

# Abschlussbericht Selbstlernende und selbsterklärende Maschine SLEM

Zuwendungsempfänger: Fraunhofer Gesellschaft zur Förderung der angewandten Forschung e. V.; Fraunhofer IPA (ausführende Stelle)	Aktenzeichen: 36-3400.7/104
Vorhabentitel: Selbstlernende und selbsterklärende Maschine - SLEM	
Projektleiter: Christian Jauch	Tel.: +49 711 970-1816 E-Mail: christian.jauch@ipa.fraunhofer.de
Laufzeit: von: 01.01.2021                      bis: 31.12.2021	Berichtszeitraum: von: 01.01.2021                      bis: 31.12.2021

## 1 Einleitung

Das Ziel des Projekts »Selbstlernende und selbsterklärende Maschine (SLEM)« war die Entwicklung eines intelligenten Assistenzsystems, das Nutzer unterschiedlichster Expertisen adaptiv, sowohl bei der Bedienung von (Sonder-)Maschinen, als auch der Führung spezieller Produktionsabläufe von komplexen Produkten, unterstützt (vgl. Abbildung 1). Im Rahmen des Projekts wurden Untersuchungen und Entwicklungen zu KI-Methoden, insbesondere in den Bereichen kamerabasierte Aktivitätserkennung von Menschen, Maschinendatenanalyse, Fusion von Daten sowie Nutzerinteraktion vorgenommen und evaluiert.

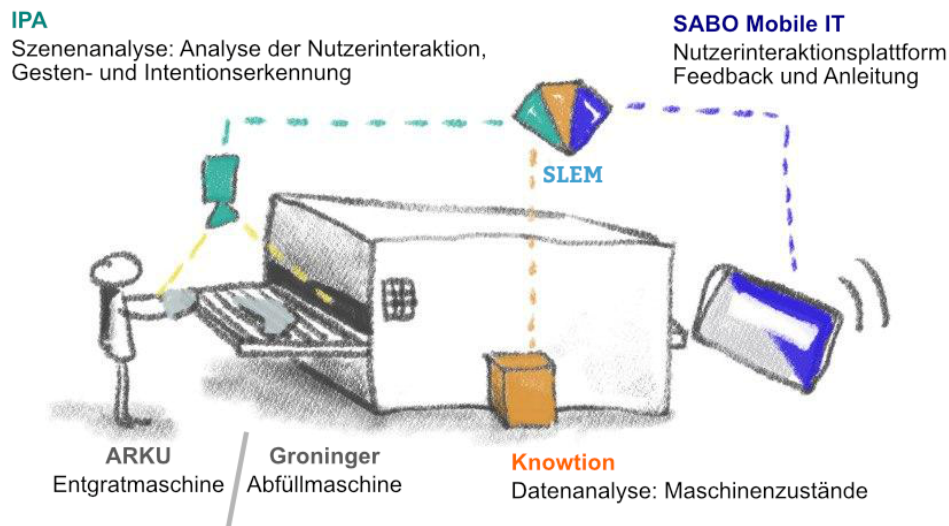


Abbildung 1: Vision SLEM

Im Folgenden werden die durchgeführten Arbeiten in den jeweiligen Arbeitspaketen erörtert. Durch die Corona-Pandemie kam es insbesondere in AP1 zu Verzögerungen. Da die folgenden Arbeitspakete auf AP 1 aufbauen, kam es auch bei diesen zu Verzögerungen, die zu Abweichungen von der ursprünglichen Planung führten. Diese werden im Weiteren ebenfalls genauer beschrieben.

## 2 Durchgeführte Arbeiten

### 2.1 Anforderungsanalyse aus Nutzer- und Systemsicht (AP1)

Zu Beginn fand zwischen den Projektpartnern Fraunhofer IPA, Knowtion und SABO Mobile IT (im weiteren SABO) sowie den Anwendungspartnern groninger und ARKU ein umfangreicher Austausch über Videokonferenzen statt. Zudem wurden von den Anwendungspartnern Handbücher und Videomaterial zur Verfügung gestellt. Dies diente dem Ziel, trotz Einschränkungen durch die Corona-Pandemie, einen Einblick in die jeweiligen Prozesse der Maschinennutzung zu bekommen und Use Cases für SLEM zu identifizieren. Als Ergebnis wurden für groninger der Use Case „Formatwechsel“ identifiziert und für ARKU der Use Case „Richten von Blechteilen“. Ersterer beinhaltet das Umrüsten der komplexen groninger-Maschinen, um sie auf eine andere Produktvariante vorzubereiten. Dieser Prozess ist aufwendig und aufgrund ähnlicher Teile fehleranfällig und fordert damit ein gewisses Maß an Erfahrung vom Nutzer. Letzterer bezeichnet die Nutzung einer ARKU Richtmaschine zum Richten von gekrümmten Blechteilen. Hier ist ebenfalls Erfahrung notwendig, um die Blechteile korrekt und möglichst effizient der Maschine zuzuführen.

Aufgrund der Corona-Pandemie konnten vor Ende Juni 2021 keine UNA durchgeführt werden. Um dennoch frühe Einblicke in die tatsächlichen Bedürfnisse der Nutzer zu bekommen, wurde ein Online-Fragebogen an groninger-Kunden verteilt. Dies lieferte erste Hinweise auf relevante Bedürfnisse bei der Maschinenbedienung, war aber aufgrund der kleinen Zahl der Rückmeldungen bedingt repräsentativ. Um dennoch mit der Bearbeitung des Projekts fortfahren zu können, wurde das Konzept und die Umsetzung des Gesamtsystems und dessen Pipeline vorgezogen. Zudem konnte mittels Dummy-Daten und Testaufnahmen am Fraunhofer IPA erste Vorbereitungstests durchgeführt werden. Ein grobes Bild der Nutzeranforderungen konnte letztlich aus den Online-Fragebögen, auf Basis eines Austauschs mit einem Ausbilder von groninger und auf Basis der Handbücher abgeleitet werden. Auch aus den Meetings mit ARKU gingen allgemeine Nutzerbedürfnisse hervor, die letztlich in den tatsächlichen UNA ausdetailliert wurden

Im Juni 2021 konnte ein UNA bei groninger vor Ort durchgeführt werden und im November vor Ort bei ARKU. Damit einhergehend erfolgte die Aufnahme eines kombinierten Mensch-Maschinendatensatzes

für beide Anwendungsfälle. In Abbildung 2 sind die verwendeten Sensoren zu sehen. Eye-Tracking-Sensorik sowie Tiefen- und RGB-Kameras dienten dazu,



Abbildung 2: Tobii Pro Glasses 3 (Eye-Tracking), Intel Realsense D435 (RGB- und Tiefenkamera), Microsoft Azure Kinect DK (RGB- und Tiefenkamera).  
Bildquellen: <sup>1 2 3</sup>

möglichst viele Informationen zu erfassen und daraus die Nutzerbedürfnisse abzuleiten. Diese Daten wurden in den Arbeitspaketen 2,3 und 4 zudem für datenbasierte KI-Verfahren verwendet.

<sup>1</sup> <https://www.tobii.com/product-listing/tobii-pro-glasses-3/>

<sup>2</sup> [https://store.intelrealsense.com/buy-intel-realsense-depth-camera-d435.html?\\_ga=2.130705289.766125610.1645433668-885824115.1631710270](https://store.intelrealsense.com/buy-intel-realsense-depth-camera-d435.html?_ga=2.130705289.766125610.1645433668-885824115.1631710270)

<sup>3</sup> <https://www.microsoft.com/en-us/d/azure-kinect-dk/8pp5vxmd9nhq>

Bei den Ergebnisreports der Nutzerbedürfnisanalyse für ARKU und groninger lag im Fokus, wie Wissen im Umgang mit der Maschine erlangt werden kann. Welches Grundwissen muss vorhanden sein, um die Maschine für einen Anfänger bedienbar zu machen? Wie kann dieses Wissen präsentiert werden (z.B. durch Einführungsvideos) und wie kann weiterführendes Wissen erlangt werden? Zusätzlich befassten sich die Reports damit, welche weiteren Phasen in der Wissensaneignung durchlaufen werden, um sich vom Anfänger zu einem fortgeschrittenen Bediener bis hin zum Experten weiterzuentwickeln.

In diesem Sinne wurden Persona und User Journey Maps erstellt, anhand deren man erkennen konnte, welches implizite Wissen gelernt wurde um Expertenstatus zu erlangen. Dies ist besonders wichtig, da ein Assistenzsystem notwendiges Grundwissen vermitteln soll um die Maschine bedienbar zu machen. Danach mussten klare Phasen abgegrenzt werden, in denen bestimmte Lerninhalte vermittelt werden.

Zusätzlich wurden die Eye-Tracking-Daten ausgewertet (vgl. Abbildung 3), um zu überprüfen, wie sich das Verhalten von Anfängern sowie von fortgeschrittenen Nutzern verhält. Dieses Vorgehen konnte in den Reports Schwächen der Displays der Maschinen aufzeigen und gab den Firmen Vorschläge zur möglichen Verbesserung der Interaktion mit den Maschinen.

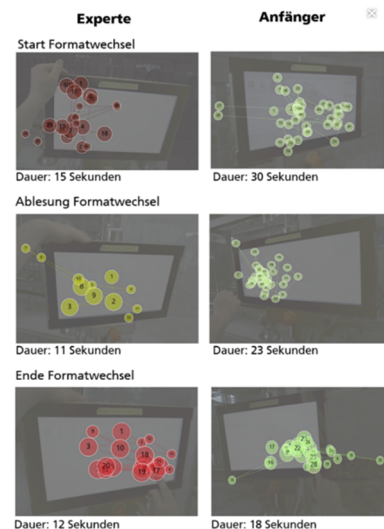


Abbildung 3: Augendatenauswertung des Displays bei groninger

Der geplante Meilenstein für AP1 wurde mit „Ergebnisreport der Nutzerbedürfnisanalyse, Anforderungen an menschenzentrierte KI, Persona und User Journey Maps für mehrere Bedienszenarien liegen vor“ festgelegt. Dieser Meilenstein konnte trotz Verzögerungen durch Corona erfüllt werden. Es liegen detaillierte Erkenntnisse über die beiden Bedienszenarien bei ARKU und groninger vor und diese dienten als Grundlage für die weiteren Arbeitspakete. Des Weiteren war die Datensatzerstellung auf Basis der UNAs ein Punkt der wichtig und mit großem Aufwand verbunden war, aber in der ursprünglichen Planung nicht aufgelistet war.

## 2.2 Szenenanalyse: Interaktions- und Intentionserkennung (AP2)

In AP2 wurden ein Ansatz zur posenbasierten Aktivitätserkennung mittels RGB-Kameras umgesetzt und für die in SLEM betrachteten Anwendungsfälle der assoziierten Partner angepasst. Durch die Verzögerungen in der Datenaufnahme mussten Teile der Interaktionserkennung

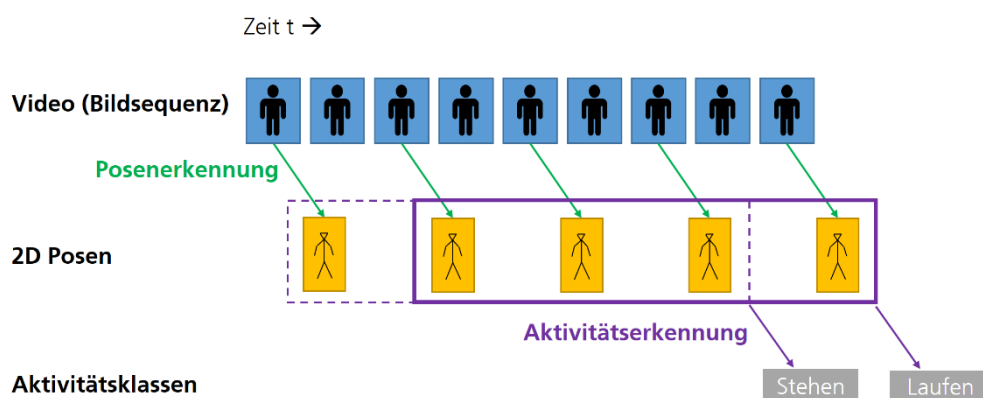


Abbildung 4: Schema posenbasierte Aktivitätserkennung im Online-Szenario

depriorisiert werden, daher wurde keine kombinierte Objekt- und Posenerkennung für die Interaktionserkennung eingesetzt. Stattdessen wurde eine rein auf Posendaten basierte Aktivitätserkennung verwendet.

Die Posendaten stellen die Positionen einzelner Gelenke des menschlichen Körpers dar und werden durch KI-gestützte Verfahren zur Posenerkennung erfasst. In Abbildung 4 ist das Schema der posenbasierten Aktivitätserkennung visualisiert. Wir betrachten für SLEM ein Online-Szenario, das heißt die Aktivitäten müssen ständig und in Echtzeit auf dem ankommenden Videostream detektiert werden. Dafür werden zu jedem Zeitpunkt die letzten N (beispielsweise N=8) 2D-Posen für die Aktivitätsklassifikation verwendet. Diese 2D-Posen können zur Verringerung der Rechenzeit bspw. nur auf jedem zweiten oder jedem dritten Frame (= Bild) des Videos berechnet werden. Weiterhin wurde eine solche posenbasierte Aktivitätserkennung mittels der Frameworks ONNXRuntime und PyTorch in das Gesamtsystem aus AP5 integriert.

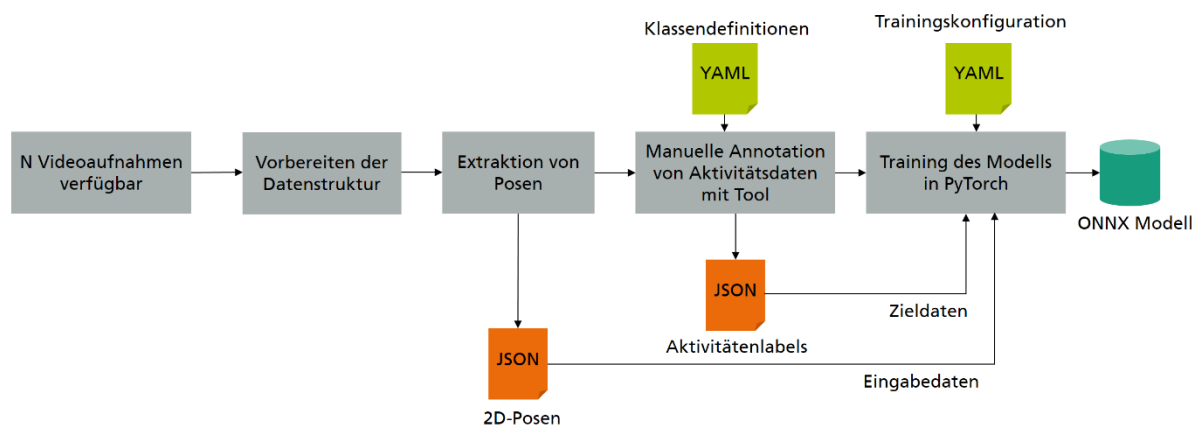


Abbildung 5: Aktivitätserkennung Trainingspipeline

Bevor das Aktivitätserkennungsmodell in SLEM eingesetzt werden kann, muss es jedoch zunächst auf einem Datensatz trainiert werden. Dazu wurde eine einfach zu verwendende Pipeline (siehe Abbildung 5) umgesetzt, mit der die posenbasierte Aktivitätserkennung auf beliebige

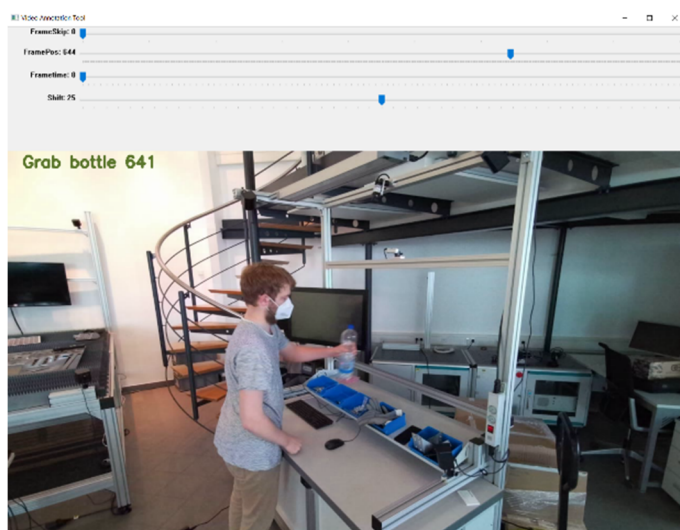


Abbildung 6: Videoannotationstool

Anwendungsfälle im Bereich der manuellen Montage anwendbar ist. Dabei können unterschiedlichste Aktivitäten an der Maschine erlernt werden. Es wurde zudem ein Annotationstool implementiert, um Videodaten mit möglichst wenig Aufwand zu annotieren (siehe Abbildung 6). Dieser Teil der Pipeline ist besonders aufwendig und bietet viel Potenzial zur Verringerung des Aufwands.

Im Verlauf des Projekts wurden zwei Datensätze aufgenommen und annotiert. Der annotierte Datensatz für den Fall „Formatwechsel“ beinhaltet Aufnahmen eines Experten und eines Anfängers, während der annotierte Datensatz „Blech richten“ aus Aufnahmen

von drei Probanden besteht: Experte, Fortgeschrittener und Anfänger. Tabelle 1 zeigt die definierten Aktivitätsklassen der jeweiligen Anwendungsfälle. Die Definition von klar getrennten Aktivitäten auf Basis von Videos aus bestehenden (kontinuierlichen) Arbeitsprozessen ist herausfordernd und beeinflusst die Qualität der Datensätze und damit der Modelle stark. Es empfiehlt sich aber dennoch, die Aktivitäten zunächst möglichst kleinteilig (bspw. Objekt aufnehmen, tragen, herausziehen) zu annotieren und falls nötig im Nachhinein wieder zusammenzulegen (bspw. Interaktion mit Objekt).

*Tabelle 1: Übersicht der Aktivitätsklassen für die Anwendungsfälle*

<b>Klassen „Formatwechsel“</b>	<b>Klassen „Blech richten“</b>
<ul style="list-style-type: none"> <li>• Nicht an der Maschine</li> <li>• Ruhend</li> <li>• Formatteil holen</li> <li>• Türe öffnen</li> <li>• Türe schließen</li> <li>• Objekt aufnehmen</li> <li>• Objekt tragen</li> <li>• Objekt herausziehen</li> <li>• Objekt ablegen</li> <li>• Schraube/Ventil öffnen</li> <li>• Schraube/Ventil schließen</li> <li>• Interaktion mit Maschinentouchdisplay</li> <li>• Herausziehen</li> <li>• Einfügen</li> <li>• Hebel entsichern</li> <li>• Hebel sichern</li> </ul>	<ul style="list-style-type: none"> <li>• Ruhend</li> <li>• Gehen</li> <li>• Messen mit Werkzeug</li> <li>• Blechteil neu arrangieren</li> <li>• Blechteil mit Händen testen</li> <li>• Objekt aufnehmen</li> <li>• Objekt ablegen</li> <li>• Interaktion mit Maschinentouchdisplay</li> <li>• Einführen des Blechteils in die Maschine</li> </ul>

Als Modellarchitekturen wurden drei verschiedene Modelle für die posenbasierte Aktivitätserkennung ausgewählt: Ein simples MLP (10 Schichten), PoseC3D<sup>4</sup> und ST-GCN<sup>5</sup>. ST-GCN lieferte ähnliche Genauigkeiten wie PoseC3D nach den Trainings und wird daher im Folgenden nicht näher betrachtet. Das MLP lieferte eine Genauigkeit von 14 % für den „Formatwechsel“ Anwendungsfall.

Im Anwendungsfall „Formatwechsel“ wurden für den Trainingsdatensatz ausschließlich Expertendaten verwendet und für die Evaluation ausschließlich Anfängerdaten. Hierbei wurden mit PoseC3D Genauigkeiten von 17 % erreicht. Dieses Ergebnis ist darauf zurückzuführen, dass reine Posendaten keine Informationen über den Kontext der Aktion darstellen. Somit kann ein Objekt, eine Türe oder eine Schraube nur schwer voneinander unterschieden werden, da viele damit zusammenhängende Aktionen sehr ähnlich aussehen. Dieses Problem kann zukünftig durch das Einbeziehen von Objektdetektion in die Aktivitätserkennung voraussichtlich gelöst werden. Aufgrund der niedrigen Genauigkeit wurden die Klassen im Datensatz reduziert und zusammengefasst, sodass am Ende 5 Klassen blieben. Auf diesem reduzierten Datensatz erreicht PoseC3D bis zu 63 % Genauigkeit.

Für den Fall „Blech richten“ wurden Trainings mit unterschiedlichen Varianten des Datensatzes durchgeführt: Ein Training mit Daten von allen 3 Probanden („gemischt“) und ein Training ausschließlich mit Expertendaten („Experte“). Das Experten-Modell übertrifft das gemischt trainierte Modell mit 50,8 % Genauigkeit im Vergleich zu 45,6 % (vgl. Tabelle 2). Der Gesamt-

<sup>4</sup> Duan et al., 2021. Revisiting Skeleton-based Action Recognition. <https://arxiv.org/abs/2104.13586v1>

<sup>5</sup> Yan et al., 2018. Spatial Temporal Graph Convolutional Networks for Skeleton-Based Action Recognition. <https://arxiv.org/abs/1801.07455v1>

Testdatensatz enthält jedoch mehr Expertendaten als Anfängerdaten, daher ist dieses Ergebnis nur bedingt repräsentativ. Klarer wird das Ergebnis, wenn die Testdaten in die jeweiligen Propandentypen aufgeteilt werden. Erwartungsgemäß erreicht das Expertenmodell bessere Werte auf dem Expertentestdatensatz. Dennoch bleibt es für den Anfängerdatensatz nah beim gemischten Modell. Obwohl es lediglich auf den Daten des Experten trainiert wurde, waren die Modelle in der Lage auf andere Personen zu generalisieren. Dies ermöglicht es, mit kleinen Datensätzen zu arbeiten und einen geringen Aufwand für die Annotation leisten zu müssen. Augmentierungsmethoden könnten diesen Aufwand bzw. die Performance der Modelle noch weiter verbessern.

Tabelle 1: PoseC3D Genauigkeitsergebnisse nach Training auf Experten- und Mischdatensatz

Modell	Genauigkeit			
	Gesamt	Anfänger	Fortgeschrittener	Experte
PoseC3D „gemischt“	0,456	<b>0,554</b>	0,418	0,449
PoseC3D „Experte“	<b>0,508</b>	0,433	<b>0,487</b>	<b>0,584</b>

In weiteren Experimenten wurde die Laufzeit des PoseC3D Modells untersucht. Bei diesem Modell werden die Posen aus der Videosequenz zu Heatmaps konvertiert und dann ein 3D-CNN verarbeitet. Dies ermöglicht hohe Genauigkeiten, benötigt jedoch eine große Menge an Rechenleistung. Daher wurden Experimente mit unterschiedlichen Konfigurationen durchgeführt (vgl. Tabelle 3). Hierbei sieht man, dass die Genauigkeit bei einer geringeren Menge an verwendeten Frames aus dem Video mit größerem zeitlichen Abstand nur geringfügig abweicht. Im Fall PC3D-16-3 wird nur in jedem dritten Frame die Pose berechnet und an PoseC3D übergeben, letztlich 16 Posen pro Ausführung. Nur mit der Konfiguration PC3D-4-12 kann eine Echtzeitfähigkeit gewährleistet werden, da die Gesamtlatenz unter 1,6 Sekunden liegt und damit innerhalb des Zeitraums der  $16 \times 3 = 48$  Frames bei 30 Frames pro Sekunde. Dies ermöglicht einen Einsatz als Echtzeitaktivitätserkennung in SLEM.

Tabelle 2: Genauigkeits- und Latenergebnisse für PoseC3D auf einer NVIDIA Tesla V100. Latenz „Einfach“ gibt die Zeit für eine einzelne Ausführung des PoseC3D Modells ohne Posenerkennung an. Latenz „Gesamt“ gibt die Gesamtzeit einschließlich Posenerkennung über einen Zeitraum von 1,6 Sekunden an (Online-Szenario).

Modell	Genauigkeit	Latenz Einfach (in ms)	Latenz Gesamt (in ms)
PC3D-48-1	0,508	161	18144
PC3D-24-2	0,504	94	7464
PC3D-16-3	<b>0,513</b>	56	4368
PC3D-8-6	0,491	31	1984
PC3D-4-12	0,498	<b>19</b>	<b>940</b>

Als ursprünglicher Meilenstein für AP2 wurde „Aufgebaute Simulationsumgebung und Erkennungsmethoden zur Erfassung von Montagebewegungen vorhanden“ definiert. Dieser Meilenstein konnte teilweise erreicht werden. Es wurden zwar unterschiedliche Erkennungsmethoden zur Erfassung von Montagebewegungen untersucht, jedoch konnte aufgrund von Verzögerungen und unvorhergesehenen technischen Herausforderungen, wie das notwendige Nachbereiten der aufgenommenen Daten, keine Simulationsumgebung umgesetzt werden. Eine weitere Abweichung von der ursprünglichen Planung ist der Verzicht auf eine

Intentionserkennung, also eine Vorhersage zukünftiger Aktivitäten. Diese wurde zugunsten einer robusten Basisaktivitätserkennung zurückgestellt.

### 2.3 Maschinendatenanalyse (AP3)

Als weitere Informationsquelle wurden die Maschinendaten neben den Videodaten aufgezeichnet. Diese wurden dann mit den Videodaten zusammengeführt, wie in Abbildung 10 zu sehen. Bei groninger wurde hierfür eine Vorauswahl an Variablen getroffen um die Anzahl der Variablen in einer überschaubaren Größe zu halten. Von ca. 15.000 Variablen der Maschine

wurden 279 Größen ausgewählt. Bei ARKU wurden alle verfügbaren Größen (10) aufgezeichnet. Nach Corona-bedingten Verzögerungen wurden während des Besuches bei groninger vor Ort besagte Maschinendaten mit aufgezeichnet. In ersten Analysen zeigte sich, dass die Hälfte der Variablen über den gesamten Aufnahmezeitraum ohne Wertänderung blieb und während des Formatwechsels lediglich Temperatursensoren

leicht schwankten. Mittels k-Means-Clustering wurden weitere Maschinendaten außerhalb des Besuchstages in Clustern zusammengefasst und diese Daten im Anschluss zum Training eines LSTM-Modells (long short-term memory) verwendet, da eine zeitliche Abhängigkeit der Arbeitsschritte gegeben ist. Für die Einteilung in Cluster wurde die Silhouette-Methode angewandt, um eine passende Anzahl Cluster zu finden (vgl. Abbildung 7). Hierbei werden die Abstände der einzelnen Datenpunkte zueinander und zum nächstgelegenen Cluster bestimmt. Eine gleichmäßige Verteilung mit einem Koeffizient  $> 0,6$  deutet auf eine geeignete Clusteranzahl hin

Die mit dem LSTM-Modell vorhergesagten Cluster wurden dann nach Möglichkeit explizit mit einem Maschinenstatus versehen wie beispielsweise „Maschine in Betrieb“ und der Fraunhofer-Software zur Zusammenführung mit den Videodaten zur Verfügung gestellt. In einer rudimentären Analyse durch SABO konnte ein zusätzlicher Informationsgehalt von etwa 3 % erreicht werden.

Die Erfahrungen im Umgang mit den Daten von groninger wurden genutzt um für den Besuch bei ARKU ein gezieltes Versuchsprogramm für die Gewinnung und spätere Verwendung der Maschinendaten durchzuführen. Hierfür wurde von ARKU eine Testmaschine zur Verfügung gestellt, die es erlaubte, gezielt Versuche durchzuführen, bei dem einzelne Größen nacheinander verändert werden. Neben der automatisierten Datenaufzeichnung wurden weitere Größen manuell festgehalten und das Ergebnis (Richtvorgang erfolgreich oder nicht) notiert. Im ersten

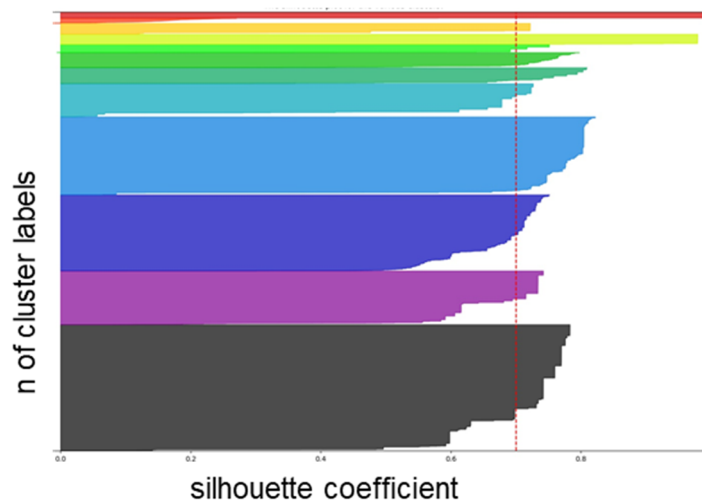


Abbildung 7: Silhouette-Methode für ein k-means-Clustering mit 12 Clustern

Schritt wurden die Zeitstempel der aufgenommenen Daten mit den manuell erfassten Daten

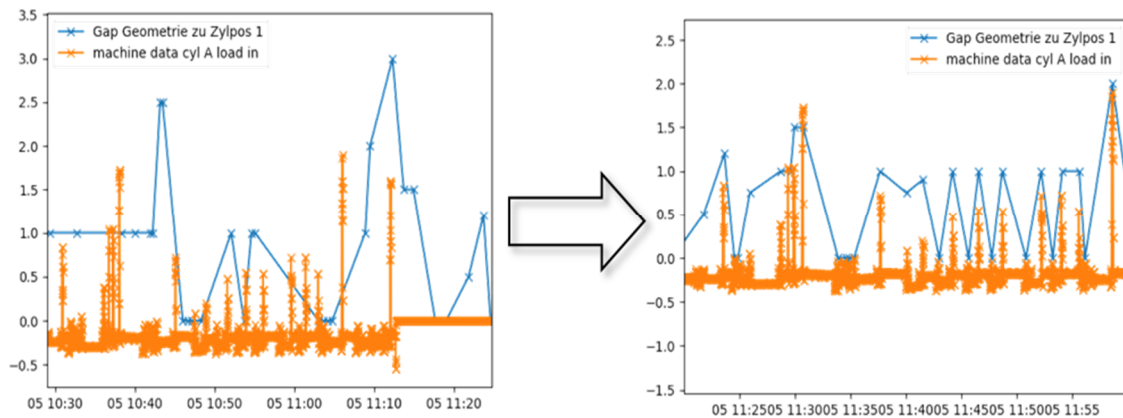


Abbildung 8: Angleichen der manuell festgehaltenen Daten mit den Maschinendaten

angeglichen wie in Abbildung 8 zu sehen. Mit den Maschinendaten wurden dann zwei neuronale Netze trainiert, eines um ein Richterergebnis bei bekannten Daten zu prognostizieren und das andere um die Dicke des Werkstücks zu bestimmen, da beide Größen nur manuell erfasst werden können.

Der Meilenstein von AP3 wurde in der ursprünglichen Planung als „Software zum Empfang der Maschinendaten und Bestimmung des Maschinenzustands prototypisch entwickelt“ festgelegt. In der Ausführung von SLEM wurden innerhalb AP3 die aufgezeichneten Maschinendaten verarbeitet und analysiert. Dabei konnten Maschinenzustände grob erkannt werden, was einer prototypischen Bestimmung des Maschinenzustands entspricht. Von der direkten Maschinenanbindung wurde auf Grund der Zeitverzögerung abgesehen, für genauere Vorhersagen sind weitere Messungen und Anpassungen an z.B. Variablenauswahl nötig.

## 2.4 Entwicklung einer Interaktionsplattform (AP4)

Im Rahmen des SLEM Projekts wurde ein Konzept zur Nutzerführung entwickelt. Die Zielsetzung war dabei vorhandenes Expertenwissen zu erfassen und mit Hilfe der Interaktionsplattform an unerfahrene Operatoren weiterzugeben. Dabei soll durch das Beobachten eines Experten und das gleichzeitige Analysieren der Maschinendaten eine Bedienungsanleitung gewonnen werden, die dann einem unerfahrenen Nutzer zur Verfügung gestellt werden kann. Außerdem werden die gewonnenen Daten genutzt um verschiedene neuronale Netze zu trainieren und deren Vorhersage für die Nutzerführung zu verwenden. Unter diesem Gesichtspunkt wurde eine Plattform entwickelt, die dies ermöglicht. Dabei wurde in einem ersten Schritt ein Verfahren entwickelt, mit dessen Hilfe die durchgeführten Arbeitsschritte eines Experten erfasst und gespeichert werden können. Hierzu wurde ein neuronales Netz trainiert, welches die von AP2 und AP3 bereitgestellten Aktivitäts- und Maschinendaten in Arbeitsschritte zusammenfasst. Hierfür wurde ein Ansatz des überwachten Lernens verwendet, bei dem die korrelierten Daten aus Aktivitäts- und Maschinendaten von AP2 und AP3 verschiedenen Arbeitsschritten zugeordnet wurden. Mit Hilfe dieser gelabelten Daten wurde dann ein LSTM Model trainiert. Diese Architektur erlaubt es, zeitliche Zusammenhänge in die Vorhersage einfließen zu lassen. Mit Hilfe des Modells ist SLEM dazu in der Lage einzelne Arbeitsschritte aus der Expertentätigkeit zu extrahieren.



Der nächste Schritt umfasst eine automatische Textgenerierung. Hier wurde mit Hilfe eines neuronalen Netzes zur Textgenerierung (in diesem Falle GPT-2) eine Bedienungsanleitung, auf Basis der zuvor detektierten Arbeitsschritte, erstellt. Diese Bedienungsanleitung kann dann dem nächsten, möglicherweise unerfahrenen, Nutzer bereitgestellt werden. Derzeit wird die Bedienungsanleitung nur in Textform bereitgestellt. Für die Zukunft ist es denkbar diese auch per Sprachausgabe oder andere innovative Darstellungsformen zu vermitteln.

Zusätzlich zur Bedienungsanleitung, welche in die einzelnen Arbeitsschritte aufgeteilt ist, wurde außerdem noch eine Anomaliedetektion trainiert. Hierfür wurde ein Ansatz des unüberwachten Lernens gewählt. Die Netzwerkarchitektur war dabei ein LSTM Autoencoder. Dieser ist dazu in der Lage Fehler des Operators zu erkennen und ihn darauf aufmerksam zu machen. Hierdurch kann ein Bediener mit wenig Erfahrung sicher und effizient durch die einzelnen Prozessschritte geführt werden. Um das mögliche Fehlverhalten des Operators zu erkennen, wird auf die zuvor in AP2 und AP3 ermittelten Daten zurückgegriffen und in Echtzeit mit dem erwarteten Arbeitsschritt verglichen.

Zusätzlich wird dabei der Aspekt der selbstlernenden Maschine berücksichtigt. Die Daten, die bei der Bedienung der Maschine durch einen erfahrenen Operator gesammelt werden, werden kontinuierlich dazu genutzt die bestehenden neuronalen Netze weiter zu optimieren und ihre Performance zu verbessern. Die komplette Pipeline zur Nutzerführung kann dabei mit real erfassten Daten oder mit simulierten Szenarien, die im Zuge des Projekts erstellt wurden, genutzt werden.

Zusammengefasst liefert AP4 die automatische Generierung einer Bedienungsanleitung sowie eine Anomaliedetektion, die Fehler im Arbeitsfluss eines unerfahrenen Nutzers erkennt und Korrekturen vorschlägt.

Der Meilenstein für AP4 war „Modell bereit für Erprobung in der Realität“. Dieser Meilenstein wurde erreicht, da die Modelle vorliegen. Allerdings sollte darauf geachtet werden, dass mehr Daten zur Verfügung stehen. Wie schon eingangs erwähnt, war die Datenaufnahme durch die Corona Pandemie nur eingeschränkt möglich. Unter diesem Gesichtspunkt ist das erreichte Ergebnis als sehr gut anzusehen.

Eine wesentliche Abweichung zum geplanten Vorgehen innerhalb des AP4 war, dass es nicht möglich war den Prozess ausführlich zu testen und zu evaluieren. Hierfür war die Datenlage, in Bezug auf reale Daten, einfach zu gering. Dies ist wesentlich auf die Einschränkungen durch die Pandemie im Jahr 2021 zurückzuführen.

## 2.5 Gesamtsystem und Demonstrator-Aufbau mit Evaluation (AP5)

Das Gesamtsystem wurde bereits in frühen Konzeptworkshops entworfen. Der Fokus lag vor allem darauf, SLEM möglichst allgemein zu entwickeln, sodass SLEM nicht nur für die Anwendungsfälle von ARKU und groninger eingesetzt werden kann, sondern für weitere Maschinen anderer Hersteller. Abbildung 9 zeigt einen Ausschnitt aus den Ergebnissen des Workshops. Hierbei wurden die Inhalte, Herausforderungen sowie Schnittstellen in und zwischen den Arbeitspaketen bzw. Softwaremodulen erarbeitet. Weiterhin wurde für eine Basisversion der Softwarearchitektur entworfen, die in Abbildung 10 zu sehen ist. SLEM wurde in einer monolithischen

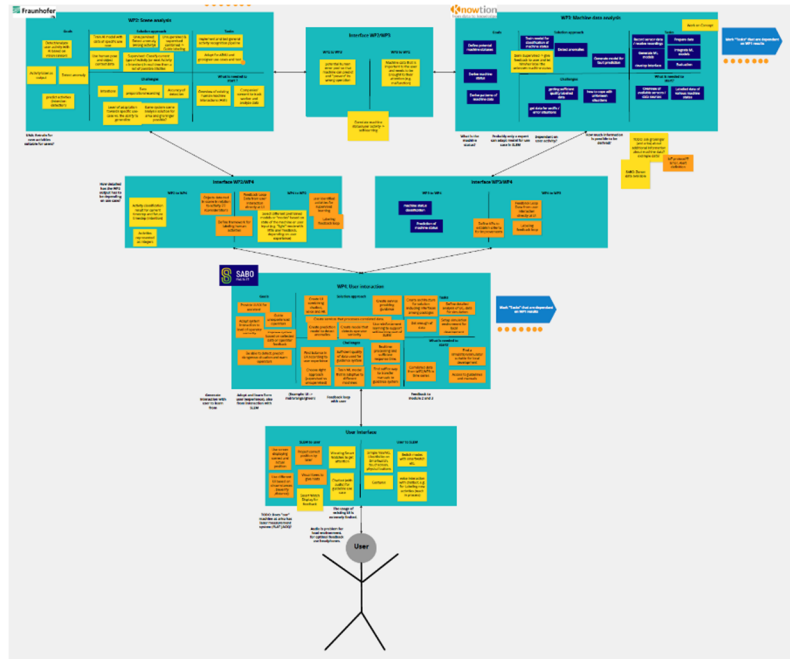


Abbildung 9: Systementwurf mit Fokus auf Zielen, Herausforderungen, Aufgaben, Voraussetzungen und Schnittstellen der einzelnen Module bzw. Arbeitspakete

Softwarearchitektur implementiert. Dabei hat jeder der drei Projektpartner ein separates Modul erstellt. Die drei Module werden parallel in drei separaten Prozessen ausgeführt. Die Prozesse können untereinander mit Hilfe mehrerer FIFO-Queues (First in - First out) kommunizieren. Dadurch können die Module ihre Ergebnisse parallel ermitteln und dann bei Bedarf an die weiteren Komponenten weitergeben. Die Prozesse werden von einem übergeordneten SLEM Service gesteuert. Außerdem wurde eine Möglichkeit geschaffen, die einzelnen Module und ihre Nachrichten, welche durch die Queues ausgetauscht werden, zu simulieren. Dadurch war es möglich, unabhängig von den Projektpartnern, eigene Features zu entwickeln. Des Weiteren wurde eine Möglichkeit entwickelt, die das Testen der kompletten SLEM Pipeline erlaubt. Dazu wurde ein Verfahren geschaffen, mit dem bestehende Videoaufnahmen und Maschinendaten in SLEM eingelesen werden können. Dadurch verhält sich SLEM genauso, als würde es die Daten direkt an einer Maschine erfassen. Es wurde ein funktionierender SLEM Software-Demonstrator entwickelt, welcher notfalls auch ohne direkte Maschinenverbindung, mit Hilfe von zuvor aufgenommenen Daten, vorgeführt werden kann.

Der Meilenstein von AP 5 „Prototyp einer SLEM, Report der qualitativen und quantitativen Evaluationsergebnisse“ konnte maßgeblich erreicht werden, da ein bestehender Software-Demonstrator inklusive quantitativer und qualitativer Ergebnisse der einzelnen Komponenten vorliegt. Um den finalen Demonstrator tiefgehend evaluieren zu können, wären aber noch weitere Daten, insbesondere auch von zusätzlichen Anwendungsfällen notwendig. Dadurch könnte auch die Generalisierungsfähigkeit des Ansatzes genauer untersucht werden. Aufgrund der Corona-Pandemie waren die Datenaufnahmemöglichkeiten stark erschwert.

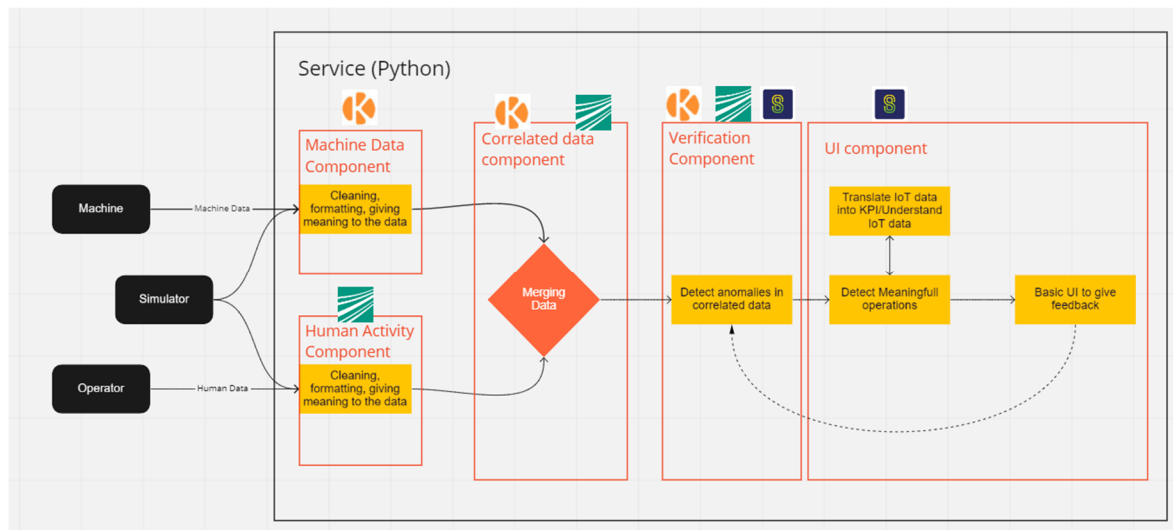


Abbildung 10: Softwarearchitektur des SLEM-Basis

## 2.6 Projektmanagement und Querschnittsaktivitäten (AP6)

Mit Regelmeetings, dedizierten technischen Meetings sowie zusätzlichen Workshops erfolgte die Zusammenarbeit vielfältig und direkt. Fraunhofer IPA organisierte und führte diese Treffen. Mit den Anwendungspartnern wurden ebenfalls gemeinsame Treffen organisiert, sowohl online als auch vor Ort. Die Kommunikation mit den Anwendungspartnern wurde gemeinsam von IPA, Knowtion und SABO durchgeführt. Zudem wurden aktuelle Aufgaben mit einem SCRUM-ähnlichen Verfahren organisiert und in 2- bzw. 3-wöchigen Sprints bearbeitet. Dabei fand ein Austausch über unterschiedlichste KI-bezogene Methoden und Werkzeuge statt, welcher allen Beteiligten einen Mehrwert bot.

AP6 beinhaltet den Meilenstein „Wissenschaftliche Ergebnisse wurden auf Fachkonferenzen und in Fachzeitschriften veröffentlicht, Workshops und Seminare wurden durchgeführt“. Dieser Meilenstein wird bis Ende des Jahres 2022 vollständig erreicht sein. Bereits am 1. Dezember 2021 wurde beim Symposium „Digitalisierung souverän gestalten II“ des Instituts für Innovation und Technik Berlin SLEM vorgestellt und im Workshop als Anwendungsbeispiel verwendet. Aktuell ist außerdem ein Artikel in einer Fachzeitschrift geplant. SLEM wird außerdem als Teil der ML Prague<sup>6</sup> Konferenz vorgestellt. Für die CPSL 2022<sup>7</sup> Konferenz wurde zudem ein Paper eingereicht, was den Aktivitätserkennungsteil von SLEM näher beschreibt. Es befindet sich aktuell in der Peer-Review-Phase. Weiterhin soll SLEM in einer Eventreihe in Form eines Vortrags vorgestellt werden.

<sup>6</sup> <https://www.mlprague.com/>

<sup>7</sup> <https://cpsl-conference.com/#home>

### **3 Verbreitung und wirtschaftliche Verwertbarkeit der Forschungsergebnisse**

Die Projektergebnisse wurden und werden in verschiedensten Formaten veröffentlicht. Es besteht bereits eine Website für SLEM<sup>8</sup>, die mit den Projektergebnissen aktualisiert wird. Auf der SPS 2021 war eine Vorstellung von SLEM durch SABO geplant, musste aber wegen der Corona-Pandemie kurzfristig abgesagt werden. Zusätzlich wurden in Abschnitt 2.6 bereits weitere geplante Veröffentlichungen und Events erwähnt.

SABO wird die Ergebnisse bei seinen aktuellen und potentiellen Kunden mit der Zielsetzung vorstellen um neue Projekte zu initiieren, die auf den Ergebnissen von SLEM aufbauen. Der Kundenkreis von SABO besteht aus einigen Hidden Champions und ist in Baden-Württemberg stark vertreten sowie international aktiv. Hier wird erwartet, dass das SLEM-Konzept auf Interesse stößt und neue Projekte entstehen, die das SLEM-Prinzip mit neuen Use Cases erweitern und die internationale Wettbewerbsfähigkeit der beteiligten Kundenunternehmen steigern.

Durch die Auswertung der Maschinendaten hat Knowtion wesentliche Erkenntnisse und Knowhow im Umgang mit Abfüllanlagen und Richtmaschinen gewonnen. Diese Erfahrung im Umgang mit den Daten dieser Anwendung sollen gezielt für Folgeprojekte eingesetzt werden. Außerdem sind die Auswertungen der Maschinendaten von ARKU besonders vielversprechend, es könnten durch weitere Versuche zuverlässige Zusatzinformationen aus den Daten gewonnen werden (z.B. Ermittlung der Werkstückdicke). Für eine Verwendung der Maschinenstaterkennung bei groninger müssen weitere Variablen gefunden werden, die im Formatwechsel Informationen zur Unterstützung des Bedieners bereithalten.

Das Fraunhofer IPA steht den Projektpartnern als weiterer Entwicklungspartner zur Verfügung und unterstützt diese bei der Verwertung der eigenen Ergebnisse soweit möglich und nötig. Im Projekt konnte das Fraunhofer IPA bestehende Methoden der Datengetriebenen User Needs Assessments in AP 1 validieren und erweitern. Damit kann Unternehmen in der Region und darüber hinaus die Möglichkeit gegeben werden, mit Hilfe des Fraunhofer IPA diese Nutzerbedürfnisanalysen speziell für industrielle Umgebungen durchzuführen.

Die entwickelten Methoden und Werkzeuge aus AP2 können ebenfalls in zukünftigen Projekten eingesetzt werden und bilden eine Basis für weitere Anwendungsfälle und Projekte. Das Fraunhofer IPA kann dadurch seinen Kunden, insbesondere KMUs zielgerichtet und preiswert Lösungen basierend auf Aktivitätserkennung in industriellen Umgebungen anbieten und diese bei der Entwicklung von Lösungen oder bei der Optimierung der eigenen Fertigung unterstützen. Dies ermöglicht auch die Entwicklung weiterer KI-Anwendungen für die Produktion und darüber hinaus.

### **4 Öffentlichkeitsarbeit, Netzwerkbildung & Wissenstransfer**

SABO hat Finanzmittel für eigne Aktivitäten zur Verbreitung der Ergebnisse des SLEM-Projektes und für Teilnahme gemeinsamen Aktivitäten der Partner eingeplant. Vorgesehen sind Artikel in Fachzeitschriften, Teilnahme an Kongressen und Messen, Publikation im Internet auf der Homepage mit SEO Maßnahmen, Blogs und Verbreitung in geschäftlich genutzten sozialen Netzwerken.

Sabo identifiziert Kunden und Marktsegmente, in denen durch SLEM Kundennutzen zu erwarten ist, und wird diese direkt ansprechen. Dabei wird SABO den Projektpartnern anbieten,

---

<sup>8</sup> <https://www.slem-projekt.de/>

entsprechend ihrer Kompetenzen und Möglichkeiten bei der Entwicklung von Mensch-Maschinen-Schnittstellen nach dem SLEM-Prinzip mitzuwirken und so das gemeinsame Wissen zum SLEM-Prinzip durch die Applikation in Kundenprojekten stetig zu erweitern.

Das Fraunhofer IPA verbreitet die Ergebnisse auf seinen wissenschaftlichen Kanälen und bei Informationsveranstaltungen und Messen. Außerdem beteiligt sich das Fraunhofer IPA auch bei der gemeinsamen Verbreitung der Ergebnisse. Zudem wird es gerne die Projektpartner bei Kundenprojekten bezüglich SLEM unterstützen und auch selbst die Projektpartner, sofern für den Kunden sinnvoll, in Entwicklungs- und Forschungsprojekte involvieren und so das in SLEM aufgebaute Netzwerk stärken und den Wissenstransfer zwischen den Partnern ausbauen und festigen.

Knowtion wird sich ebenfalls an der Verbreitung der Ergebnisse durch gemeinsame Aktivitäten beteiligen. Die Projektpartner werden für SLEM-Anwendungen in Entwicklungs- und Forschungsprojekten sowie auch Kundenprojekten bestmöglich unterstützt, um das gemeinsam errichtete Netzwerk zu stärken und den Wissenstransfer aufrecht zu erhalten. Auch wird Knowtion die Projektpartner entsprechend ihrer Kompetenzen zu Aktivitäten, die sich mit der SLEM-Architektur befassen selbstverständlich hinzuziehen.

## **5 Zusammenfassung**

Im Projekt SLEM wurde ein aus mehreren Komponenten bestehendes KI-Assistenz-System entwickelt, welches von Experten lernt und diese Informationen an unerfahrene Werker weitergeben kann. Mehrere KI-Methoden (posenbasierte Aktivitätserkennung, Maschinendaten-Clustering, diverse LSTMs) wurden kombiniert, um das Gesamtsystem in zwei realen Industrieszenarien umzusetzen. SLEM kann zukünftig auch in anderen Anwendungsfällen eingesetzt werden und dort mit neuen Daten angepasst werden. In zukünftigen Arbeiten könnten der Datenannotierungsprozess bspw. durch umfangreiches Augmentieren verbessert werden und das System direkt an eine reale Maschine angebunden werden.

Das Projekt SLEM wird im Rahmen des KI-Innovationswettbewerbs Baden-Württemberg vom Ministerium für Wirtschaft, Arbeit und Wohnungsbau Baden-Württemberg gefördert.



**IW4.0**  
Initiative Wirtschaft 4.0 BW



Baden-Württemberg

MINISTERIUM FÜR WIRTSCHAFT, ARBEIT UND WOHNUNGSBAU