18. AALE-Konferenz
Hochschule Pforzheim, 09.03.–11.03.2022

Prozessinnovation: Selbstlernendes Assistenzsystem für die manuelle Montage

J. Jung[1], A. Hubert[2], K. Doll[3], M. Kröhn[4], J. Stadler[5]

Zusammenfassung

Manuelle Montageprozesse sind nach wie vor unverzichtbar in vielen Bereichen der produzierenden Industrie. Vor allem die Qualitätskontrolle, sowie das Einlernen neuer Mitarbeitenden stellen Betriebe durch die voranschreitende Digitalisierung vor neue Herausforderungen. Assistenzsysteme können hier helfen, die Lücke zwischen Anforderungen und Qualifikation zu überbrücken. Wir stellen einen Ansatz zur intelligenten Assistenz vor, welcher auf einer kamerabasierten Erkennung von Arbeitsabläufen mit Hilfe von Methoden des maschinellen Lernens beruht. Das Assistenzsystem erzeugt automatisiert Hilfsmaterial zur Unterstützung der Werkenden. Zusätzlich zur Darstellung der technischen Aspekte, werden psychologische Aspekte, wie Akzeptanz und Motivation untersucht.

Stichwörter

Machine Learning, Montage, Assistenzsystem, Mensch-Maschine-Schnittstelle

1 Einleitung

1.1 Problemstellung

Kleine und mittelständische Unternehmen aus dem produzierenden Gewerbe stehen heute vor der Aufgabe, sich durch Digitalisierung für die Zukunft aufzustellen, aber gleichzeitig und oftmals in Nischenbereichen wirtschaftlich zu bleiben. In diesem Zusammenhang stellen manuelle Montageprozesse eine besondere Herausforderung dar. Durch meist kleine Losgrößen oder komplexe Bauteile ist ein hoher Automatisierungsgrad nicht wirtschaftlich umsetzbar, sodass Werkende unentbehrlich bleiben. Allerdings werden die Montageprozesse immer komplexer, während Mitarbeitende in diesen Bereichen häufig gering qualifiziert und teilweise keine Muttersprachler sind. Daher sind Assistenzsysteme, die die Lücke zwischen Anforderungen und Qualifikation überbrücken können, ein zunehmend bedeutenderer Faktor, um Unternehmen erfolgreich zu halten. Zudem lässt sich dadurch auch ein Bruch in der Digitalisierungskette vermeiden, indem wertvolle Daten zum Fertigungsprozess erfasst werden können. Daraus ergibt sich ein Bedarf an intelligenten Montagearbeitsplätzen, die die Wirtschaftlichkeit der Prozesse des Unternehmens und die Qualifikation der Mitarbeitenden berücksichtigt. Vor diesem Hintergrund bearbeitet die Technische Hochschule Aschaffenburg gemeinsam mit den Industriepartnern Fischer Information, ROBUR Automation und weiteren Partnern das Forschungsprojekt GesInA40 (Gestenbasiertes Intelligentes Assistenzsystem) [1]. Die Partner beforschen seit 2012 das Thema Unterstützungssysteme, mit deren Hilfe Beschäftigten komplexe Montageprozesse vermittelt werden, indem sie kontextbezogene Hilfe durch das System erhalten. Aus diesen Arbeiten ist bereits ein

1 Janis Jung, TH Aschaffenburg – University of Applied Sciences, Aschaffenburg
2 Andreas Hubert, TH Aschaffenburg – University of Applied Sciences, Aschaffenburg
3 Konrad Doll, TH Aschaffenburg – University of Applied Sciences, Aschaffenburg
4 Michael Kröhn, ROBUR Automation GmbH, Niedernberg
5 Jochen Stadler, Fischer Information Technology AG, Radolfzell

Open Access. © 2022 J. Jung, A. Hubert, K. Doll, M. Kröhn, J. Stadler
This work is licensed under the Creative Commons Attribution 4.0 License. +

Musterarbeitsplatz hervorgegangen, an dem ein intelligentes, automatisiert lernendes Assistenzsystem einen einfachen Ablauf unterstützt. Technisch blieben dabei noch einige Themen unberührt, die im aktuellen Vorhaben bearbeitet werden.

1.2 Ziele

Im Vorhaben wird die Verbesserung des bestehenden Assistenzsystems hinsichtlich technischer Robustheit und Usability angestrebt. Hierzu wurde der vorhandene Musterarbeitsplatz (Bild 1) adaptiert, um damit unterschiedlichste Montageprozesse nachzustellen. Mit dem bestehenden Assistenzsystem können gelernte Abläufe an einem manuellen Montagearbeitsplatz beobachtet, validiert und unterstützt werden. Konkret soll die Erweiterung des Systems erlauben, neue Abläufe ohne Prozesswissen durch das Beobachten von erfahrenen Mitarbeitern einzulernen und unerfahrenen diese aufgezeichneten Beobachtungen abhängig vom Prozessschritt zur Verfügung zu stellen. Visuelle Hilfestellung werden hierzu während des Einlernens automatisiert aufgezeichnet und können bei Bedarf prozessgerecht und im Kontext des aktuellen Arbeitsschritts dargestellt werden. Eine besondere technische Herausforderung stellen Aktivitätserkennung und Handerkennung dar, weil diese den von System erfassten Montageprozess darstellen. Aktivitäten innerhalb eines Montageprozesses werden als Gesten und Bewegungen erfasst. Für eine erweiterte Abdeckung von Anwendungsfällen ist es wichtig, dass dies ohne zusätzliche Hilfsmittel, wie bspw. eingefärbte Handschuhe oder Marker, möglich ist. Andererseits sollen Handschuhe auch kein Problem für die Erkennung darstellen, da viele Assemblierungsprozesse aus Reinheitsgründen damit durchgeführt werden. Ziel ist, dass Beschäftigte ihre Arbeitsroutine weiterhin beibehalten können. Dies wird auch dadurch unterstützt, dass Informationen im Sichtfeld des Arbeitenden dargestellt werden. Hier kommt ein Beamer zum Einsatz, der automatisiert generierte Videosequenzen auf die Arbeitsfläche projiziert.

Bild 1: Musterarbeitsplatz an der Technischen Hochschule Aschaffenburg, an dem sich unterschiedliche Montageprozesse oberhalb der gesamten Tischplatte simulieren lassen. Kamera und Beamer sind überkopf montiert, Eye-Tracker auf Höhe der blauen Boxen.

Open Access. © 2022 J. Jung, A. Hubert, K. Doll, M. Kröhn, J. Stadler
This work is licensed under the Creative Commons Attribution 4.0 License. +

In diesem Zusammenhang ist zu klären, welche Form der Darstellung an welcher Position auf der Arbeitsfläche am effektivsten ist. Außerdem ist wichtig, wie lange eine Information dargestellt werden soll, um die Gefahr von Reizüberflutung und Sättigung zu vermeiden. Hier kommt unter anderem ein Eyetracker zum Einsatz, der von den psychologischen Fakultäten der Universitäten Würzburg und Heidelberg verwendet wird, um diese Sachverhalte zu untersuchen. Als weitere Konsortialpartner beschäftigen diese sich mit Usability, Arbeitsmotivation und Arbeitsplatzgestaltung. Unter anderem werden Aufmerksamkeitsspannen untersucht, um die Nützlichkeit verschiedener Unterstützungselemente zu validieren. Begleitet werden diese Untersuchungen durch Kampagnen mit Mitarbeitenden der Firmen und Studierenden. Im Folgenden werden zunächst die Komponenten des Assistenzsystems, wie die verbindende Software-Architektur und die Kommunikationsflüsse, beschrieben. Dann werden die technischen Herausforderungen und Lösungen bei der Umsetzung von Hand- und Aktivitätserkennung diskutiert. Im Anschluss daran werden technische Ergebnisse sowie die Resultate aus einer Usability-Kampagne vorgestellt. Aus den Ergebnissen kann weiterer Handlungsbedarf abgeleitet werden, der in Zusammenfassung und Ausblick mit nächsten Schritten unterlegt wird.

2 Assistenzsystem

Der innovative Kern der Arbeit entwickelt sich um ein Assistenzsystem zur Unterstützung manueller Montageprozesse. Dieses System, seine Komponenten und seine Abläufe sollen zunächst betrachtet werden, um darauf aufbauend die Entwicklungsarbeiten zu diskutieren. Das Assistenzsystem besteht aus einer 3D Kamera (Intel RealSense D435) und einem Projektor, die beide überkopf am Arbeitsplatz montiert werden, sowie einem Eye-Tracker und einem Industrie-PC. Visualisierung und Konfiguration erfolgen über eine Smartphone-App. Die Systemarchitektur besteht softwareseitig aus einem Steuersystem und einem Informationssystem mit semantisch strukturierter Wissensdatenbank. Das Setup ist in Bild 2 zu sehen.

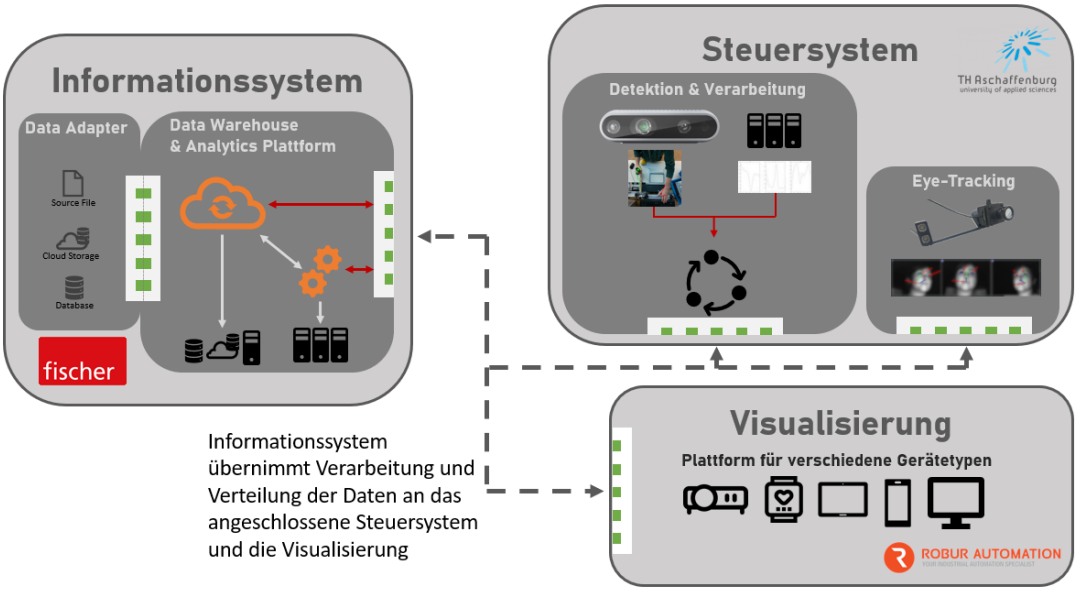

Bild 2: Schematisches Setup der Systemarchitektur und Verknüpfung der Einzelsysteme, bestehend aus Informationssystem, Steuersystem, Visualisierung.

Die 3D-Kamera ermöglicht eine Erfassung von Bild- und Tiefendaten. Zusätzlich kann ein Infrarot Bild abgerufen werden. Damit lassen sich die Bewegungen auf der Arbeitsfläche auf verschiedene Weisen auswerten. Hierzu wurden Algorithmen erprobt und verglichen, die markerbasiert (bspw. eingefärbte Handschuhe) oder markerlos eine Erkennung der Hände in unterschiedlichen Detailstufen ermöglichen und eine Umwandlung von Videostreams in Zeitreihen von Positionsdaten ermöglichen. Diese

Open Access. © 2022 J. Jung, A. Hubert, K. Doll, M. Kröhn, J. Stadler
This work is licensed under the Creative Commons Attribution 4.0 License. +

Zeitreihen werden weiter analysiert, um daraus Aktivitäten abzuleiten. Auf Basis dieser Aktivitäts- und Gestenerkennung, wird das System so erweitert, dass ein Arbeitsprozess durch den Einsatz von Machine Learning Verfahren selbständig belernt werden kann. Das System lernt, indem es einen erfahrenen Werkenden bei der Durchführung von unbekannten Aktivitäten beobachtet. In der Lernphase wird der Arbeitsprozess solange wiederholt, bis das System ein zuvor festgelegtes Konfidenzlevel übersteigt. In einem ersten Test hat sich gezeigt, dass ein Arbeitsprozess mit 15 Einzelschritten bereits nach 5 Wiederholungen soweit eingelernt werden kann, dass mit hoher Sicherheit alle Aktivitäten des Arbeitsprozesses korrekt erkannt werden. Während der Lernphase baut das System den Arbeitsprozess als gerichteten, azyklischen Graphen auf. Dieser Graph stellt durch Knoten (Prozesszustände) und Kanten (Aktivitäten) verschiedene mögliche Arbeitsabläufe dar. Hilfematerial wird automatisiert in Form von Videos der Handgriffe und entsprechenden Arbeitsanweisungen produziert, welche dann in situ als Augmented Reality (AR) über den Projektor bereitgestellt werden. Der passende automatisierte Zuschnitt auf die Teilschritte ist dabei für die Akzeptanz durch den Anwender ebenso wichtig, wie die Klarheit und Einfachheit der Darstellung. Wissenschaftliche Auswertungen bezüglich Mensch-Maschine-Schnittstellen werden in Kooperation mit dem Lehrstuhl für Kognitive Psychologie der Julius-Maximilians-Universität Würzburg und dem Psychologischen Institut der Universität Heidelberg durchgeführt. Der Eye-Tracker, der in erster Linie zur Bestimmung von Aufmerksamkeitsspannen verwendet wird, soll in diesem Zuge auch als Eingabemöglichkeit untersucht werden. Erste Tests mit dem Auslösen von Events durch Blickrichtungssteuerung, wie dem Einblenden von Hilfevideos, gestalteten sich sehr positiv. Komplettiert wird das Assistenzsystem von einer Wissensdatenbank, die die redaktionelle Pflege der automatisiert generierten Inhalte ermöglicht. Mitarbeitende der technischen Redaktion können die aufgezeichneten Abläufe Teilschritt für Teilschritt abrufen und um weitere Inhalte, wie etwa beschreibende Texte, ergänzen.

2.1 Automatisierung

Ein wesentlicher Bestandteil des Assistenzsystems ist das automatische Einlernen und Beobachten von unbekannten Prozessen. Das System durchläuft für einen unbekannten Prozess drei Betriebszustände. Im ersten Schritt wird der grundlegende Arbeitsprozess durch Beobachtung gelernt. Im zweiten Schritt wird durch wiederholte Ausführung des Arbeitsprozesses eine mögliche Variabilität trainiert, um eine höhere Robustheit gegenüber möglichen Abweichungen zu schaffen. Der dritte Schritt ist der Normalbetrieb, der Prozess wird beobachtet und kann mit der eingelernten Vorlage abgeglichen werden.

Einlernen. Zu Beginn durchläuft das Assistenzsystem für neue Umgebungen einmalig einen Kalibrierprozess, um sich an unterschiedlichen Umgebungen (z.B. Belichtung) anzupassen. Das System übergibt nach der Initialisierung an den Werker, der den konkreten Arbeitsablauf durchführt. In dieser Phase lernt das System anhand der Analyse von Livebildern die korrekte Abfolge von Montageschritten. Durch Machine Learning Methoden wird aus den Beobachtungen des Systems ein Zustandsautomat erstellt. Es findet dabei kein aktiver Eingriff durch das System statt, auch muss der Werkende keine zusätzlichen Informationen beitragen.

Weiterlernen. Der initial erstellte Ablauf ist auf Basis einer oder weniger Beobachtungen des Montageprozesses gelernt worden. Beim wiederholten Durchführen des Montageprozesses kommt es für einzelne Prozessschritte zu unterschiedlich großen Abweichungen, die für eine robuste Erkennung der Aktivitäten abgeschätzt und berücksichtigt werden müssen. Der Montageprozess wird so lange durchgeführt, bis das System ein bestimmtes Konfidenzlevel erreicht hat. Darüber hinaus können während des Weiterlernens, der Zustandsautomat dynamisch um weitere Prozessschritte und veränderte Abläufe ergänzt werden.

Normalbetrieb. Einlernen und Weiterlernen müssen einmalig erfolgt sein, damit das Assistenzsystem im Normalbetrieb arbeiten kann. In diesem Betriebsmodus beobachtet das System den Werker und gleicht den erfassten Prozess mit dem im Zustandsautomaten hinterlegten Prozess ab. Dem Wer-

Open Access. © 2022 J. Jung, A. Hubert, K. Doll, M. Kröhn, J. Stadler
This work is licensed under the Creative Commons Attribution 4.0 License. +

ker ist es möglich auf Hilfematerial für den aktuellen Prozessschritt zurückzugreifen und sich dieses visuell darstellen zu lassen. Abweichungen werden vom System in einer Wissensdatenbank abgelegt. Das System lässt sich durch weitere Metainformationen anreichern, um Prozessschritte mit Labeln und Beschreibungen zu versehen.

2.2 Handdetektion

Die in dieser Arbeit verwendeten Algorithmen zur Erfassung und zum selbstständigen Lernen von Aktivitäten, bauen auf einem Modul zur Handerkennung auf. Die Erfassungsgenauigkeit dieser Handerkennung beeinflusst die Genauigkeit des gesamten, darauf aufbauenden Systems. Aus diesem Grund soll an dieser Stelle genauer auf verschiedene Techniken und deren Vor- und Nachteile eingegangen werden, die zur Erkennung der Hände herangezogen worden sind.

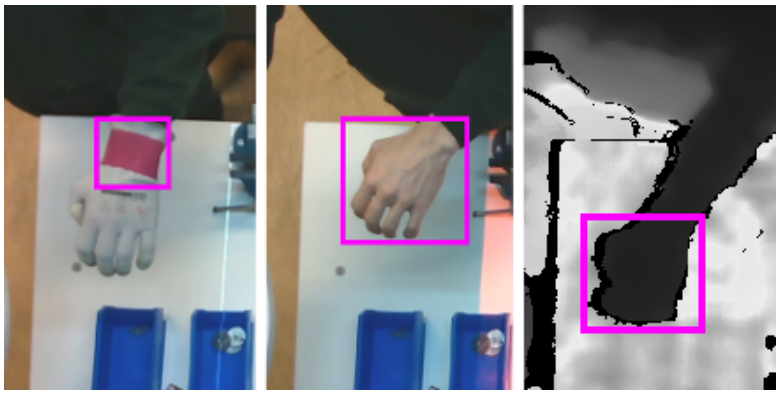

Bild 3: Darstellung der Handerkennung auf Basis verschiedener Verfahren und verschiedener Eingangsdaten. Links: Farbbasierte Erkennung. Die Erkennung findet anhand der rosa Farbe des Handschuhs statt. Mitte: Erkennung im Farbbild. Deep Learning Verfahren zur Erkennung von Händen auf Farbbildern. Rechts: Erkennung im Tiefenbild. Deep Learning Verfahren zur Erkennung von Händen auf Tiefenbildern.

Farbbasierte Erkennung. Durch das Tragen von speziellen, farbig markierten Handschuhen können diese robust und sehr effizient im Bild durch Blob Detection erkannt werden. Die Berechnungsrate ist stark abhängig von der verwendeten Hardware und Optimierung, aber im Vergleich zu den weiteren vorgestellten Methoden deutlich schneller. Wie in Bild 3, links, dargestellt, ist die Markierung am Handgelenk angebracht, da dieses in dem normalen Bewegungsraum der Hände am Arbeitsplatz, im Gegensatz von z.B. den Fingern, fast immer sichtbar ist. Eine weitere gängige Methode zur robusten Erkennung ist das Einbringen von speziellen Mustern oder Makern auf die Handschuhe [2]. Je nach Anforderung kann aus diesen Mustern auch die Pose der Hand bestimmt werden.

Vorteile:

- Effizient zu berechnen, benötigt keine dedizierte Hardware, wie z.B. Grafikkarten
- Robuste Erkennung, auch für schnelle Bewegungen oder schlechten Lichtverhältnissen
- Übertragbar, da sich das Erscheinungsbild der Hand nicht ändert

Nachteile:

- Durch die Erkennung des Handgelenks kommt es zu einer Abweichung von der eigentlichen Handposition. Kleine Bewegungen der Hand sind dadurch nicht erkennbar.
- Farbe ist vorgegeben, darf nicht in Farbspektrum des Arbeitsplatzes vorkommen
- Tragen von Handschuhen ist erforderlich, dies ist nicht immer möglich. Vorgaben zum Tragen von speziellen Handschuhen kann auch auftragsseitig vorgegeben sein.

Open Access. © 2022 J. Jung, A. Hubert, K. Doll, M. Kröhn, J. Stadler
This work is licensed under the Creative Commons Attribution 4.0 License. +

Erkennung im Farbbild. Tiefe neuronale Netze bilden den de facto Standard in der Objekterkennung [3]. Bei der Handerkennung handelt es sich um einen prominenten Spezialfall, für den es eine große Anzahl an wissenschaftlichen Publikationen und öffentliche Datensätze gibt. Durch die vielen Freiheitsgrade der Erscheinungsform von Händen, wie z.B. Blickwinkel, Textur, Entfernung zur Kamera, Verdeckung, wird eine sehr große Anzahl an Trainingsdaten benötigt, um einen hinreichend guten, allgemeingültigen Klassifikator zu erstellen [4]. Die daraus resultierenden, großen Modelle mit hoher Genauigkeit sind im Vergleich langsam (<30 fps).

Vorteile:

- Funktioniert auch bei Handinteraktion verlässlich
- Funktioniert auch für Handschuhe
- Muss nicht trainiert werden, verfügbare Modelle liefern bereits Out-of-the-box sehr gute Erkennungsqualität, auch für komplexe Szenarien, wie getragene Handschuhe

Nachteile:

- Große Modelle, grundsätzlich langsam, kleinere und schnellere Modelle brauchen spezifisches Training oder liefern schlechte Performance
- Benötigt dedizierte Hardware, wie z.B. Grafikkarten

Erkennung im Tiefenbild. Handerkennung findet direkt auf der Disparitätenkarte einer Stereokamera statt. Ein eminentes Problem bei der Verwendung von Handerkennern, die ohne Hilfsmittel wie Farbe oder Marker auskommen ist die große Variabilität der Hände. Variation in Farbe und Muster senken die Genauigkeit von Handerkennern immens. Ein konkretes Problem ist das Tragen von Handschuhen, die zur Trainingszeit unbekannt sind. Tiefeninformationen beinhalten die Form und sind unabhängig von Farbe und Muster.

Vorteile:

- Übertragbar, da sich die Handform grundlegend nicht verändert
- Invariant gegenüber Farbe und Muster

Nachteile:

- Geringe Auflösung und Genauigkeit, z.B. Interaktion zwischen Händen oder mit Werkzeug
- Vergleichsweise hohe fehlerhafte Erkennungen

Erkennung der Gelenkpunkte. Für verschiedenen Anwendungsszenarien kann es erforderlich sein, über die reine Lokalisierung der Hand im Bild hinaus, Informationen zu erfassen. Sollen Gesten und Handpositionen ausgewertet werden, kann dies durch die Erfassung einzelner Gelenkpunkte, aus denen sich ein Handskelett erstellen lässt, erfolgen [5, 6].

Vorteile:

- Detaillierte Analyse zum Zustand der Hände, z.B. Hand offen, Objekt gegriffen
- Ergebnis kann plausibilisiert werden

Nachteile:

- Grundsätzlich langsamer als die reine Handerkennung, da die Schätzung der Gelenkpunkte auf eine Handerkennung folgt.
- Hoher Annotationsaufwand beim Erstellen von Datensätzen

Für das Vorhaben kommt die farbbasierte Erkennung und die Erkennung auf den Tiefenwerten zum Einsatz. Die Farberkennung liefert mit Abstand die robustetste Erkennung und ist daher wann immer möglich bevorzugt. Die Erkennung auf Basis der Tiefenwerte kommst in allen anderen Fällen zum

Open Access. © 2022 J. Jung, A. Hubert, K. Doll, M. Kröhn, J. Stadler
This work is licensed under the Creative Commons Attribution 4.0 License. +

Einsatz, da diese unabhängig vom Erscheinungsbild (Farbe und Muster) der Hand und somit für ein breites Anwendungsgebiet einsetzbar ist.

2.3 Aktivitätserkennung

Aktivitäten sind im Kontext von manuellen Montageprozessen dadurch definiert, dass diese Bestandteile einer zumeist vordefinierten Abfolge von eindeutig zuordenbaren Tätigkeiten darstellen. Dabei können Aktivitäten einfach oder auch mehrfach in einer solchen Abfolge auftreten. Um einen möglichst hohen Grad an Assistenz zu bieten, müssen die Aktivitäten innerhalb einer Abfolge robust und frühzeitig erkannt werden, damit bei möglichen fehlerhaften Ausführungen schnellstmöglich eingeschritten werden kann. Die Detektion der Hände und der resultierende Positionsverlauf dienen als Basis für die Erkennung der Aktivitäten während der Montage.

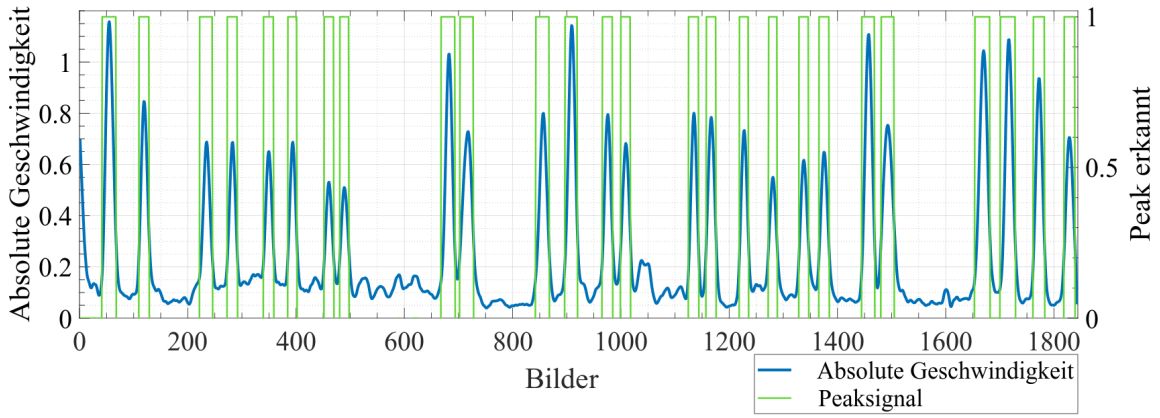

Bild 4: Ausgangssignal des Peakdetektions-Algorithmus (grün) auf die geglättete absolute Geschwindigkeit (blau) für einen beispielhaften Arbeitsablauf.

Aktuell wird die Aktivitätserkennung mit Hilfe eines k-Nearest-Neighbor (kNN) Klassifikators realisiert [7]. Dieser Ansatz wurde bereits in vorherigen Arbeiten experimentell evaluiert [8, 9]. Die Grundlage ist die einfache Annahme, dass sich eine oder beide ausführenden Hände zur Aktivität bewegen, bei der Durchführung nahezu ruhig an einer Stelle bleiben, womöglich auch nur für einen sehr kurzen Moment, und sich danach zur nächsten Aktivität bewegen. Durch diese Annahme lassen sich alle Aktivitäten eines Prozesses bestimmen, in dem die Geschwindigkeitsverläufe der Hände segmentiert werden, siehe auch Bild 4. Dadurch entstehen ortsgebundene Aktivitäten, die sich durch mehrfache Wiederholungen zu einer Datenbasis, bestehend aus Aktivitätsclustern, entwickeln, beispielhaft dargestellt in Bild 5. Diese Vorgehensweise erlaubt das Einlernen des Assistenzsystems durch mehrfache Wiederholung eines Arbeitsablaufs von geschultem Personal. Diese Art der Aktivitätserkennung ist rein positionsbasiert, d.h. alle unterschiedlichen Aktivitäten eines Prozesses müssen räumlich getrennt sein, ansonsten ist eine Unterscheidung nicht möglich.
Um diese Einschränkung zu überwinden, wird eine trajektorienbasierte Aktivitätserkennung implementiert. Diese bezieht zusätzlich zur aktuellen Position auch die vergangenen Positionen der Hände mit ein. Dazu werden nicht die Positionen geclustert, sondern die Trajektorien zwischen den Aktivitäten verwendet. Eine Aktivität ist der Bereich zwischen zwei Peaks, wie in Bild 4 in grün hervorgeboben. Die einzelnen Trajektorien werden mittels Dynamic Time Warping (DTW) [10] verglichen und geclustert. Für die Aktivitätserkennung werden diese Cluster mit DTW Barycenter Averaging (DBA) [11] gemittelt, um repräsentative Trajektorien zu erhalten. Im Livebetrieb des Assistenzsystems werden die repräsentativen Aktivitätstrajektorien mit den erfassten Handtrajektorien über DTW verglichen.

Open Access. © 2022 J. Jung, A. Hubert, K. Doll, M. Kröhn, J. Stadler
This work is licensed under the Creative Commons Attribution 4.0 License. +

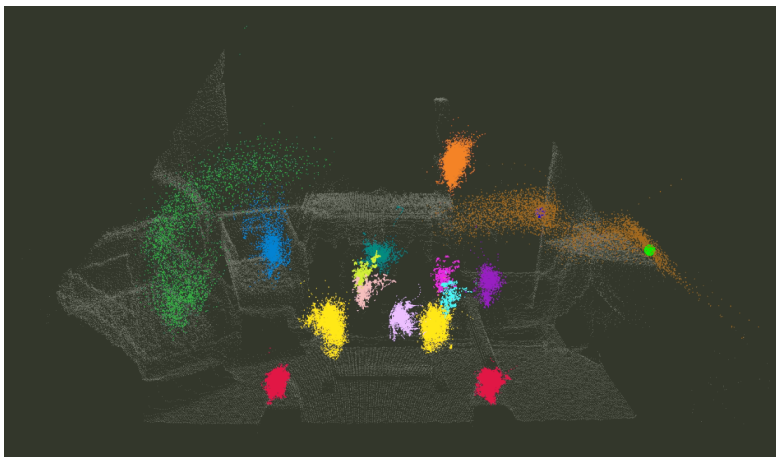

Bild 5: Beispielhafte Darstellung der farbigen Cluster, welche die einzelnen Aktivitäten repräsentieren. Der Arbeitsplatz ist durch die weiße Punktewolke angedeutet.

3 Evaluation

Im laufenden Vorhaben werden insbesondere die Hand- und Aktivitätserkennung gegenüber dem ursprünglichen Assistenzsystem verbessert. Der Vorgänger arbeitete mit einer farbbasierten Handschuherkennung, da sie, wie zuvor beschrieben, robust und kostengünstig bzgl. Hardware und Energieeffizienz ist. Allerdings hängt sie vom Einsatz farbiger Handschuhe ab. Dies ist ggf. nicht möglich. Unter Umständen lässt sich auch kein ausreichend isolierter Farbort für die Erkennung finden. Daher wird das System an dieser Stelle verallgemeinert. Die Vorteile einer farbbasierten Erkennung bestehen in der Robustheit nicht zuletzt auch hinsichtlich der erkannten Aktivitäten. Um diese auch mit anderen Ansätzen zu erreichen, wird die zuvor beschriebene neue Methode der trajektorienbasierten Aktivitätserkennung verwendet. Nachfolgend werden erste Ergebnisse und Limitationen besprochen.

3.1 Aktivitätserkennung

Der erste Schritt zur Evaluierung der trajektorienbasierten Aktivitätserkennung, ist die Beurteilung der Ähnlichkeit zweier Instanzen anhand eines Similarity-Scores, der durch ein Matching mittels DTW ermittelt werden soll. Um diesen methodischen Ansatz auf dessen Eignung zu untersuchen, wurden zwei komplette Arbeitsdurchläufe ID 0 und ID 1, von unterschiedlichen Personen durchgeführt und miteinander verglichen. Beide beinhalten jeweils 26 Aktivitätssegmente, von denen jede mögliche Kombination mit DTW verglichen und ein Score berechnet wurde. Bild 6 zeigt das Ergebnis der Untersuchung. Der Marker (roter Stern) in den Feldern zeigt den geringsten Score für die jeweilige Spalte – also die Kombination mit der höchsten Ähnlichkeit. Eine Diagonale durch die Matrix deutet auf ein perfektes Matching der Segmente hin. Für die Kombinationen (5, 5) und (6, 6) wurde dies nicht erreicht, jedoch liegt bei Beachtung der in Klammern geschriebenen ClassIDs keine fehlerhafte Erkennung vor. Dieser Umstand ist auf die Redundanzen in den Klassen zurückzuführen. Ein weiteres Phänomen – bei Betrachtung der gesamten Heatmap – sind schachbrettartige Interferenzmuster im linken oberen und rechten unteren Bereich, aber auch abgeschwächt in anderen Teilen des Diagramms. Diese kommen durch die zugrundeliegende Abfolge an Tätigkeiten und den aufeinanderfolgenden Interaktionsgruppen von Take-Use-Interaktionen auf, die sehr ähnliche Trajektoriensegmente aufweisen und teilweise derselben Klasse zugehörig sind. Dabei sticht das Feld (9, 9) beziehungsweise dessen Spalte dadurch heraus, dass es eine allgemein geringere Ähnlichkeit zeigt verglichen mit anderen gematchten Feldern. Dies ist ein Indiz dafür, dass in dieser aufgenommenen Sequenz ID 1 parasitäre Variationen in Segment ID 9 auftreten, die zu einer erschwerten oder sogar fehlerhaften Erkennung führen können. Aus diesem Grund ist es essenziell, eine repräsentative Datenbasis aufzubauen, um zum einen tatsächliche Variationen in der Aktivitätsausführung zu berücksichtigen und zum anderen solche parasitären Effekte durch ein geeignetes Averaging zu reduzieren. Detailliertere Evaluationen zur trajektorienba-

Open Access. © 2022 J. Jung, A. Hubert, K. Doll, M. Kröhn, J. Stadler
This work is licensed under the Creative Commons Attribution 4.0 License. +

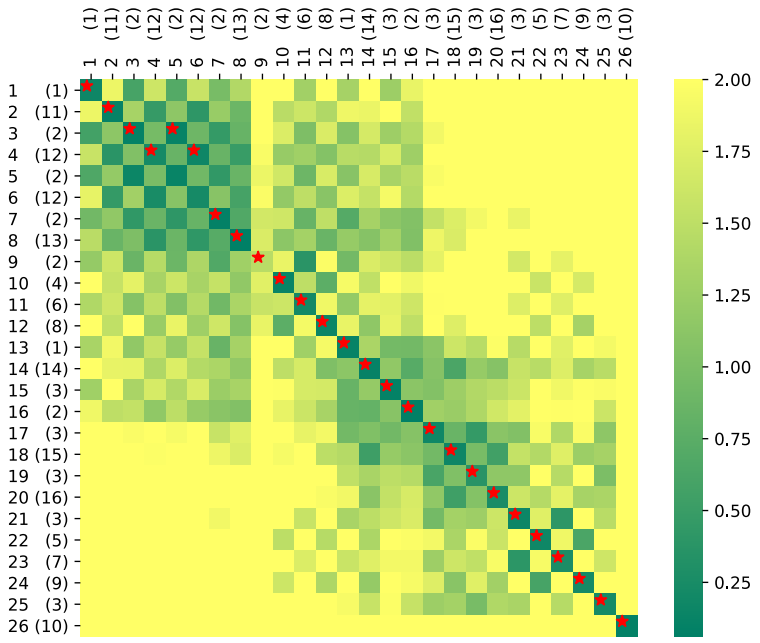

Bild 6: Similarity-Scores der DTW Berechnungen aller Kombinationen der extrahierten Segmente aus den Sequenzen ID0 (links aufgetragen) und ID1 (oben aufgetragen) – betitelt mit der Segment ID und in Klammern die dazugehörige ClassID. Je höher die Ähnlichkeit, desto dunkler ist das jeweilige Feld. Der geringste Score jeder Spalte ist mit einem Stern versehen.

sierten Aktivitätserkennung und ihrer Robustheit beim Einsatz im Assistenzsystem stehen noch aus.

3.2 Limitation

Die aktuell implementierten Verfahren, für sowohl trajektorien- wie auch positionsbasierten Aktivitätserkennung lassen nur eine gewisse Auflösung von Aktivitäten zu. Reale Prozessschritte, die nicht eindeutig aufgelöst werden können, werden vom System zu einer Aktivität zusammengefasst. Wie groß die Auflösung des Systems ist, liegt an mehreren Faktoren. Der wesentliche Faktor ist dabei die gelernte Abweichung für ein und dieselbe Aktivität. Beispielsweise ist der Mittelpunkt der Aktivität nicht der Mittelpunkt der erfassten Hand, bei veränderter Griffposition verändert sich entsprechend der Ort der Aktivität und die Unsicherheit nimmt zu.

3.3 Nützlichkeit und Motivation

Entscheidend für die erfolgreiche Einführung neuer Technologien ist die Akzeptanz der Anwendenden. Wird das System als nützlich empfunden? Motiviert es im Arbeitsprozess? Aus diesem Grund wird begleitend zur technischen Entwicklung eine psychologische Evaluation durchgeführt. In der ersten Phase einer Kampagne wurde das Assistenzsystem dem klassischen Lernen des Arbeitsablaufs mit Papierdokumentation gegenübergestellt. Die Aufgabe bestand darin, den vorgegebenen Montageprozess drei Mal korrekt auszuführen, entweder auf Basis der Videos oder der Dokumentation. Ein Montagefehler führt zum Abbruch und erneuten Start von vorne. In einer zweiten Phase wurde dann das Erlernte ohne Hilfsmittel angewandt. Ziel war es, den Montageprozess fünf Mal korrekt auszuführen. Die Studie, die mit 29 Teilnehmenden durchgeführt wurde, ergab, dass mit der aktuellen Version des Assistenzsystems Durchführungszeiten und Fehlerhäufigkeiten im Vergleich zum klassischen Ansatz in ähnlichen Bereichen lagen. Auch in der User Experience, die mit begleitenden Fragebögen erhoben wurde, ist das Assistenzsystem der Papierdokumentation aktuell noch nicht überlegen. Allerdings

Open Access. © 2022 J. Jung, A. Hubert, K. Doll, M. Kröhn, J. Stadler
This work is licensed under the Creative Commons Attribution 4.0 License. +

ließen sich daraus konkrete Handlungsempfehlungen ableiten, die in der nächsten Iteration des Entwicklungsprozesses umgesetzt werden, beispielsweise die präzisere Vermittlung von Feedback und die Verwendung zusätzlicher, intuitiver Darstellungselemente.

4 Zusammenfassung

In diesem Beitrag wurden die zentralen technischen Bausteine eines Assistenssytems für manuelle Montageprozesse vorgestellt. Zum einen wurde auf Methoden zur Handerkennung, sowie deren Vor- und Nachteile eingegangen. Zum anderen wurde die darauf basierende Aktivitätserkennung geschildert und in einem ersten Ansatz evaluiert. Erste Auswertungen haben Limitationen des Gesamtsystems sowohl auf technischer, aber auch aus psychologischer Sicht ergeben. Diese fließen in die weitere Entwicklung des Systems ein. Die zukünftigen Entwicklungen zielen vor allem auf die Robustheit der Aktivitätserkennung, sowie die Darstellung und Aufbereitung der Hilfsmaterialien zur Steigerung der Akzeptanz der Werkenden bei Anwendung im Echtzeitbetrieb ab.
Für den Einsatz einer Lösung zur Unterstützung von manueller Montage ist die Betrachtung der Wirtschaftlichkeit unabdingbar. Die vorgestellte Lösung macht einen Vorschlag zur Entwicklung flexibler Systeme, die robuste KI-Modelle verwenden, welche den hohen Initialaufwand und laufenden Support durch selbstständiges Lernen deutlich reduzieren. Konkret bedeutet das für Unternehmen, dass sich durch eigenes Wissen selbstständig und schnell digital abgebildete Montageprozesse ändern oder bestehende erweitern lassen, ohne dass Fachkenntnisse zu den KI-Modellen notwendig sind.

5 Danksagung

Wir bedanken uns beim Bundesministerium für Bildung und Forschung (BMBF) für die Förderung des Forschungsprojektes GesInA40 (01IS19043C) im Rahmen der Fördermaßnahme "KMU-innovativ" sowie beim DLR Projektträger Datenwissenschaften / Softwareintensive Systeme.

6 Literatur

[1] GesInA40: Projekt im Rahmen der Förderinitiative "KMU-innovativ", die vom Bundesministerium für Bildung und Forschung (BMBF) unterstützt wird. Zuteilungsnummer: 01IS19043C
[2] Han, S., Liu, B., Wang, R., Ye, Y., Twigg, C., Kin, K.: Online optical marker-based hand tracking with deep labels. ACM Transactions on Graphics **37**, 1–10 (2018). DOI 10.1145/3197517.3201399
[3] Kebinger, S., Kröhn, M.: A brief history of deep neural networks for image processing tasks (2021). DOI 10.13140/RG.2.2.29804.80005
[4] Shan, D., Geng, J., Shu, M., Fouhey, D.: Understanding human hands in contact at internet scale (2020)
[5] Simon, T., Joo, H., Matthews, I., Sheikh, Y.: Hand keypoint detection in single images using multiview bootstrapping. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2017)
[6] Zhang, F., Bazarevsky, V., Vakunov, A., Tkachenka, A., Sung, G., Chang, C., Grundmann, M.: Mediapipe hands: On-device real-time hand tracking. CoRR **abs/2006.10214** (2020). URL https://arxiv.org/abs/2006.10214
[7] Altman, N.S.: An introduction to kernel and nearest-neighbor nonparametric regression. The American Statistician **46**(3), 175–185 (1992). URL http://www.jstor.org/stable/2685209
[8] Köhler, S., Doll, K., Kebinger, S., Schmitt, D., Kröhn, M., Fried, M., Börsig, R.: Prozessinnovation: Hochautomatisiert lernendes Assistenzsystem für die manuelle Montage. AALE pp. 153 – 164 (2018)
[9] Köhler, S., Doll, K., Kebinger, S., Schmitt, D., Kröhn, M., Fried, M., Börsig, R.: Gestenerkennung in einem hochautomatisiert lernenden Assistenzsystem für manuelle Montageprozesse. In: AUTOMATION 2018, VDI-Berichte 2330, pp. 145 – 156 (2018)
[10] Petitjean, F., Forestier, G., Webb, G.I., Nicholson, A.E., Chen, Y., Keogh, E.: Dynamic time warping averaging of time series allows faster and more accurate classification. In: 2014 IEEE International Conference on Data Mining, pp. 470–479 (2014). DOI 10.1109/ICDM.2014.27
[11] Petitjean, F., Ketterlin, A., Gançarski, P.: A global averaging method for dynamic time warping, with applications to clustering. Pattern Recognition **44**(3), 678–693 (2011). DOI https://doi.org/10.1016/j.patcog.2010.09.013. URL https://www.sciencedirect.com/science/article/pii/S003132031000453X

Open Access. © 2022 J. Jung, A. Hubert, K. Doll, M. Kröhn, J. Stadler
This work is licensed under the Creative Commons Attribution 4.0 License. +