf zukünftige Forschungsfragen abgeschlossen wird.

## 2 STAND DER FORSCHUNG

Im Folgenden werden Forschungsprojekte beschrieben, die sich mit der Steuerung durch Personenverfolgung oder Gesten- und Spracherkennung beschäftigen. Wäh-

rend bei der Personenverfolgung aus Position und Bewegung des Benutzers Befehle generiert werden, nutzt die Gestensteuerung einzelne Bewegungen der Gliedmaßen zur Befehlserzeugung. Die Steuerung durch Spracherkennung wird hier nur kurz behandelt. Die Verfahren zur Spracheingabe sind – vor allem durch die Konsumgüterindustrie – schon weit entwickelt und kommen bei FiFi nur zusätzlich zu der hauptsächlichen Steuerung durch Personen- und Gestenerkennung zum Einsatz.

## 2.1 PERSONENVERFOLGUNG

Die Personenverfolgung ist eine in der Robotik angesiedelte Disziplin. „Person Following Robots“ (PFR) erfassen die Umgebung mittels Sensoren, interpretieren die Rohdaten mittels Algorithmen und generieren daraus Steuerbefehle.

Der in [Kob06] vorgestellte PFR (Abbildung 1) nutzt eine Segwayplattform als Fahrantrieb, verarbeitet 2D-Daten eines Laserscanners und Videodaten einer omnidirektionalen Kamera. Im Rahmen des Forschungsprojektes wurde die Verfolgung von Personen im Außenbereich bei Geschwindigkeiten von bis zu 2,3 m/s realisiert.



Abbildung 1. PFR mit Segway-Plattform [Kob06]

Der in (Abbildung 2) dargestellte PFR nutzt Algorithmen zur Verkürzung des Fahrwegs bei Personenverfolgung durch einen Hindernisparcours [Doi12]. Die Kinect ist durch einen zusätzlichen Mechanismus horizontal und vertikal schwenkbar und kann Personen dadurch ohne Fahrbewegungen optisch verfolgen. Bei Vergrößerung des Abstandes führt das Fahrzeug Fahrbewegungen zur Verfolgung aus. Die Kinect wird zur Personenverfolgung und ein Laserscanner zur Lokalisierung bzw. zur A-priori-Erstellung einer Umgebungskarte genutzt. Läuft die Person einen Umweg, kürzt das Fahrzeug ab. In einem Hindernisparcours wurde mit einem kombinierten Verfahren zwischen Personenverfolgung und Pfadplanung eine Verkürzung der Pfade um 29,6 % erreicht.

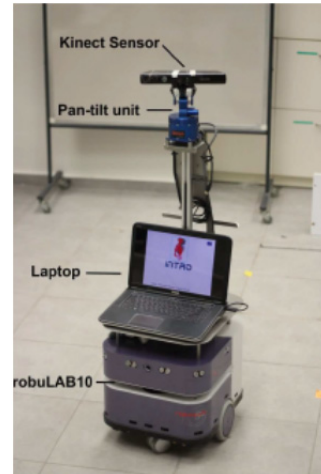


Abbildung 2. PFR für Experimente hinsichtlich Verkürzung des Fahrweges [Doi12]

Im Gegensatz zu den vorgestellten PFR soll FiFi in Anwendungen der Intralogistik zum Einsatz kommen, woraus sich weitere Anforderungen ergeben. Durch den Transport von Ladungsträgern sind beispielsweise die Montagepositionen der Kamera eingeschränkt. Außerdem muss die Personenverfolgung die Be- und Entladung erlauben. Bei manuellen Tätigkeiten spielt der Durchsatz eine wichtige Rolle weshalb die Gestensteuerung schnell und robust funktionieren muss.

## 2.2 GESTENSTEUERUNG

Bei der Gestensteuerung ist neben den Verfahren zur Erkennung auch die Definition der Gesten entscheidend. Gesten müssen zum einen intuitiv und zum anderen in der jeweiligen Einsatzumgebung nutzbar zu sein.

Gestensteuerung wird in zahlreichen Forschungsprojekten der Robotik und Anthropomatik untersucht. In [Abi13] wird eine Kombination zwischen Gesten- und Sprachsteuerung zur Interaktion mit einem Serviceroboter im Heim- und Healthcare-Bereich vorgestellt. Eine andere Anwendung ist die TV-Steuerung mittels Handgesten, bei der sowohl Handbewegungen als auch Handposen genutzt werden [Shi13].

[Lam11] stellt Ansätze zur Programmierung von Industrierobotern durch Gesten vor. Dabei werden die Bewegungen des Bedieners vom Industrieroboter durch synchrones Tracking direkt nachgeahmt oder mittels asynchronem Tracking abgespeichert und im späteren Programmablauf verwendet.

Bei den vorgestellten Systemen kann jede Bewegung des Benutzers interpretiert und umgesetzt werden – FiFi birgt die Herausforderung, Gesten von natürlichen Bewegungen im Arbeitsablauf zu unterscheiden.

## 2.3 SPRACHSTEUERUNG

Verfahren zur nutzerunabhängigen Spracherkennung sind bereits in zahlreichen Konsumgütern wie Smartphones oder Kraftfahrzeugen vorhanden. Entscheidend für die Einsetzbarkeit von Sprachsteuerung ist die Auswahl der Befehle. Neben der Erkennung von Sprache bietet die Kinect auch die Möglichkeit, die Richtung zu bestimmen, aus der die Sprache kommt. Hierfür wird ein Array von Mikrofonen genutzt.

Die Kombination von 3D-Sensordaten ermöglicht die Reduktion von Hintergrundgeräuschen und die Verbesserung der Spracherkennung durch Zuordnung der Äußerungen zur betreffenden Person [Gal12].

Da die Qualität der Spracherkennung vom Geräuschpegel der Umgebung abhängt, kann bei FiFi jeder durch Sprache geäußerte Befehl auch durch eine Geste artikuliert werden, weshalb die Sprachsteuerung keine Kernfunktion darstellt.

## 3 ZIELSETZUNG UND ANWENDUNGSMÖGLICHKEITEN

Wie in der Einleitung erläutert, zielt FiFi darauf ab, den Bediener bei manuellen Transporten sowohl körperlich als auch geistig zu entlasten. Die körperliche Entlastung geschieht durch die Substitution der Deichsel mit einem Kamerasystem. Das Ziel der mentalen Entlastung soll durch die Reduzierung der Aufmerksamkeit, die für das Transportsystem nötig ist, erreicht werden.

Zur Erreichung dieser Ziele wurden die nachfolgend dargelegten Betriebsarten und Funktionen entwickelt, woraus beispielhaft Anwendungen abgeleitet wurden.

### 3.1 BETRIEBSARTEN UND -FUNKTIONEN

Um den Einsatz von FiFi flexibel und effizient zu ge-

stalten, wurden vier unterschiedliche Betriebsarten definiert (siehe Abbildung 3).

Der Benutzer kann nach Anmeldung je nach Bedarf zwischen den Betriebsarten wechseln.

Im **Folgemodus** ist vor allem die Personenerkennung aktiv. FiFi folgt einem sich bewegenden Nutzer indem es einen konstanten Abstand zur Person hält. Der Nutzer kann sich bewegen ohne auf FiFi zu achten und kann auch andere Tätigkeiten ausführen. Innerhalb des Folgemodus gibt es die Übergabefunktion. Sobald FiFi den Folgeabstand erreicht hat und der Benutzer sich zwei Sekunden nicht bewegt, nähert sich FiFi ihm an, um eine Kistenübergabe realisierbar zu machen.

Der **Rangiermodus** ermöglicht eine Feinpositionierung z.B. zur Aufnahme oder Abgabe von Ladungsträgern. Hierbei wird statt der Personen- die Gestenerkennung eingesetzt. Als Referenz werden die Hände des Benutzers verwendet.

Der **Linienfolgemodus** ist für längere Transportstrecken geeignet, in denen keine Interaktion mit dem Nutzer stattfinden. Wenn der Nutzer FiFi an die Leitlinie abgibt, folgt es dieser selbstständig unter Nutzung einer Spurführungskamera.

Um größere Warenmengen zu transportieren, können im **Clustermodus** mehrere Fahrzeuge gleichzeitig dem Bediener folgen. Dabei folgt das erste Fahrzeug dem Bediener im Folgemodus und alle darauffolgenden dem Vorgängerfahrzeug.

Die **Hubfunktion** erlaubt das ergonomische Anheben der transportierten Kisten. Dabei soll sich das Fahrzeug zum einen selbstständig an die Größe des Bedieners anpassen. Zum anderen ist aber auch eine aktive Ansteuerung des Hubes durch den Benutzer mit Gesten möglich.

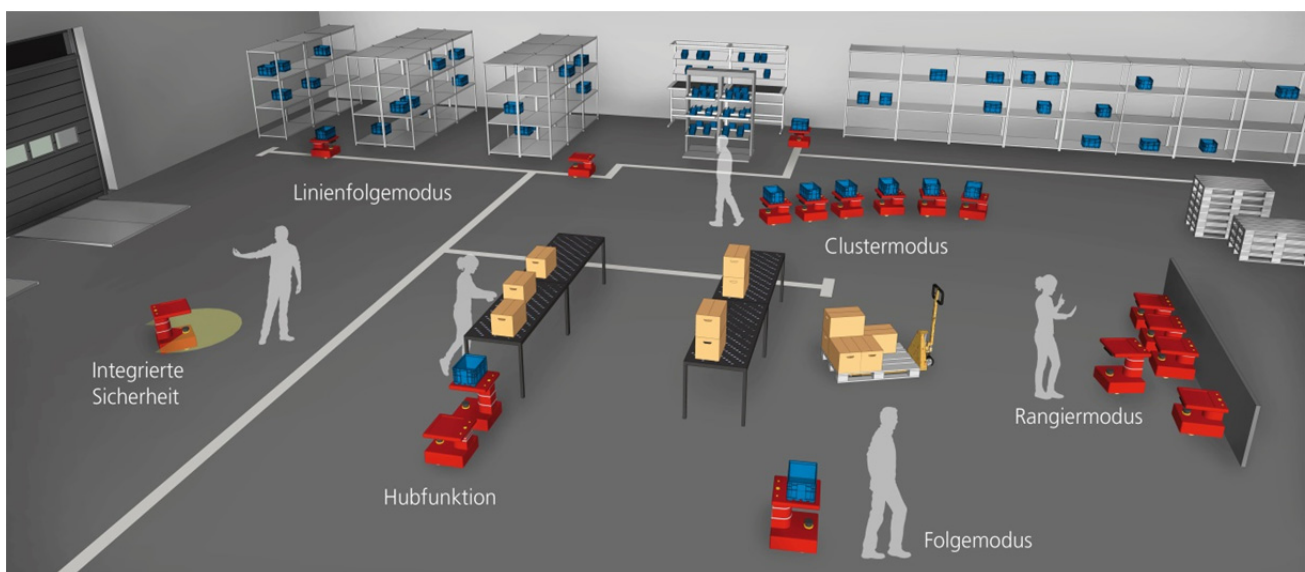


Abbildung 3. Betriebsarten und Funktionen von FiFi

**Integrierte Sicherheit** ist notwendig, weil FiFi aufgrund der Personenverfolgung und der direkten Interaktion ständig im Kontakt mit Personen ist. Daraus ergeben sich Risiken, die entstehen, wenn eines der Teilsysteme versagt.

### 3.2 ANWENDUNGEN UND NUTZEN

Durch Kombination der Betriebsarten können die Funktionen von FiFi zusätzlich zum realisierten Einzeltransport auch in anderen Anwendungen eingesetzt werden.

Eine Anwendung ist die manuelle Steuerung von Fahrerlosen Transportfahrzeugen (FTF) bei Inbetriebnahme und Konfiguration. Durch das Ablaufen der später zu fahrenden Fahrstrecke lernt das Fahrzeug durch Verfolgung der Person neue Wege und Übergabepositionen. Der Vorteil liegt in der intuitiven Bedienung ohne lange Einlernzeiten.

Bei manueller Kommissionierung ermöglicht FiFi dem Bediener die Fokussierung seiner Aufmerksamkeit auf den eigentlichen Kommissionierprozess, da das Fahrzeug selbstständig folgt. Hierbei nutzt er den Folgemodus um sich durch die Regalgassen von FiFi verfolgen zu lassen. Bleibt der Bediener stehen, nähert sich FiFi im Übergabemodus, damit die Ware aufgelegt bzw. abgenommen werden kann. Sind alle Kommissionieraufgaben erledigt, wird der Linienfolgemodus aktiviert. FiFi fährt dann selbstständig in den nächstgelagerten Bereich, z.B. der Verpackung und ist an der Endposition der Leitlinie wieder für die Bedienung durch einen anderen Benutzer bereit. FiFi bietet hier neben den Vorteilen der körperlichen und mentalen Erleichterung beim Kommissionieren durch die Verknüpfung von verschiedenen Arbeitsbereichen mittels Linienfolgemodus den Nutzen von deutlich verkürzten Laufstrecken.

## 4 INTERAKTIONSKONZEPTE

Das Interaktionskonzept umfasst die in den verschiedenen Modalitäten zwischen Mensch und FiFi ausgetauschten Signale (Abbildung 4). Der Mensch interagiert mit FiFi über Sprache, Gesten und seine Laufbewegungen. Von FiFi gehen optische Lichtsignale und Audio-Signale aus. Auch durch die Fahrbewegungen findet eine implizite Interaktion statt, da der Bediener anhand des Verhaltens den Rückschlüsse über den Zustand ziehen kann.

Aufbauend auf den nachfolgend dargelegten Kriterien für Interaktionsmechanismen werden für die bereits beschriebenen Betriebsarten- und Funktionen Interaktionsmechanismen erarbeitet.

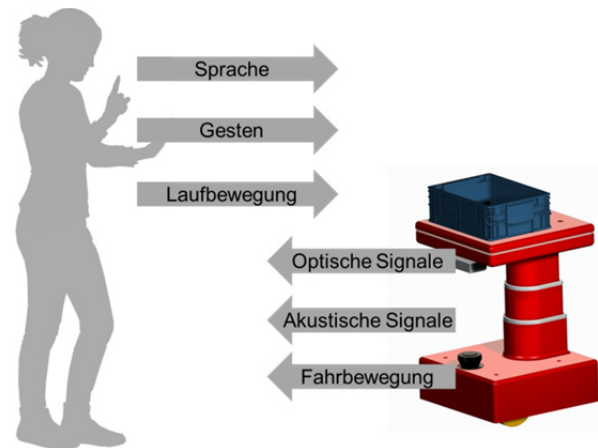


Abbildung 4. Interaktion zwischen Bediener und FiFi

### 4.1 DEFINITION DER INTERAKTIONSMECHANISMEN

Gesten und Sprache sind vom Menschen explizit durchgeführte Tätigkeiten, mit dem direkten Ziel der Steuerung von FiFi. Das Vorausgehen kann hingegen als implizite Tätigkeit betrachtet werden.

Die Gestenerkennung setzt die Definition eines geeigneten Gestenvokabulars voraus, das die Menge der von der Benutzerschnittstelle zu erkennenden Gesten darstellt. Zur Definition des Gestenvokabulars existieren zwei Ansätze: Es kann entweder ein Systemorientiertes oder ein Benutzerorientiertes Vokabular definiert werden [Nie92]. Bei beiden Ansätzen ist die Eindeutigkeit der Gesten erforderlich, damit sie vom Erkennungssystem unterschieden werden können.

Der systemorientierte Ansatz wird hat zum Ziel, für das Erkennungssystem technisch gut erkennbare Gesten zu definieren. Solche Gesten können für beliebige Anwendungen genutzt werden, haben dann jedoch keinen semantischen Bezug zur eigentlichen Tätigkeit. Der Systemorientierte Ansatz bietet den Vorteil der einfachen Erkennung, da für das System eindeutige Gesten definiert werden. Nachteilig ist jedoch, dass die Gesten keinen direkten Bezug zum Befehl haben, dadurch nicht intuitiv sind und somit vom Benutzer schwerer zu merken und durchzuführen sind [Nie04]. Abhängig vom Kontext müssen Gesten so gewählt werden, dass der Benutzer die Gesten nicht bereits bei seiner Tätigkeiten durchführt um Falsch-Positiv Erkennungen zu vermeiden.

Beim benutzerorientierten Ansatz wird die Nutzerschnittstelle nach [Nie92] mit neun Gestaltungsrichtlinien bewertet:

- Einfache und natürliche Dialoge
- Nutzung der Sprache des Benutzers
- Minimale Beanspruchung des Gedächtnisses des Benutzers



- Konsistenz
- Rückgabe von Feedback
- Ausstiegs- und Abbruchmöglichkeiten
- Angebot von Abkürzungen
- Gute Fehlermeldungen
- Vermeidung von Fehlern

Der Kern des Ansatzes besteht darin, dass Gesten intuitiv, einfach zu merken und nachvollziehbar sind, einen metaphorischen und semantischen Bezug zur Funktionalität haben und ergonomisch sind. Hierdurch werden die Zeit und der Einlernaufwand reduziert, wodurch die effiziente Nutzung der Nutzerschnittstelle möglich ist. Hierbei muss die aus dem systemorientierte Ansatz geforderte Eigenschaft der eindeutigen Erkennung der Gesten gegeben sein [Nie04].

Für FiFi wurden die nachfolgend beschriebenen Gesten nach dem Benutzerorientierten Ansatz definiert. Die Abbildungen zeigen die Tiefenbilder der 3D-Kamera sowie die segmentierten Skelettdaten.

#### 4.1.1 ANMELDUNG DES BENUTZERS

Bevor eine Anmeldegeste erfolgt, werden keine Befehle ausgeführt. Die Erkennung von Personen ist im Wartezustand aktiv. Sobald eine Person erkannt wird, wird dies optisch angezeigt, um dem Nutzer zu signalisieren, dass nun eine Anmeldung möglich ist. Befinden sich mehrere Personen im Sichtbereich der Kamera, werden ausschließlich Anmeldegesten von der am nächsten beim Fahrzeug erkannten Person akzeptiert. Nach Ausführung der Anmeldegeste werden ein optisches und ein akustisches Signal ausgegeben, um dem Benutzer die erfolgreiche Anmeldung zu bestätigen. Die Anmeldung ist mit einem Wechsel in den Folgemodus verknüpft, der Grundlage für die Aktivierung aller weiteren Modi ist. Nach Ausführung der Abmeldegeste wird der Benutzer abgemeldet wodurch – außer der erneuten Anmeldegeste – keine Gesten mehr verarbeitet werden.

#### Gesten zur An- und Abmeldung:

Die An- bzw. Abmeldung erfolgt durch dreimaliges Winken mit der rechten bzw. der linken Hand (Abbildung 5). Die Winkbewegung erfolgt mit Unterarm und Hand. Ein Winken mit dem gesamten Arm oder nur der Hand wird ignoriert. Die Geste hat durch das Winken semantischen Bezug zur Funktion der An- und Abmeldung. Die dreimalige Durchführung der Geste ermöglicht den vom systemorientierten Ansatz geforderte Robustheit der Erkennung sowie die Vermeidung von Fehlerkennungen. Der Sprachbefehl „FiFi Abmelden“ führt zum gleichen Ergebnis.

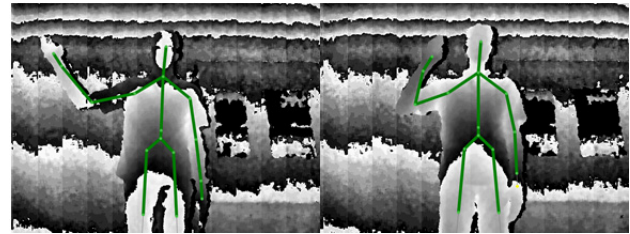


Abbildung 5. Start- und Endposition der Anmeldegeste

#### 4.1.2 HUBFUNKTION

Die Hubvorrichtung kann im Folgemodus angehoben und abgesenkt werden.

#### Geste zur Anhebung und Absenkung des Hubs:

Durch das Öffnen der Arme wird der Hub angehoben, durch das Schließen abgesenkt (Abbildung 6). Beide Unterarme werden aus der Grundstellung mit Ausrichtung nach vorne um jeweils 90 Grad nach außen bewegt um den Hub anzuheben. Bei Vertauschung von Start- und Endgeste wird der Hub abgesenkt.

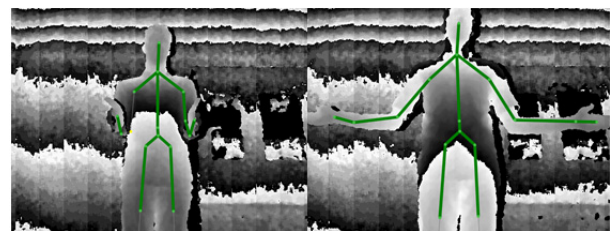


Abbildung 6. Start- und Endposition der Geste zur Anhebung des Hubs

Um Fehlerkennungen während des Betriebes zu vermeiden, wird auf eine Geste mit direkterem semantischem Bezug – wie beispielsweise das Anheben oder Absenken einer Hand – verzichtet. Stattdessen wird durch eine robustere Geste ein indirekter Bezug durch die Verknüpfung von Heben mit öffnen und Senken mit schließen hergestellt.

#### 4.1.3 FOLGEMODUS

Im Folgemodus hält FiFi einen konstanten Abstand zur Person. Er ist durch den begrenzten Öffnungswinkel der Kamera auf einen Soll-Abstand von ca. 1,60 cm festgelegt, da bei kleineren Abständen keine Erfassung der gesamten Person möglich ist. Nach Ausführung der Anmeldegeste regelt FiFi den Abstand und den Winkel zum Bediener durch Fahrbewegungen. Abhängig von der Anwendung kann FiFi entweder auf bis zu 4 m entfernte Anmeldungen reagieren und zum Benutzer fahren oder der Nutzer muss sich in einem definierten Bereich vor FiFi aufhalten, um die Anmeldung durchführen zu können. Als Referenzpunkt für den Folgemodus wurde der Beckenmittelpunkt festgelegt, da er in verschiedenen Entfernungen im Sichtbereich der Kamera liegt und bei Laufbewegungen dem Laufweg entspricht. Außerdem wird er

bei manuellen Tätigkeiten, z.B. dem Griff in ein Regal, im Vergleich zu anderen Gliedmaßen nur wenig bewegt, wodurch unnötige Bewegungen von FiFi vermieden werden.

Neben einer visuellen Signalisierung über den aktiven Folgemodus wird ein akustisches Warnsignal ausgegeben, wenn der Benutzer sich seitlich aus dem Sichtbereich bewegt oder sich zu schnell vom Fahrzeug entfernt. Damit kann sich der Nutzer frei bewegen, ohne sich umschauen zu müssen, ob FiFi noch ordnungsgemäß hinterherfährt.

Bewegt sich die Person trotz der Warnhinweise aus dem Sichtbereich der 3D-Kamera, bleibt FiFi stehen und wartet auf eine erneute Anmeldegeste.

#### 4.1.4 ÜBERGABEFUNKTION

Eine direkte Aufnahme oder Abgabe wird durch die Übergabefunktion ermöglicht. Sie wird aktiviert, sobald FiFi den Folgeabstand erreicht hat und die Person sich für zwei Sekunden nicht bewegt. Dann fährt FiFi auf einen Abstand von 40cm an die Person heran, wodurch die Aufstellfläche zum Be- und Entladen zugänglich gemacht wird. Bedingt durch den geringen Abstand können hierbei keine Körpermerkmale erfasst werden. Während der Übergabe wird deshalb nicht mehr auf Gesten reagiert.

Da bei diesem Abstand nicht das gesamte Skelet erfasst wird, wird bei Aktivierung des Übergabemodus während der Übergabe ein Abstandswert in der Mitte der Person verfolgt. Während sich FiFi annähert, muss sich dieser Abstandswert entsprechend der Fahrgeschwindigkeit stetig verringern. Ist der Übergabevorgang abgeschlossen vergrößert die Person den Abstand wobei der Wertverlauf auf Stetigkeit geprüft wird. Zusammenhängende Punkte mit dem Abstandswert dürfen nicht seitlich verschwinden, da die Person dann das Sichtfeld verlassen hätte. Wird bei vergrößertem Abstand aus den Punkten mit den verfolgten Abstandswerten wieder eine Person erkannt, können diese der Person zugeordnet werden. Dadurch ist sichergestellt, dass die angemeldete Person den Sichtbereich nicht verlassen hat und eine erneute Anmeldegeste für den dann aktivierten Folgemodus entfällt. Während der Übergabefunktion wird der Nutzer durch die Sprachausgabe von „Bitte zurücktreten“ aufgefordert, den Abstand zu vergrößern, bis das Skelett wieder erkannt wird. Die Sprachausgabe hat hier den Vorteil, dass die Person auch von FiFi weglaufen kann ohne Sichtkontakt halten zu müssen.

Alternativ kann FiFi durch den Sprachbefehl „FiFi Warte“ in einen Wartezustand versetzt werden. Die Personenverfolgung ist dann nicht mehr aktiv bis die Person wieder in den Sichtbereich der Kamera tritt und die Anmeldegeste ausführt oder den Sprachbefehl „FiFi Weiter“ äußert.

#### 4.1.5 RANGIERMODUS

Um FiFi exakt zu positionieren, wird der Rangiermodus eingesetzt. Dabei werden statt dem Becken die Handflächen als Referenz genutzt (Abbildung 7). In diesem Modus sind die Reaktionstoleranzen minimal, wodurch bereits geringe Bewegungen der Handflächen in Fahrbefehle umgewandelt werden.



Abbildung 7. FiFi und Bediener im Rangiermodus

Bei Aktivierung des Rangiermodus wird die Winkelposition der Hände nullgesetzt. Dadurch kann der Bediener seine Position frei wählen, dann die Hände heben und beispielsweise seitlich an FiFi vorbeischaun um bei Rückwärtsfahrt die Fahrstrecke genau einsehen zu können.

#### Geste zur Aktivierung des Rangiermodus:

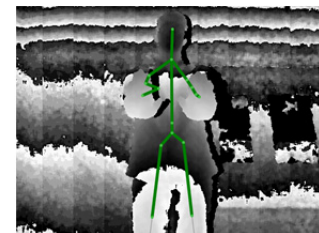


Abbildung 8. Geste zum Starten des Rangiermodus

Der Rangiermodus wird durch das Ausstrecken der Arme mit den Handflächen nach vorne gestartet (Abbildung 8).

Durch das Heben der Handflächen ist ein metaphorischer Bezug zur Bedienung eines Handhubwagens gegeben. Da die Geste in üblichen Arbeitsprozessen nicht auftritt, kann sie nach der kurzen Ausführungszeit von einer Sekunde direkt in Fahrbefehle umgesetzt werden.

#### 4.1.6 LINIENFOLGEMODUS

Im Linienfolgemodus folgt FiFi unter Nutzung einer Spurführungskamera selbstständig einer Leitlinie. Durch eine Geste des Bedieners wird die Spurführungskamera aktiviert. Nun bewegt der Bediener FiFi über die Leitlinie, die – sobald von der Kamera erfasst – verfolgt wird.

### Geste zur Aktivierung des Linienfolgmodus:

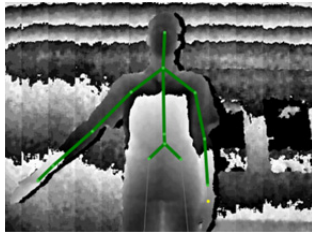


Abbildung 9. Geste zur Aktivierung des Linienfolgmodus

Die Geste zur Aktivierung des Linienfolgmodus (Abbildung 9) wird durch das Zeigen auf den Boden mit dem gestreckten Arm durchgeführt. Das Ausstrecken des Armes in Richtung Boden ist eine benutzerorientierte Linie, da das Zeigen mit dem Arm auch im Alltag als Aufforderung zur Zielverfolgung verwendet wird..

#### 4.1.7 PERSONENSICHERHEIT

Beim Transport großer Lasten geht bei Geschwindigkeiten von bis zu 1,6 m/s ein hohes Risiko von der Kollision mit Personen aus. Da die Komponenten, die die Befehle generieren und verarbeiten – 3D-Kamera, Rechner, SPS und Motoren – keine Sicherheitsbauteile sind, muss davon ausgegangen werden, dass sie Fehlzustände erzeugen können, die gefahrbringend sind.

Maßnahmen zur Verbesserung der Robustheit bei der Gestenerkennung und Personenverfolgung sind sicherheitstechnisch nicht verwertbar, weshalb zur Vermeidung von Kollisionen mit Personen ein Sicherheitslaserscanner zum Einsatz kommt. Die unterschiedlich großen Schutzfelder werden durch sichere Drehgeber umgeschaltet. Hierdurch kann die Schutzfeldgröße bei geringen Geschwindigkeiten reduziert werden. Bei Verletzung des Schutzfeldes durch Objekte oder Personen werden die Antriebe stillgesetzt. Sollte FiFi aufgrund der Falschinterpretation der 3D-Daten eine unerwünschte Bewegung ausführen, wird diese durch den Laserscanner gestoppt, bevor es zu einer Kollision kommen kann.

## 5 REALISIERUNG VON FiFi

Zur Entwicklung und Untersuchung der Verfahren zur Erkennung und Verfolgung von Personen wurden die Anforderungen an den Versuchsträger definiert und Konzepte für mechanische und elektrische Komponenten erarbeitet.

### 5.1 TECHNISCHE ANFORDERUNGEN

Als Grundlage für die Entwicklung und den Test der Algorithmen zur Personenverfolgung wurde eine Fahrzeugplattform entwickelt, für die folgende Anforderungen definiert wurden:

- Aufstellfläche zum Transport von Kleinladungsträgern (300mm x 400mm) mit einem Gewicht von 20kg mit Sicherung gegen Herabfallen und Verrutschen
- Bauraum für Rechner im oberen und Motorsteuerung im unteren Teil der Plattform. Nutzung des unteren Teils der Plattform (Chassis) auch für andere Projekte möglich
- Zwei Montagepunkte für die Anbringung des Kamerasystems an verschiedenen Positionen des Fahrzeuges zur Untersuchung der Qualität der Gesten- und Personenerfassung auf unterschiedlicher Montagehöhe
- Fahrgeschwindigkeit von 1,6 m/s um Personen in Schrittgeschwindigkeit folgen zu können
- Berührungslos wirkende Schutzvorrichtung (Laserscanner) zur Vermeidung von Kollisionen mit Personen und Hindernissen
- Not-Halt Gerät für den sicheren Stopp (EN ISO 13850)
- Hubvorrichtung zum Anheben/Absenken des Förderguts auf Arbeitshöhe
- Abgerundetes, das gesamte Fahrzeug umschließendes, robustes Gehäuse. Vermeidung von scharfen Kanten und Quetschstellen (insbes. beim Hub)
- Einrichtung zur visuellen und akustischen Darstellung des Systemzustandes
- Kommunikation über WLAN zur Diagnose und zur manuellen Steuerung über Tablet-Computer
- Akkulaufzeit von mindestens zwei Stunden
- Berücksichtigung der Wartbarkeit: Steckbarer Kabelbaum und einfacher und schneller Austausch von Komponenten

### 5.2 BAU DER PLATTFORM

Das sich selbst tragende Chassis (Abbildung 10) besteht aus fünf gekanteten und gelaserten Blechen. Sie sind mittels Setzmuttern und Schrauben miteinander verbunden sind, was ein Aufbau der steifen Gesamtstruktur ohne Schweißverbindungen ermöglicht. Die beiden Hauptteile der Blechbiegekonstruktion können zur Montage auseinandergezogen werden, wodurch Motorsteuerung, Antriebe, Akku und Elektroverkabelung gut zugänglich sind. Aufgrund des Gewichtes wurden die Bleiakkus im Chassis mittig angebracht, um einen niedrigen Schwerpunkt sicherzustellen und Standsicherheit zu gewährleisten.



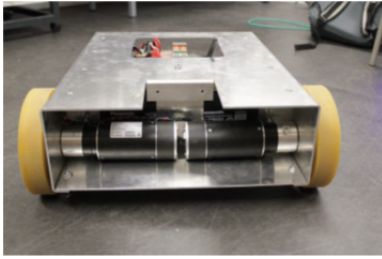


Abbildung 10. Chassis des FTF

Als Antrieb wurde ein differenzieller Radseitenantrieb mit passivem Kastorrad realisiert. Hierdurch ist kein weiterer Lenkantrieb nötig und die Antriebe können ohne zusätzliche Halterung direkt am Chassis angebracht werden. Durch eine auf das Chassis montierte Hubsäule kann die Aufstellfläche für Kleinladungsträger auf ergonomische Höhe gebracht werden.

Die Anbindung der elektrischen Komponenten an die Energieversorgung erfolgt über eine Hutschiene. Hierbei werden durch einen separaten Not-Halt-Kreis bei Betätigung des Not-Halt Gerätes oder Verletzung der Laserscanner-Schutzbereiche die Antriebe stillgesetzt.

Die Verkleidung der Plattform wurde aus glasfaserverstärktem Kunststoff (GFK) hergestellt. Hierfür wurde eine Negativform gefräst, die mittels Handlaminierverfahren ausgekleidet wurde.



Abbildung 11. FiFi mit ein- und ausgefahrener Hubsäule

Um dem Benutzer Rückmeldung über den Betriebsmodus bzw. den Status der Gesten und Personenerkennung zu geben, wurde im oberen Teil des Fahrzeuges eine umlaufende Lichtleiste (Abbildung 11) angebracht. Zur Abgabe von akustischen Warnsignalen und zur Sprachausgabe wurden Lautsprecher integriert. Zur Diagnose und Wartung kann FiFi über eine mobile Mensch-Maschine-Schnittstelle (MMS) bedient werden.

Zur Erfassung der Personen und Gesten wird eine Kinect-Kamera genutzt, die unterhalb des Hubtisches angebracht wurde, um Verdeckungen durch Ladungsträger zu vermeiden. Sie liefert neben einem VGA-Videobild 3D-Daten mit einer Auflösung von 640 x 480 Pixeln mit

einer Bildrate von 30 fps. Die Pixel des Tiefenbildes werden jeweils durch 11 Bit dargestellt, woraus sich ein Bereich für die Tiefeninformationen von 2048 Werten ergibt. Die Kamera erfasst Tiefeninformationen in einem Abstandsbereich von 0,5m bis 5m. Nach [Kho10] beträgt die Tiefenaufklärung bei einem Abstand von 5m lediglich 7 cm. Der optimale Bereich für die Erfassung von Gesten liegt zwischen 1m und 3m. Dies ergibt sich aus der relativ groben Infrarotmuster und den eingeschränkten Sichtbereichen von horizontal 57° und vertikal 43°. Die Tiefeninformationen werden unter Einsatz eines CMOS-Sensors generiert, der Infrarotlicht in einem definierten Wellenlängenbereich erfassen kann, das von einem Infrarot-Muster Laserstrahler emittiert wird. Mittels einer Variante der passiven Stereo-Triangulation, der Stereometrie mit strukturiertem Licht wird die 3D-Szene rekonstruiert [Bes88]. Ein Vorteil gegenüber Stereokameras ist die Projektion mit Infrarotlicht, wodurch auch bei schlechten Lichtverhältnissen eine Erfassung möglich ist. Bei Sonneneinstrahlung auf die Szene oder den Sensor verschlechtert sich die Auflösung – bei hoher Intensität ist keine Tieferfassung mehr möglich. Ein Mikrofonarray bestehend aus vier Mikrofonen liefert einen Winkel für die Geräuschquelle und kann Hintergrundgeräusche ausmaskieren.

Zum Schutz vor Kollisionen mit Personen kommt ein Sicherheitslaserscanner zum Einsatz. Er erlaubt die Konfiguration von Schutzfeldern, und löst den Sicheren Stopp bei Verletzung eines Schutzfeldes aus. Der Laserscanner lässt sich an sichere Drehgeber anbinden, wodurch eine geschwindigkeitsabhängige Schutzfeldumschaltung realisiert werden kann.

Neben einer SPS, die die beiden Antriebe mittels CAN ansteuert, kommt ein Computer zum Einsatz, auf dem die Verarbeitung der 3D-Kamera-Daten stattfindet. Die vom PC generierten Steuerdaten werden an die SPS weitergeleitet.

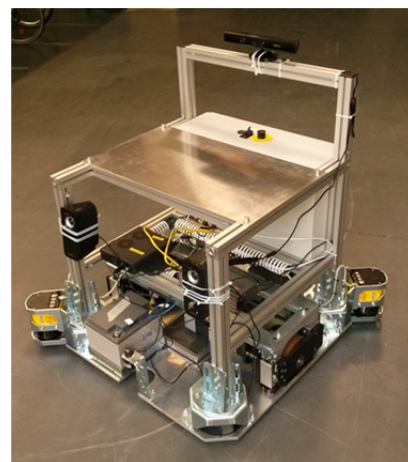


Abbildung 12. Versuchsträger zum Transport von Lasten bis zu 250 kg



Ein zweites Fahrzeug mit einer Traglast von 250 kg wurde vom Industriepartner BÄR Automation GmbH entwickelt und aufgebaut (Abbildung 12). Im Unterschied zum oben beschriebenen Versuchsträger ist es mit zwei Sicherheitslaserscannern ausgestattet, was eine sichere Rückwärtsfahrt ermöglicht. Durch eine Kamera zur optischen Spurführung kann der Linienfolgemodus realisiert werden. Die Position der Kamera auf dem zweiten Versuchsfahrzeug erlaubt – bedingt durch das größere Sichtfeld bei gleichem Abstand zum Fahrzeug – eine robustere Übernahme von Ladungsträgern und eine kürzere Verfolgungsdistanz.

### 5.3 IMPLEMENTIERUNG UND HERAUSFORDERUNGEN

Die Software für die Gestenerkennung und Personenverfolgung wurde mittels Kinect SDK von Microsoft entwickelt. Der Folge- und der Schiebemodus wurden implementiert und – bisher lediglich nach subjektiven Kriterien – getestet. Neue Bediener konnten FiFi ohne Übung auf dem Versuchsfeld direkt benutzen. In einer ersten Untersuchung kann man also vermuten, dass zum einen die Kinect nicht auf den Bediener eingelernt werden muss, und dass auch die Bedienung intuitiv ist. Hinsichtlich der Robustheit der berührungslosen Steuerung lassen sich folgende Aspekte beobachten:

Bei den Sprachbefehlen wurde die Einschränkung eingeführt, dass sie nur dann umgesetzt, wenn die Richtungsinformation der Sprache von der durch die Kinect erfassten Bedienerposition weniger als 25° abweicht. Der relativ große Winkel ergibt sich aus der Ungenauigkeit der durch die Mikrofonarrays generierten Richtungsinformation. Der Sprachbefehl „Stopp“ wird aus jeder Richtung und ohne Aktivierungswort akzeptiert. Die Funktion der Spracherkennung hängt stark von den Umgebungsgeräuschen ab.

Auch das Vorhandensein von Hindernissen stellt eine Herausforderung dar. Hohe Hindernisse, wie z.B. Wände, verdecken den Sichtbereich der Kinect. Wird die Person nicht mehr gesehen, bleibt FiFi stehen. Hindernisse von geringerer Höhe verdecken zwar nicht den Sichtbereich, erschweren jedoch die Kurvenfahrt. Da FiFi immer versucht, direkt in Richtung der Person zu navigieren, schneidet es Kurven. In diesem Falle werden die Hindernisse vom Laserscanner erkannt und lösen den sicheren Stopp aus. Bei der Fahrt durch Bereiche mit Hindernissen ist somit die Aufmerksamkeit des Nutzers gefordert, der – wie beim Ziehen eines Hubwagens – den Nachlaufweg beachten muss.

Auch ist die 3D-Erfassung anfällig für Fehlerkennungen, was Auswirkung auf die Definition der Gesten hat. So werden z.B. manche Gegenstände als Personen erkannt, wie die Palette in Abbildung 13. Auch Bewegungen im Arbeitsablauf des Bedieners können als Befehle erkannt. Zur Vermeidung wurden die Gesten, die zu schnellen Fahrbewegungen führen, komplexer gestaltet.

Die Anmeldegeste wurde deswegen z.B. auf drei Winkbewegungen festgelegt, da die Wahrscheinlichkeit eine dreifache gleichartige Bewegung in einem Gegenstand oder Arbeitsablauf zu erkennen gering ist.

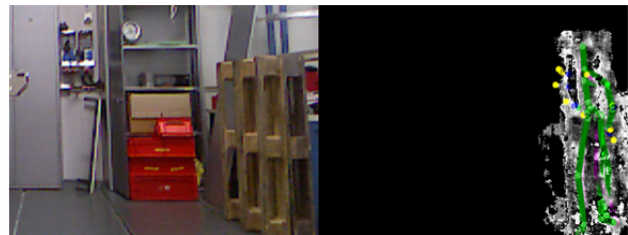


Abbildung 13. Falsch-Erkennung einer Palette als Person

## 6 ZUSAMMENFASSUNG UND AUSBLICK

Aufgrund der entwickelten Betriebsarten und Interaktionskonzepte konnten der Folgemodus und der Rangiermodus wie konzipiert umgesetzt werden. Die Robustheit der Personenverfolgung in Bereichen ohne Hindernisse ist gegeben. Im weiteren Verlauf des Forschungsprojektes sollen, als Basis der Weiterentwicklung der Verfahren, objektive Kriterien zur Bewertung des Folgeverhaltens und der Gestensteuerung erarbeitet werden.

Eine Herausforderung beim Clustermodus ist die Erkennung der Vorgängerfahrzeuge. Hier bietet sich die Kombination von 3D- und Videodaten an. Es sollen eindeutige Merkmale der Vorgänger wie z.B. die Farbe des Fahrzeuges und die LED-Signale zusammen mit den 3D-Daten genutzt werden.

Die Anpassung dieser Betriebsarten auf intralogistische Anwendungen wie dem Kommissionieren, mit dem Ziel, dass der Bediener seine Aufmerksamkeit nicht mehr auf FiFi lenken muss, ist Gegenstand künftiger Arbeiten. Damit bei Hindernissen ein durch den Laserscanner ausgelöster sicherer Stopp vermieden werden kann, soll eine Hinderniserkennung mittels Kinect das Schneiden der Kurven verhindern. FiFi soll – statt Kurven zu schneiden - bei Erkennung eines Hindernisses den Pfad des Nutzers nachfahren.

Durch die Umsetzung der Konzepte zur Gestensteuerung und Personenverfolgung konnte gezeigt werden, dass die berührungslose Steuerung durch Gesten und Personenverfolgung – neben dem Einsatz im Konsumerbereich – großes Potential für Anwendungen in der Intralogistik bietet.

## LITERATUR

- [Fur10] Furmans, Kai; Schönung, Frank; Gue, Kevin R.: *Plug-and-work material handling systems*. In Proceedings of the International Material Handling Research Colloquium. 2010.
- [Kob06] Kobilarov, Marin; Sukhatme, Gaurav; Hyams, Jeff; Batavia, Parag: *People tracking and following with mobile robot using an omnidirectional camera and a laser*. IEEE International Conference on Robotics and Automation. 2006.
- [Doi12] Doisy, Guillaume; Jevtic, Aleksandar, Lucet, Eric; Edan, Yael: *Adaptive Person-Following Algorithm Based on Depth Images and Mapping*. Workshop on Robot Motion Planning: Online, Reactive, and in Real-time. 2012.
- [Shi13] Shimada, Atsushi; Yamashita, Takayoshi; Taniguchi, Rin-ichiro: *Hand Gesture based TV Control System*. 19th Korea -Japan Joint Workshop on Frontiers of Computer Vision. 2013.
- [Lam11] Lambrecht, Jens; Kleinsorge, Martin; Krüger, Jörg: *Markerless Gesture-Based Motion Control and Programming of Industrial Robots*. IEEE ETFA. 2011.
- [Abi13] Abid, Muhammad R.; Santiago Melo, Lidia B.; Petriu, Emil M.: *Dynamic Sign Language and Voice Recognition for Smart Home Interactive Application*. IEEE International Symposium on , vol., no., pp.139,144, 4-5. 2013.
- [Gal12] Galatas, Georgios; Potamianos, Gerasimos; Makedon, Fillia: *Audio-Visual Speech Recognition incorporating Facial Depth Information captured by Kinect*. 20th European Signal Processing Conference. 2012.
- [Nie92] Nielsen, J.: *The Usability Engineering Life Cycle*. Computer 25 (3), p.12–22. 1992.
- [Nie04] Nielsen, M.; Störring, M.; Moeslund, T.; Granum, E.: *A Procedure for Developing Intuitive and Ergonomic Gesture Interfaces for HCI*. In: A. Camurri und G. Volpe (Hrsg.), *Gesture-Based Communication in Human-Computer Interaction*, Volume 2915 of Lecture Notes in Computer Science, p. 105–106. Springer Berlin /Heidelberg. 2004.
- [Kho10] Khoshelham, K.: *Accuracy Analysis of Kinect Depth Data*. GeoInformation Science 38 (5/W12), S. 6. 2010.
- [Bes88] Besl, P. J.: *Active, Optical Range Imaging Sensors*. Machine Vision and Applications 1, S. 127–152. 1988