

(12) NACH DEM VERTRAG ÜBER DIE INTERNATIONALE ZUSAMMENARBEIT AUF DEM GEBIET DES PATENTWESENS (PCT) VERÖFFENTLICHTE INTERNATIONALE ANMELDUNG

(19) Weltorganisation für geistiges Eigentum
Internationales Büro

(43) Internationales Veröffentlichungsdatum
06. September 2024 (06.09.2024)



(10) Internationale Veröffentlichungsnummer
WO 2024/179994 A1

(51) Internationale Patentklassifikation:
G05B 13/02 (2006.01)

(21) Internationales Aktenzeichen: PCT/EP2024/054821

(22) Internationales Anmeldedatum:
26. Februar 2024 (26.02.2024)

(25) Einreichungssprache: Deutsch

(26) Veröffentlichungssprache: Deutsch

(30) Angaben zur Priorität:
23159065.4 28. Februar 2023 (28.02.2023) EP

(71) Anmelder: SIEMENS AKTIENGESELLSCHAFT
[DE/DE]; Werner-von-Siemens-Straße 1, 80333 München (DE).

(72) Erfinder: POL, Sebastian; Grasweg 19, 90556 Seukendorf (DE). TURNER, Danielle; Waite Close, 14, Pocklington, York Yorkshire YO42 2YU (GB). GEIPEL, Markus Michael; Melanchthonstr. 38a, 81739 München (DE). THON, Ingo; Stefflweg 5, 85630 Grasbrunn (DE).

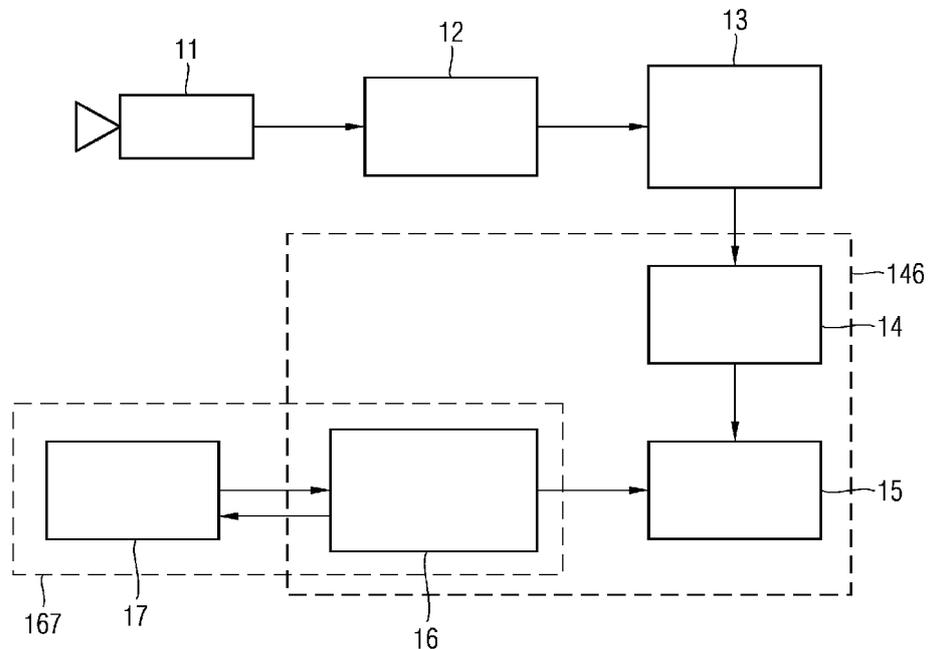
(74) Anwalt: SIEMENS PATENT ATTORNEYS; Postfach 22 16 34, 80506 München (DE).

(81) Bestimmungsstaaten (soweit nicht anders angegeben, für jede verfügbare nationale Schutzrechtsart): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MU, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT,

(54) Title: METHOD FOR OPTIMIZING MOTION CONTROL USING REINFORCEMENT LEARNING, CORRESPONDING COMPUTER PROGRAM PRODUCT, AND CORRESPONDING APPARATUS

(54) Bezeichnung: VERFAHREN ZUR OPTIMIERUNG EINER BEWEGUNGSSTEUERUNG UNTER VERWENDUNG VON VERSTÄRKENDEM LERNEN, ENTSPRECHENDES COMPUTERPROGRAMMPRODUKT UND ENTSPRECHENDE VORRICHTUNG

FIG 1



(57) Abstract: In order to find an optimum strategy for assigning a receiving device to an object on a transport apparatus, a system based substantially on reinforcement learning (RL) is proposed. While it is generally known to train the agent against the digital twin of the system, the approach in the present method also involves training the agent in a surrogate simulation, that is to say in an artificially generated environment, since the training process can therefore be considerably accelerated.

(57) Zusammenfassung: Um eine optimale Strategie für die Zuweisung eines Aufnahme gerätes zu einem Objekt auf einer Transport-

WO 2024/179994 A1

QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST,
SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ,
VC, VN, WS, ZA, ZM, ZW.

- (84) Bestimmungsstaaten** (soweit nicht anders angegeben, für jede verfügbare regionale Schutzrechtsart): ARIPO (BW, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, SD, SL, ST, SZ, TZ, UG, ZM, ZW), eurasisches (AM, AZ, BY, KG, KZ, RU, TJ, TM), europäisches (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, ME, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Veröffentlicht:

- mit internationalem Recherchenbericht (Artikel 21 Absatz 3)

vorrichtung zu finden, wird ein im Wesentlichen auf Reinforcement Learning (RL) basierendes System vorgeschlagen. Während es im Allgemeinen bekannt ist, den Agenten gegen den digitalen Zwilling des Systems zu trainieren, besteht der Ansatz in dem vorliegenden Verfahren weiterhin darin, den Agenten in einer Surrogatsimulation, also in einer künstlich generierten Umgebung, zu trainieren, da der Trainingsprozess so erheblich beschleunigt werden kann.

Beschreibung

5 VERFAHREN ZUR OPTIMIERUNG EINER BEWEGUNGSSTEUERUNG UNTER
VERWENDUNG VON VERSTÄRKENDEM LERNEN, ENTSPRECHENDES
COMPUTERPROGRAMMPRODUKT UND ENTSPRECHENDE VORRICHTUNG

In Automatisierungsanlagen gibt es vielfältige Aufgaben, welche die Aufnahme und Bewegung von Objekten beinhalten. Fertigungsanlagen, darunter beispielsweise Verpackungsanlagen, in unterschiedlichsten Branchen der Industrieautomatisierung verwenden beispielsweise Förderbänder, auf denen sich 10 Objekte, Produkte oder Teile von Produkten, in der Regel unsortiert oder unregelmäßig befinden. Beispiele finden sich in der Primär- und Sekundärverpackungsbranche, wie in der Nahrungs- und Genussmittelindustrie. Die Anzahl und die Anordnung der Produkte pro Förderfläche kann dabei stark variieren. Für einen Verpackungsvorgang werden oftmals mehrere Industrieroboter, so genannte Delta-Picker oder Knickarmroboter, in Reihe verwendet, um ein „Pick and Place“, also das 20 Aufnehmen des Produkts vom Förderband und Neupositionieren durchzuführen.

Unter einer Handhabungsmaschine, im Folgenden auch als Gerät bezeichnet, versteht man beispielsweise eine sogenannte Deltakinematik oder einen Roboter, der in der Automatisierung 25 eingesetzt wird; das beschriebene Vorgehen ist aber nicht auf diese Beispiele beschränkt. Zu den Anwendungen gehören neben den Pick-and-Place-Operationen das Verpacken (beispielsweise das Palettieren) von Produkten und vieles mehr.

30 Beim Betrieb einer Reihe von, mehr oder weniger gleichartigen, Geräten (z. B. den oben aufgeführten Robotern oder Delta-Kinematiken), die die - mehr oder weniger - gleiche Aufgabe ausführen können, ergibt sich die Herausforderung, zu entscheiden welche Maschine zu einem bestimmten Zeitpunkt je-

weils die Aufgabe ausführt. Wenn man sich zum Beispiel eine Reihe von Pick-and-Place-Maschinen vorstellt, die eine Schachtel oder Kiste mit Produkten füllen, welche auf einem Förderband antransportiert werden, so muss entschieden werden, welches Objekt von welcher Maschine kommissioniert, also von einem Förderband oder einer anderen Transportvorrichtung aufgenommen, ggf. bearbeitet und an einem anderen Platz abgelegt wird. In vielen Konfigurationen kann dabei jedes Produkt von jeder Aufnahmeeinrichtung kommissioniert werden, was zu einem sehr hohen Maß an Flexibilität für die Lösung der Aufgaben führt. Ein Gerät ist jedoch möglicherweise nicht in der Lage, rechtzeitig alle Objekte auszuwählen, die sich beispielsweise sehr nahe beieinander befinden, so dass nach Durchlauf Objekte nicht kommissioniert und beispielsweise auf dem Förderband liegen geblieben sind. Mehrere Kommissionierer werden daher üblicherweise die Kommissionierung übernehmen und diese Aufgabe verteilt lösen.

Darüber hinaus werden in der Regel unterschiedliche sogenannte Kommissionierlinien für unterschiedliche Aufgaben eingerichtet, die unterschiedliche Optimierungskriterien erfordern. Es kann in einem Anwendungsfall wichtig sein, dass alle Objekte erfolgreich ausgewählt und aufgenommen werden (z. B. beim Herausfiltern von bestimmten Materialien wie Wertstoffen oder auch Verunreinigungen aus einem Müllstrom). In anderen Anwendungsfällen ist die Aufgabe zu lösen, eine Verpackungseinheit vollständig zu befüllen (z. B. eine Schachtel Pralinen mit der richtigen Art und Anzahl von vorher festgelegten Schokoladenstücken), während es dabei weniger problematisch ist, wenn einige Einheiten auf dem Förderband am Ende nicht kommissioniert, also aufgenommen und abgelegt, werden. Ein weiteres Szenario wäre, Joghurtbecher in Paletten zu platzieren, so dass die Geschmacksrichtungen innerhalb einer Palette gleichmäßig verteilt werden.

Zusätzlich zu diesen Optimierungszielen gibt es ein weiteres Problem, dass bei aktuellen Einsatzstrategien der dadurch verursachte Verschleiß der Maschinen nicht berücksichtigt wird. Wenn die Arbeitsbelastung ungleichmäßig ist, also beispielsweise das erste Gerät an der Strecke immer zuerst berücksichtigt wird, so verschleiben die Geräte mit unterschiedlichen Geschwindigkeiten, was zu verschiedenen Wartungsintervallen, höheren Ausfallzeiten und dadurch letztendlich zu häufigerem Stillstand in der Anlage und zu erhöhten Betriebskosten führt. Darüber hinaus lässt sich auch der Verschleiß individueller Delta-Picker minimieren, indem Geschwindigkeit, Beschleunigung, und Ruck intelligent gesteuert werden. Ebenso kann eine Optimierung der Fahrwege (z. B. keine scharfen Kurven) den Verschleiß eines Delta-Pickers reduzieren.

Aus der Druckschrift EP 3456485A1 ist bereits ein Verfahren zum Optimieren eines automatisierten Vorgangs zum Auswählen und Greifen eines Objekts durch einen Roboter bekannt. Es wird ein Optimierungskriterium festgelegt und bei der Ermittlung eines Prioritätskenners berücksichtigt. Dabei geht die Kapazität der verschiedenen Roboter der Anordnung mit in die Berechnung ein. Der Prioritätskenner regelt die Priorität derjenigen Objekte, welche gemeinsam einem Roboter zugeordnet werden.

In US 2018/0089589A1 wird eine Machine Learning Vorrichtung beschrieben, die den Betriebszustand eines Roboters lernt. Der Roboter greift eine Vielzahl von Objekten, die auf einer Trägervorrichtung angeordnet sind, und legt sie in einem Behälter ab. Ziel ist die Reduzierung der Auslastung des Roboters und die Minimierung der Zeit für die Ablage.

US 2023/0040623A1 offenbart ein Reinforcement Learning System für ein Pick-and-Place System, wobei das Training Entscheidungen für eine Echtzeit-Umgebung erzeugt werden, indem ein Bewegungspfad ermittelt wird, der verschiedene Hardware- und
5 Echtzeitanforderungen erfüllt.

Während bekannte Zuweisungsstrategien die Hauptaufgabe in den meisten Szenarien erfüllen können, berücksichtigen sie nicht die oben erwähnten Herausforderungen. Das liegt daran, dass
10 die bekannten Zuweisungsstrategien eine Koordination zwischen mehreren Maschinen, welche die gleiche Aufgabe erfüllen sollen oder können, bislang nicht berücksichtigen. Stattdessen ist jede Maschine (Kommissionierer) mit einer individuellen Richtlinie ausgestattet, die entscheidet, welche Objekte in
15 seiner Arbeitszone ausgewählt und gegriffen werden sollen.

Gesucht ist also beispielsweise

- eine Zuordnung der Objekte zu den Zielpositionen, sowie
- eine Abfolge von Transport-, Leerlauf- sowie Pick-and-
20 Place-Operationen der Kommissionierer.

In der Praxis treten hier unterschiedliche Konfigurationen auf:

- Die Geräte können sich ständig oder abrupt bewegen.
- 25 - Die Kommissioniervorgänge können von beweglichen oder temporären Stoppbändern aus durchgeführt werden.
- Die Anordnung der Transportvorrichtungen können parallel (in die gleiche Richtung oder entgegengesetzt laufend) oder orthogonal zueinander sein.
- 30 - Die Geräte können eine oder auch mehrere Komponenten gleichzeitig aufnehmen.
- Die Zielpositionen können eines oder mehrere Objekte aufnehmen.

Auch hinsichtlich der logistischen Randbedingungen können unterschiedliche Anforderungen bestehen:

- Es kann Einschränkungen geben, welches Objekt in welcher Zielposition platziert wird, oder es kann Bedingungen für
- 5 Klassen auf Objekten und Klassen von Zielpositionen geben, z. B. ein grünes Objekt in jedem Feld.

Bekannte Strategien für die Bestimmung der nächsten Aktion, und damit im oben zitierten Beispiel die Auswahl des nächsten

10 Gerätes zur Entnahme eines Objekts vom Förderband sind beispielsweise:

- Auswahl des Objekts, welches sich räumlich dem Maschinenarm am nächsten befindet,
 - Auswahl des Objekts, das am weitesten vorne oder am weitesten
- 15 hinten auf dem Förderband liegt.

Diese individuellen Zuweisungs-Strategien führen jedoch häufig zu dem Problem, dass die die Arbeitsbelastung individuell und damit ungleichmäßig auf die Maschinen in einer Arbeits-

20 Linie verteilt werden und damit beispielsweise die Maschinen am Beginn eines Förderbands eine viel höhere Arbeitsbelastung haben als die Maschinen am Ende des Förderbands.

Keine der oben beschriebenen Methoden erlaubt es dem Benutzer

25 jedoch, verschiedene Optimierungskriterien gegeneinander abzuwägen, um so zu einem verbesserten Steuerungs- und Auswahlverfahren zu gelangen.

30 Es ist daher eine Aufgabe der Erfindung, ein verbessertes Verfahren anzugeben, welches die oben genannten Nachteile überwindet. Es ist weiterhin eine Aufgabe der Erfindung, eine Vorrichtung anzugeben.

Diese Aufgabe wird gelöst durch ein Verfahren aufweisend die Merkmale des unabhängigen Patentanspruch 1 und ein Computerprogrammprodukt mit den Merkmalen des Patentanspruchs 9.

Weiterhin wird die Aufgabe gelöst durch eine Vorrichtung welche die Merkmale des unabhängigen Patentanspruchs 10 aufweist.

Weitere Ausführungsformen der Erfindung werden durch die Unteransprüche abgedeckt.

10

Das Verfahren zur Optimierung einer Bewegungssteuerung eines Gerätes unter Verwendung von verstärkendem Lernen weist eine Trainingsphase auf, welche unter Verwendung von generierten Trainingsdaten erfolgt, mit einem ersten Optimierungsziel und mindestens einem zweiten, vom ersten Optimierungsziel verschiedenen Optimierungsziel umfasst mit jeweils einer Reward Funktion, und wobei die Trainingsphase einem Trainingsanteil auf dem ersten Optimierungsziel und einen Trainingsanteil auf das zumindest zweite Optimierungsziel richtet, mittels einer gewichteten Summe der Reward Funktionen des jeweiligen Optimierungsziels.

Genauer gesagt kann für jedes gewünschte Optimierungsziel eine Reward-Funktion aufgestellt werden. Mehrere Optimierungsziele können dann einem RL Agenten während des Trainings über eine nach Priorität gewichtete Summe der einzelnen Rewards zugeführt werden.

Für das Training nur eines Optimierungsziels ist dann die Gewichtung für das oder die anderen Optimierungsziele 0. Weitere Ausgestaltungen sind in den Unteransprüchen festgehalten.

Die Erfindung wird im Folgenden beispielhaft durch die Figuren dargestellt, dabei zeigen

Figur 1 ein Architekturkonzept
Figur 2 einen zeitlichen Ablauf des Kontrollflusses, unterteilt in zwei Phasen und
Figur 3 eine beispielhafte Simulation einer Pick&Place Situation mit Aufnahmegeräten in Draufsicht.

Um eine optimale Strategie für die Zuweisung eines Aufnahme-
gerätes zu einem Objekt auf einer Transportvorrichtung zu
finden, wird ein im Wesentlichen auf Reinforcement Learning
10 (RL) basierendes System vorgeschlagen.

Reinforcement Learning RL, auch verstärkendes Lernen genannt,
steht für eine Reihe von Methoden des sogenannten maschinellen
Lernens, bei denen ein Agent selbstständig eine Strategie
15 (policy) erlernt, um erhaltene Belohnungen (rewards) zu maximieren. Dabei wird dem Agenten nicht vorgezeigt, welche Aktion in welcher Situation die beste ist, sondern er erhält durch die Interaktion mit seiner Umwelt zu bestimmten Zeitpunkten eine Belohnung, welche anzeigt, wie gut er eine Aufgabe gelöst hat. Diese Belohnung kann auch negativ sein.
Die mathematischen Grundlagen des verstärkenden Lernens bilden die folgenden Begriffe: Der Agent (agent), die Umwelt (environment), die Zustände (states), die Aktionen (actions) und die Belohnungen (rewards). Die Methoden des bestärkenden
25 Lernens betrachten die Interaktion des lernenden Agenten mit seiner Umwelt. Die Umwelt besteht aus einer Menge von Zuständen und einer Menge von Aktionen, sowie einer Dynamik und einer Startverteilung. Die Interaktion des Agenten mit der Umwelt findet zu diskreten Zeitpunkten statt. Zu jedem Zeitpunkt befindet sich der Agent in einem Zustand, wählt eine
30 Aktion aus und erhält dafür eine Belohnung, wie oben beschrieben.

Die verwendeten Komponenten für das Verfahren sind (neben anderen) in Figur 1 dargestellt. Die außerhalb der gestrichelten Boxen 146, 167 liegenden Komponenten Kamera 11, Objekterkennung 12, und PLC 13 sind dabei für das ausführende System
5 erforderliche Systemkomponenten, aber kein direkter Bestandteil der beschriebenen Erfindung. Die Komponenten, die sich innerhalb der gestrichelten Boxen 146, 167 befinden, werden im Folgenden näher beschrieben.

Die Box 167 beschreibt schematisch die wesentlichen Komponenten des Reinforcement Learning RL-Agent 16, welcher das Verfahren der Bewegungssteuerung in einer Trainingsphase auf Basis von Trainingsdaten 17 erlernt, um dann in einer Produktivphase auf die Steuerung 13 einzuwirken. Die Konstellation 146 wie in Figur 1 entspricht dabei einer Realisierung auf
10 einem Edge Gerät, mittels Proxy Server 14 und Rest API 15. Andere Konstellationen sind jedoch ebenfalls denkbar.

Figur 2 zeigt einen Kontroll-Fluss des vorgeschlagenen Verfahrens basierend auf der in Figur 1 vorgestellten Architektur. Die am Verfahren beteiligten Komponenten sind, soweit
20 möglich, mit den gleichen Bezugszeichen versehen wie in Figur 1.

In der ersten Phase, der Trainingsphase P1 des Verfahrens wird der RL-Agent in einer sogenannten Surrogatsimulation
25 trainiert. Die Surrogatsimulation zeichnet sich dadurch aus, dass sie signifikant schneller als Echtzeit ablauffähig ist und die zu simulierende Umgebung hinreichend akkurat abbildet. Dies geschieht beispielsweise durch eine gegenüber der Realität vereinfachte Darstellung der Steuerung des Gerätes
30 in der Simulation. Während des Trainings interagiert der RL-Agent kontinuierlich mit der Umgebung, d. h. jede ausgeführte Aktion 256 wird anhand des zu erreichenden Ergebnisses bewertet, mit einer Belohnung (positiv oder negativ) 265, und führt zur Verstärkung des Verhaltens (bei positiver Bewer-

tung) oder zur Änderung des Verhaltens (bei negativer Bewertung).

In dem vorher angeführten Beispiel der Bewegungssteuerung für
5 ein Aufnahmegesetz erhält der RL-Agent für das Training jeweils eine Beobachtung aus der Umgebung und führt eine Aktion durch, beispielsweise die Zuordnung eines neuen Gegenstands, welche sich auf einer Beförderungsvorrichtung befindet, zu einer Maschine, welche den Gegenstand aufnehmen soll. Diese
10 Trainingsschleife kann beliebig oft durchlaufen werden. In Figur 3 werden die in dem Beispiel verwendeten Szenarien weiter unten detaillierter dargestellt.

Der Beobachtungsraum des RL-Agenten - also die Umgebungseigenschaften, die dem RL-Agenten zur Entscheidungsfindung zur
15 Verfügung stehen - ist dabei so konzipiert, dass er alle relevanten Informationen enthält, die für eine optimale Entscheidungsfindung für die Auswahl des passenden Geräts erforderlich sind.

20 In dem genannten Beispiel umfasst das beispielsweise:

- Eine Liste der Koordinaten jedes Gegenstands auf der Transportvorrichtung (dem Förderband) oder den Transportvorrichtungen (Transport-Roboter, AGV...),
- Bereits bestehende Zuordnungen von weiteren aufzunehmenden
25 Gegenständen zu jeweils einer der Aufnahmegeräte,
- Eine Liste der Koordinaten der Ziel- /Platzierungspositionen, und/oder
- Eine Variable für jede Aufnahmegeräte, die den jeweiligen Verschleiß der Vorrichtung angibt.

30

In einer zweiten Phase P2 in Figur 2 wird die Produktiv- oder Online Phase des Verfahrens in Anwendung auf das genannte Beispiel beschrieben, also wenn das System in einer Anlage integriert zur Anwendung kommt. Eine Kamera mit Objekterken-

nung 1112 erkennt einen neuen Gegenstand inklusive der Ortsinformation zu diesem Gegenstand, und liefert diese Informationen 212 an die Steuerung 13. Die Steuerung 13 serialisiert ein Produktregister (d. h. es erfolgt eine Konvertierung der Daten in einen Bytestream, um das Objekt zu speichern oder in den Arbeitsspeicher, eine Datenbank oder eine Datei zu übertragen.

Hauptzweck ist, den Zustand des Objekts zu speichern, um es bei Bedarf neu erstellen zu können) und sendet es 223 (beispielsweise mittels Übertragungsprotokoll TCP) an einen Proxy-Server 14. Das Produktregister stellt einen Speicherbaustein dar, welcher den Zustand einer Kommissionierlinie beschreibt und insbesondere die Ortsangaben der erkannten Produkte auf der Transportvorrichtung, sowie ggf. weitere Informationen, wie die Orientierung, Größe, Gewicht, oder den gewünschte Ablageort enthält. In dem besprochenen Beispiel bildet ein Ausschnitt des Produktregisters den Beobachtungsraum des RL-Agenten.

Die Daten werden in dem dargestellten Beispiel in einer Edge Konfiguration über den Proxy 14 und die Rest API Schnittstelle verarbeitet:

Produkt Register (von der Steuerung) 223
Beobachtungsstatus 234
Abfrage an den Agenten (Forward request), 245
Antwort des Agenten (Forward response), 254
Zuordnung Objekt (product assignment), 243, 232 an die Steuerung, von der die Anfrage getriggert war.

In Figur 3 sind die vorbereiteten Trainingsdaten der Surrogatsimulation dargestellt, inklusive der daraus berechneten optimalen Fahrwege für die Bewegungssteuerung. Verschiedene Varianzen 31 der Lage (Anordnung, Häufigkeit) von Objekten 331, 332, ... auf einer Transportvorrichtung 33 werden simu-

liert und als Basis für das Training eines RL Agenten verwendet. Dabei befindet sich beispielsweise auf einem Förderband 33 eine Anzahl an Objekten 331, 332, ... 33x, welche durch eine Aufnahmevorrichtung (in der Figur nicht dargestellt) aufgenommen und jeweils auf einem Zielplatz 321, 322, ... 32n beispielsweise auf einem weiteren Förderband abgelegt werden sollen. Dargestellt ist die jeweilige Bewegungsspur der Aufnahmevorrichtungen. In dem vorliegenden Beispiel sind das drei Aufnahmevorrichtungen mit den jeweiligen Bewegungsspuren 341, 342, 343. Eine Spur stellt die nacheinander folgende Aufnahme von Gegenständen von der Transportvorrichtung, also beispielsweise wird der Gegenstand 333 von der Aufnahmevorrichtung G1, G2 auf den Platz 322 bewegt, die Spur wechselt von Aufnahme zu Ablage und dann wieder zurück zu Aufnahme. Die Bewegungsrichtung des Förderbands ist in der Darstellung von rechts nach links, 35. Vorstellbar sind z. B Lebensmittel, wie Schokoladentafeln oder Pralinen, welche zur Verpackung in eine passende Schachtel abgelegt werden sollen.

Bei der Ablage sind ebenfalls einige Randbedingungen zu beachten, beispielsweise soll auf jedem Platz nur genau ein Objekt abgelegt werden. Es kann auch vorkommen, dass die Objekte nicht alle gleichartig sind, und dass von jeder Art nur eine vorgegebene Anzahl in ggf. auch vordefinierten Aufnahmeplätzen abgelegt werden soll. Dabei ist auch zu beachten, dass die Aufnahmeplätze am Ende des Vorgangs alle belegt sein sollten.

Befindet sich nun eine sehr große Anzahl von (ggf. auch verschiedenartigen) Objekten auf dem Förderband, kann der entstehende Beobachtungsraum sehr groß werden, was die Lernkomplexität im Training negativ beeinflusst. In solchen Fällen ist es sinnvoll, für die Trainingsphase nur einen Teil aller

Objekte, z. B. die neuesten n Objekte, auf dem Förderband zu beachten.

Ein geeigneter Wert für n kann über eine Hyperparameteroptimierung gefunden werden. Ein Hyperparameter ist ein Parameter, der zur Steuerung des Trainingsalgorithmus verwendet wird und dessen Wert im Gegensatz zu anderen Parametern vor dem eigentlichen Training des Modells festgelegt werden muss. Es gibt dabei verschiedene Vorgehensweisen, wie die Rastersuche, die Zufallssuche, die Bayessche Optimierung, Gradientenbasierte Optimierung oder die Evolutionäre Optimierung.

Ein RL-Agent lernt, das von der Umgebung gegebene Belohnungssignal zu maximieren, daher muss die Belohnungsfunktion mit dem Optimierungsziel übereinstimmen. In einer weiteren Ausgestaltung der Erfindung kann die Belohnung eine gewichtete Summe von Belohnungen für das Erreichen der verschiedenen Optimierungsziele sein.

Dies ermöglicht eine sehr feingranulare Verhaltensabstimmung durch den Benutzer. Um beispielsweise eine Richtlinie zu erlernen, in der alle Objekte erfolgreich kommissioniert, d. h. aufgenommen und an einem gewünschten Zielort abgelegt werden, und die Arbeitsbelastung jedes Kommissionierers dabei gleichmäßig gehalten wird, könnte eine geeignete Belohnungsfunktion sein:

- Der Agent erhält eine große negative Belohnung, sobald ein Objekt am Ende des Trägerbandes herunterfällt, weil es nicht ausgewählt bzw. aufgenommen wurde.
- Der Agent erhält eine negative Belohnung, die der Varianz der Arbeitsbelastung jedes Kommissionierers entspricht, d. h. abhängig davon, wie gleichmäßig die Arbeitslast auf die einzelnen Aufnahmeverrichtungen verteilt wird.

Während es im Allgemeinen bekannt ist, den Agenten gegen den digitalen Zwilling des Systems zu trainieren, ist der Ansatz

in dem vorliegenden Verfahren, den Agenten in einer Surrogatsimulation, also in einer künstlich generierten Umgebung, zu trainieren, da der Trainingsprozess so erheblich beschleunigt werden kann. Dies ist vorteilhaft, da der Rechenaufwand
5 in der digitalen Zwillingsimulation (Physik, 3D-Rendering etc.) erheblich ist, aber für das beschriebene Lernproblem völlig irrelevant. Daher kann dieser Rechenaufwand in einer Surrogatsimulation leicht weggelassen werden.

10 Darüber hinaus ermöglicht in einer vorteilhaften Ausgestaltung ein Konfigurator die automatische Erstellung einer solchen Surrogatsimulation. In diesem Konfigurator kann der Benutzer eine physische Beschreibung einer Kommissionierlinie (Anzahl der Kommissionierer, Position der Kommissionierer,
15 Förderbandgeschwindigkeit usw.) importieren oder diese erforderlichen Informationen manuell eingeben. Optional enthält der Konfigurator Schieberegler, in denen der Anwender festlegen kann, wie wichtig verschiedene Optimierungskriterien sind. Je nach Konfiguration dieser Schieberegler kann automatisch
20 eine passende Belohnungsfunktion deaktiviert werden.

Darüber hinaus ist es beispielsweise möglich, eine sogenannte Slider-Konfiguration in den Beobachtungsraum des RL-Agenten zu integrieren. Das Konfrontieren des Agenten anhand ver-
25 schiedener Kombinationen von Optimierungskriterien während der Trainingsphase ermöglicht es dem Benutzer, das Verhalten des Agenten so, während des Einsatzes, mit sofortiger Wirkung anzupassen, ohne dass ein erneutes Training des Agenten erforderlich ist.

30

Das Training des RL-Agenten erzeugt ein Regelwerk in Form eines parametrisierten Modells, z. B. eines neuronalen Netzes. Der untere Teil P2 des Sequenzdiagramms in Figur 2 zeigt, wie

dieses erlernte Regelwerk in ein tatsächliches System zur Online-Entscheidungsfindung integriert wird. Sobald ein neues Objekt von der Kamera 1121 erkannt wird, serialisiert die Steuerung 13 das Produktregister und sendet es, 223 (beispielsweise über TCP) an einen Proxy-Server 14. Das Produktregister enthält die Produktpositionen (vergleiche Figur 3, 321, 322, 32n auf dem Förderband 33) sowie die anderen Informationen, die den Status des Agenten ausmachen. Der Proxy-Server deserialisiert die empfangenen Daten und extrahiert die Informationen, die die neuronale Netzwerkrichtlinie als Eingabe erwartet. Die neuronale Netzwerkrichtlinie selbst ist als Microservice verpackt, der eine REST-API bietet.

<https://de.wikipedia.org/wiki/Microservices>

[https://de.wikipedia.org/wiki/Representational State Transfer](https://de.wikipedia.org/wiki/Representational_State_Transfer)

Der Proxy-Server fordert eine Zuweisung für das neue Produkt von der neuronalen Netzwerkrichtlinie an und gibt die Zuweisung an die SPS zurück.

Der oben erwähnte Proxy-Server 14 sowie der Agenten-Microservice, der eine Programmierungsschnittstelle bietet, beispielsweise als REST-API 15 ausgestaltet, werden vorteilhafterweise auf einem externen Rechner, beispielsweise einem Industrie-PC oder einer Edge-Box ausgeführt. Denkbar ist aber auch, das Agentenverhalten direkt in die SPS 13 selbst zu integrieren. Dass der Agent als Microservice implementiert ist, bzw. auf einem IPC oder in der SPS laufen kann ist lediglich eine beispielhafte Implementierung.

Der beschriebene Ansatz ermöglicht es, die Auslastung von Aufnahmevorrichtungen unter Berücksichtigung verschiedener Optimierungsziele zu verbessern. Während sich Lösungen derzeit nur auf die Erreichung der Kernaufgabe einer Kommissionierlinie konzentrieren, erlaubt die vorgeschlagene Lösung

die Berücksichtigung zusätzlicher Kriterien wie den gleichmäßigen Verschleiß aller Maschinen einer Linie / an einem Förderband. Dies hat beispielsweise den großen Vorteil, dass alle Maschinen ein einziges Wartungsintervall haben, das planbarer ist, Stillstandszeiten reduziert und Kosten senkt.

Bezugszeichenliste:

	11	Kamera
	12	Objekterkennung
5	1112	Kamera / Objekterkennung
	13	Steuerung, PLC, SPS
	14	Proxy Server
	15	Programmierschnittstelle, Rest API
	16	Reinforcement Learning Agent
10	17	Simulation
	146	Bereitstellung auf Edge Geräten
	167	Training (High Performance Computer)
	212	Objekt Lokalisierung
15	223	Objekt (Produkt?) Registrierung
	232, 243	Objekt (Produkt?) Zuordnung
	234	Beobachtungsstatus
	245	Weiterleiten Anfrage (forward request)
	254	Weiterleiten Antwort (forward response)
20		
	31	Aufsicht schematisch, Trainingsdaten
	32, 33	Förderband
	331, 33x	Objekt
	321, 32n	Platz zur Aufnahme von Objekten
25	341, 342	Fahrstrecke von Maschine zur Aufnahme und Ablage
	G1, G2	Gerät (Roboter, Maschine)
	P1	Trainingsphase
	P2	Produktivphase

Patentansprüche

1. Verfahren zur Optimierung einer Bewegungssteuerung eines Gerätes (G1, G2), unter Verwendung von verstärkendem Lernen, wobei das verstärkende Lernen eine Trainingsphase (P1) umfasst, welche in einer Surrogatsimulation trainiert, unter Verwendung von generierten Trainingsdaten (17) erfolgt, die eine vereinfachte Darstellung der Steuerung des Gerätes darstellen,
- 5
- 10 wobei die Trainingsphase (P1) des verstärkenden Lernens ein erstes Optimierungsziel mit einer ersten Rewardfunktion umfasst und
- wobei die Trainingsphase (P1) des verstärkenden Lernens mindestens ein zweites, vom ersten Optimierungsziel verschiedenes Optimierungsziel mit einer zweiten Rewardfunktion umfasst,
- 15
- wobei die Trainingsphase (P1) einem Trainingsanteil auf dem ersten Optimierungsziel und einen Trainingsanteil auf das zumindest zweite Optimierungsziel richtet, mittels einer gewichteten Summe der Rewardfunktionen der jeweiligen Optimierungsziele und
- 20
- im Training der Bewegungssteuerung das erste Optimierungsziel des Gerätes (G1) in Abhängigkeit von dem Optimierungsziel eines zweiten Geräts (G2) erfolgt.
- 25
2. Verfahren gemäß Patentanspruch 1, dadurch gekennzeichnet, dass
- das Gerät (G1) ein Aufnahmegerät ist, welches dazu geeignet ist, eine Pick & Place Aktion eines Objektes (331, 332, ...) durchzuführen.
- 30
3. Verfahren gemäß Patentanspruch 2, dadurch gekennzeichnet, dass

das Objekt (331, 332, ...) durch das Gerät (G1, G2) von einer Transportvorrichtung (33) aufgenommen und auf einer Aufnahmevorrichtung (32) abgelegt wird.

- 5 4. Verfahren gemäß einem der vorherigen Patentansprüche, dadurch gekennzeichnet, dass die Bewegungssteuerung für das Gerät (G1, G2) folgende Verfahrensschritte umfasst:
- Auswahl des Objekts (331, 332, ...) auf der Transportvorrichtung (33),
 - Auswahl des Geräts aus einer Menge von Geräten (G1, G2),
 - Aufnahme des ausgewählten Objekts (331, 332, ...) durch das ausgewählte Gerät (G1, G2),
 - Auswahl eines Ablageplatzes (321, 322) für das Objekt, und
 - 15 - Ablage des Objekts (331, 332, ...) durch das Gerät an dem ausgewählten Ablageort (321, 322, ...).

5. Verfahren gemäß einem der vorherigen Patentansprüche, dadurch gekennzeichnet, dass
- 20 das erste Optimierungsziel und das zweite Optimierungsziel jeweils einen der folgenden Aspekte betrifft:
- Transportzeit des Objekts (331, 332, ...) bis zur Ablage,
 - Fahrweg des Geräts (341, 342, ...),
 - Aufnahmequote der zutransportierenden Objekte,
 - 25 - Verschleiß des Geräts (G1, G2),
 - Korrekte Zuordnung des Objekts (331, 332, ...) zum Ablageort (321, 322, ...),
 - Energieverbrauch des Geräts (G1, G2),
 - Durchsatz.

30

6. Verfahren gemäß einem der vorherigen Patentansprüche, dadurch gekennzeichnet, dass die Generierung der Trainingsdaten konfigurierbar ist und

automatisiert erfolgt.

7. Verfahren gemäß einem der vorherigen Patentansprüche, dadurch gekennzeichnet, dass

5 zur Vereinfachung einer Simulation während der Trainingsphase nur eine Anzahl n Testfälle oder Objekte (331, 332, ...), insbesondere die letzten n Objekte, berücksichtigt werden.

8. Verfahren gemäß Patentanspruch 7,

10 dadurch gekennzeichnet, dass

die Anzahl n mittels Hyperparameteroptimierung ermittelt wird.

9. Computerprogrammprodukt zur Optimierung einer Bewegungs-

15 steuerung unter Verwendung von verstärkendem Lernen zur Durchführung des Verfahrens gemäß einem der Patentansprüche 1 bis 8.

10. Vorrichtung zur Optimierung einer Bewegungssteuerung ei-

20 nes Gerätes ($G1$, $G2$) unter Verwendung von verstärkendem Lernen (146)

umfassend einen Agenten (16) geeignet für die Durchführung eines Trainings mittels verstärkendem Lernen, wobei eine

Trainingsphase ($P1$) als Surrogatsimulation unter Verwendung

25 von generierten Trainingsdaten (17) erfolgt, wobei die generierten Trainingsdaten eine vereinfachte Darstellung der Steuerung des Gerätes simulieren

wobei die Trainingsphase ($P1$) des verstärkenden Lernens ein erstes Optimierungsziel mit einer ersten Rewardfunktion um-

30 fasst und

wobei die Trainingsphase ($P1$) des verstärkenden Lernens mindestens ein zweites, vom ersten Optimierungsziel verschiede-

nes Optimierungsziel mit einer zweiten Rewardfunktion umfasst und

wobei die Trainingsphase (P1) einem Trainingsanteil auf dem ersten Optimierungsziel und einen Trainingsanteil auf das zumindest zweite Optimierungsziel richtet, mittels einer gewichteten Summe der Reward Funktionen der jeweiligen Optimierungsziele,
5 eine Schnittstelle (14, 15) zur Ausgabe der von Steuersignalen an das Gerät (G1) und
der Agent (16) zur Optimierung der Bewegungssteuerung auf das Optimierungsziel des Gerätes (G1) in Abhängigkeit von dem Optimierung
10 ziel eines zweiten Geräts (G2) trainiert ist.

11. Vorrichtung gemäß Patentanspruch 10,
dadurch gekennzeichnet, dass
das Gerät (G1, G2) ein Aufnahmegerät ist, welches dazu geeignet
15 net ist, eine Pick & Place Aktion eines Objektes (331, 332, ...)
durchzuführen.

12. Vorrichtung gemäß Patentanspruch 10,
dadurch gekennzeichnet, dass
20 das Objekt (331, 332, ...) durch das Gerät (G1, G2) von einer Transportvorrichtung (33) aufgenommen, und auf einer Aufnahmeforrichtung (32) abgelegt wird, wobei die Bewegungssteuerung für das Gerät (G1, G2) dazu ausgelegt ist, folgende Verfahrensschritte durchzuführen:

- 25 - Auswahl des Objekts (331, 332, ...) auf der Transportvorrichtung (33),
- Auswahl des Geräts aus einer Menge von Geräten (G1, G2),
- Aufnahme des ausgewählten Objekts (331, 332, ...) durch das ausgewählte Gerät (G1, G2),
30 - Auswahl eines Ablageplatzes (321, 322) für das Objekt, und
- Ablage des Objekts (331, 332, ...) durch das Gerät an dem ausgewählten Ablageort (321, 322, ...).

13. Vorrichtung gemäß einem der vorherigen Patentansprüche 10 bis 12,
dadurch gekennzeichnet, dass
das erste Optimierungsziel und das zweite Optimierungsziel
5 jeweils einen der folgenden Aspekte betrifft:
- Transportzeit des Objekts (331, 332, ...) bis zur Ablage,
- Fahrweg des Geräts (341, 342,...),
- Aufnahmequote der zutransportierenden Objekte,
- Verschleiß des Geräts (G1, G2),
10 - Korrekte Zuordnung des Objekts (331, 332, ...) zum Ablageort
(321, 322, ...),
- Energieverbrauch des Geräts (G1, G2),
- Durchsatz.

15 14. Vorrichtung gemäß einem der vorherigen Patentansprüche 10 bis 13,
dadurch gekennzeichnet, dass
die Generierung der Trainingsdaten konfigurierbar, automatisiert erfolgt.

20 15. Vorrichtung gemäß einem der vorherigen Patentansprüche 10 bis 14,
dadurch gekennzeichnet, dass
während der Trainingsphase nur eine Anzahl n Testfälle oder
25 Objekte (331, 332, ...), insbesondere die letzten n Objekte,
berücksichtigt werden.

16. Vorrichtung gemäß einem der vorherigen Patentansprüche 10 bis 15,
30 dadurch gekennzeichnet, dass
die Anzahl n mittels Hyperparameteroptimierung gefunden wird.

17. Vorrichtung gemäß einem der vorherigen Patentansprüche 10 bis 16,

dadurch gekennzeichnet, dass
die Realisierung auf einem Edge Gerät (146), mittels Proxy
Server (14) und Rest API (15) erfolgt.

FIG 1

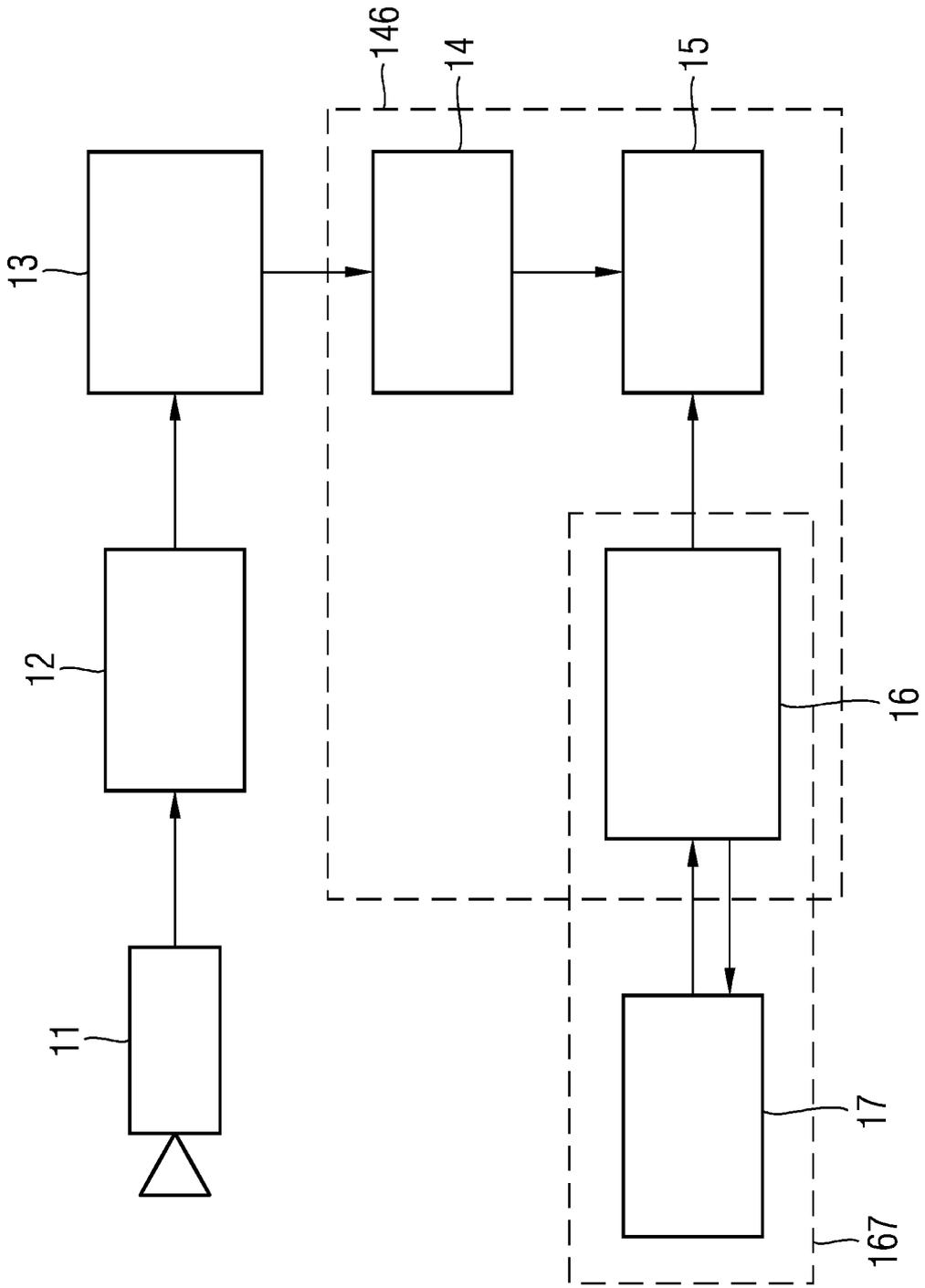


FIG 2

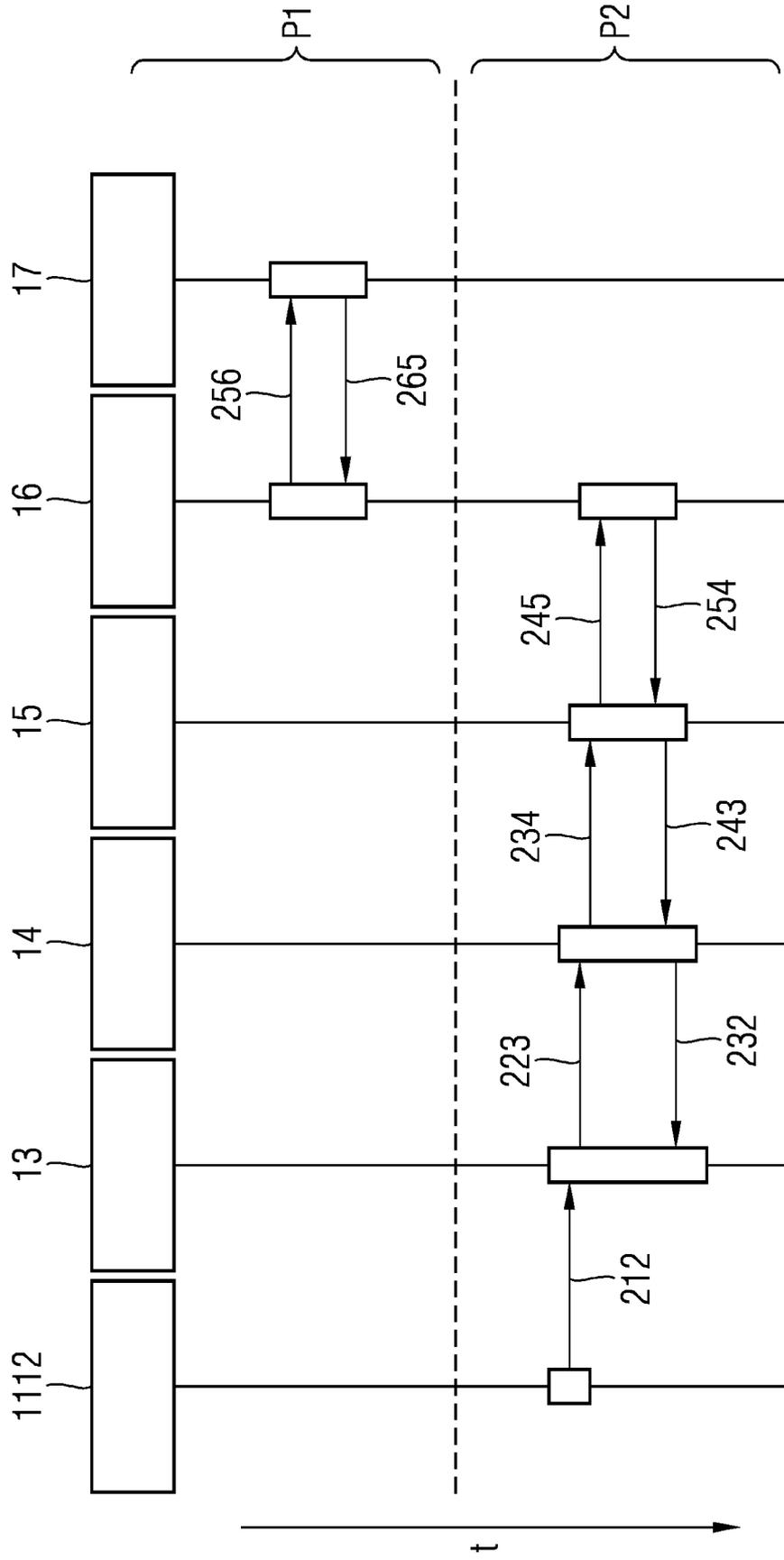
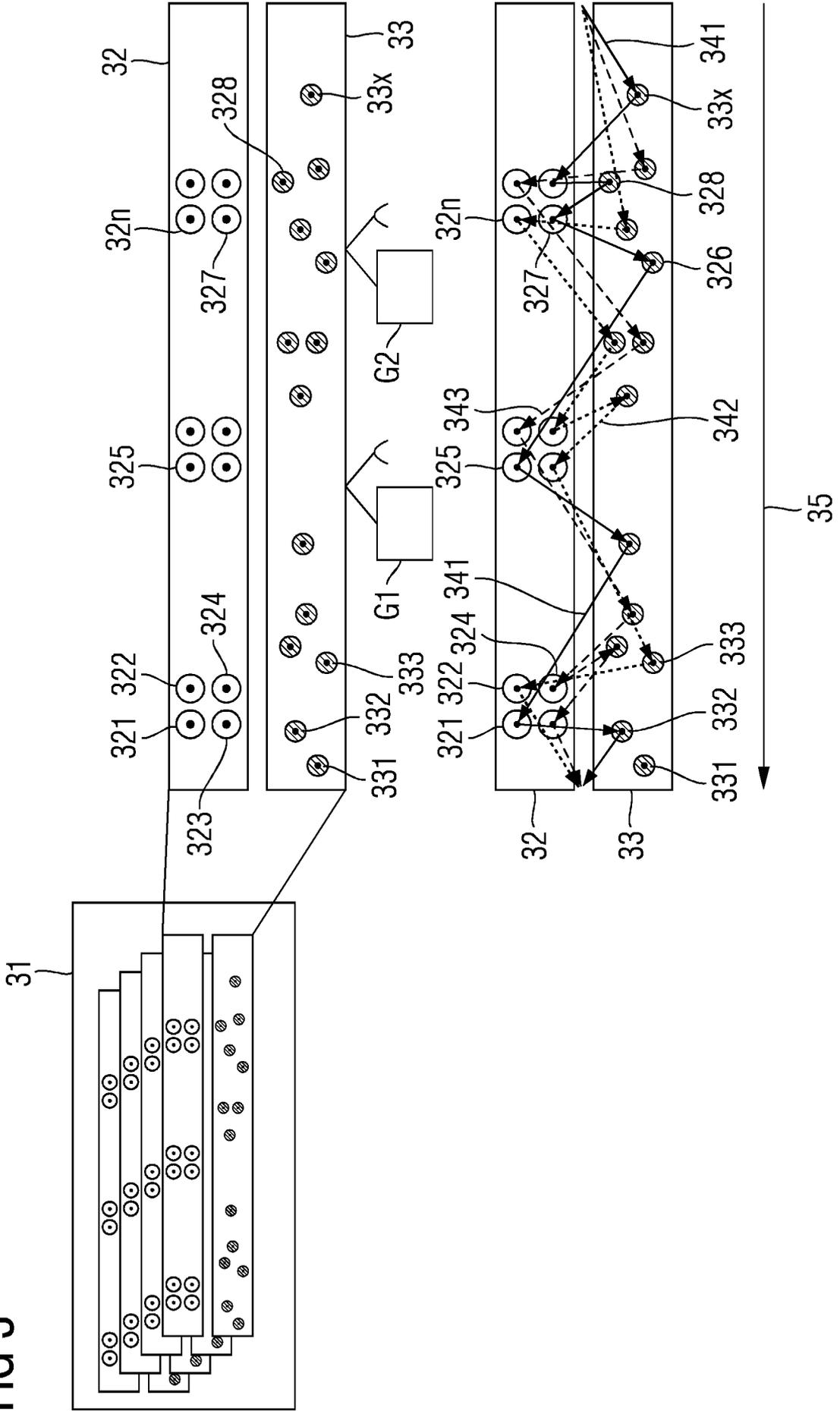


FIG 3



INTERNATIONAL SEARCH REPORT

International application No.

PCT/EP2024/054821

A. CLASSIFICATION OF SUBJECT MATTER <i>G05B 13/02</i> (2006.01) According to International Patent Classification (IPC) or to both national classification and IPC		
B. FIELDS SEARCHED Minimum documentation searched (classification system followed by classification symbols) G05B; B25J; G06N Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) EPO-Internal, WPI Data		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2018089589 A1 (OOBA MASAFUMI [JP]) 29 March 2018 (2018-03-29) cited in the application the whole document	1-17
X	US 2023040623 A1 (LE PHAM-TUYEN [KR] ET AL) 09 February 2023 (2023-02-09) cited in the application the whole document	1-17
A	LI PEIJIN ET AL. "A Q-learning Control Method for a Soft Robotic Arm Utilizing Training Data from a Rough Simulator" <i>2021 IEEE INTERNATIONAL CONFERENCE ON ROBOTICS AND BIOMIMETICS (ROBIO), IEEE, 27 December 2021 (2021-12-27), pages 839-845</i> DOI: 10.1109/ROBIO54168.2021.9739524 XP034106495 * I. Introduction ** II. Methods *	1, 6-10, 14-16
A	US 2005075752 A1 (BAN KAZUNORI [JP] ET AL) 07 April 2005 (2005-04-07) paragraph [0004] - paragraph [0010] paragraph [0043] - paragraph [0046] figure 6	1-5,9,10,12,13
<input checked="" type="checkbox"/> Further documents are listed in the continuation of Box C. <input checked="" type="checkbox"/> See patent family annex.		
<p>* Special categories of cited documents:</p> <p>“A” document defining the general state of the art which is not considered to be of particular relevance</p> <p>“E” earlier application or patent but published on or after the international filing date</p> <p>“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)</p> <p>“O” document referring to an oral disclosure, use, exhibition or other means</p> <p>“P” document published prior to the international filing date but later than the priority date claimed</p> <p>“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention</p> <p>“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone</p> <p>“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art</p> <p>“&” document member of the same patent family</p>		
Date of the actual completion of the international search 31 May 2024		Date of mailing of the international search report 19 June 2024
Name and mailing address of the ISA/EP European Patent Office p.b. 5818, Patentlaan 2, 2280 HV Rijswijk Netherlands (Kingdom of the) Telephone No. (+31-70)340-2040 Facsimile No. (+31-70)340-3016		Authorized officer Gardella, Simone Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/EP2024/054821

C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 2021276188 A1 (TANG HAORAN [US] ET AL) 09 September 2021 (2021-09-09) the whole document	1, 5, 9, 10, 13
A	RYO MATSUMURA ET AL. "Learning Based Industrial Bin-picking Trained with Approximate Physics Simulator" <i>ARXIV.ORG, CORNELL UNIVERSITY LIBRARY, 201 OLIN LIBRARY CORNELL UNIVERSITY ITHACA, NY 14853, 23 May 2018 (2018-05-23), XP080881337</i> * I. Introduction *	1, 9, 10

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. Claims Nos.: **1-8 (in part)**
because they relate to subject matter not required to be searched by this Authority, namely:
Method for performing purely mental acts - see Rule 39.1 (iii)

2. Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/EP2024/054821

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
US	2018089589	A1	29 March 2018	CN	107866809	A	03 April 2018
				DE	102017008836	A1	29 March 2018
				JP	6514171	B2	15 May 2019
				JP	2018051664	A	05 April 2018
				US	2018089589	A1	29 March 2018

US	2023040623	A1	09 February 2023	JP	7398830	B2	15 December 2023
				JP	2023024296	A	16 February 2023
				KR	102346900	B1	04 January 2022
				US	2023040623	A1	09 February 2023

US	2005075752	A1	07 April 2005	EP	1522911	A2	13 April 2005
				JP	2005111607	A	28 April 2005
				US	2005075752	A1	07 April 2005

US	2021276188	A1	09 September 2021	EP	4115340	A1	11 January 2023
				US	2021276188	A1	09 September 2021
				WO	2021178872	A1	10 September 2021

A. KLASSIFIZIERUNG DES ANMELDUNGSGEGENSTANDES		
INV. G05B13/02		
ADD.		
Nach der Internationalen Patentklassifikation (IPC) oder nach der nationalen Klassifikation und der IPC		
B. RECHERCHIERTE GEBIETE		
Recherchierter Mindestprüfstoff (Klassifikationssystem und Klassifikationssymbole)		
G05B B25J G06N		
Recherchierte, aber nicht zum Mindestprüfstoff gehörende Veröffentlichungen, soweit diese unter die recherchierten Gebiete fallen		
Während der internationalen Recherche konsultierte elektronische Datenbank (Name der Datenbank und evtl. verwendete Suchbegriffe)		
EPO-Internal, WPI Data		
C. ALS WESENTLICH ANGESEHENE UNTERLAGEN		
Kategorie*	Bezeichnung der Veröffentlichung, soweit erforderlich unter Angabe der in Betracht kommenden Teile	Betr. Anspruch Nr.
X	US 2018/089589 A1 (Ooba Masafumi [JP]) 29. März 2018 (2018-03-29) in der Anmeldung erwähnt das ganze Dokument -----	1 - 17
X	US 2023/040623 A1 (Le Pham-Tuyen [KR] ET AL) 9. Februar 2023 (2023-02-09) in der Anmeldung erwähnt das ganze Dokument ----- - / - -	1 - 17
<input checked="" type="checkbox"/> Weitere Veröffentlichungen sind der Fortsetzung von Feld C zu entnehmen <input checked="" type="checkbox"/> Siehe Anhang Patentfamilie		
* Besondere Kategorien von angegebenen Veröffentlichungen : "A" Veröffentlichung, die den allgemeinen Stand der Technik definiert, aber nicht als besonders bedeutsam anzusehen ist "E" frühere Anmeldung oder Patent, die bzw. das jedoch erst am oder nach dem internationalen Anmeldedatum veröffentlicht worden ist "L" Veröffentlichung, die geeignet ist, einen Prioritätsanspruch zweifelhaft erscheinen zu lassen, oder durch die das Veröffentlichungsdatum einer anderen im Recherchenbericht genannten Veröffentlichung belegt werden soll oder die aus einem anderen besonderen Grund angegeben ist (wie ausgeführt) "O" Veröffentlichung, die sich auf eine mündliche Offenbarung, eine Benutzung, eine Ausstellung oder andere Maßnahmen bezieht "P" Veröffentlichung, die vor dem internationalen Anmeldedatum, aber nach dem beanspruchten Prioritätsdatum veröffentlicht worden ist		"T" Spätere Veröffentlichung, die nach dem internationalen Anmeldedatum oder dem Prioritätsdatum veröffentlicht worden ist und mit der Anmeldung nicht kollidiert, sondern nur zum Verständnis des der Erfindung zugrundeliegenden Prinzips oder der ihr zugrundeliegenden Theorie angegeben ist "X" Veröffentlichung von besonderer Bedeutung; die beanspruchte Erfindung kann allein aufgrund dieser Veröffentlichung nicht als neu oder auf erfinderischer Tätigkeit beruhend betrachtet werden "Y" Veröffentlichung von besonderer Bedeutung; die beanspruchte Erfindung kann nicht als auf erfinderischer Tätigkeit beruhend betrachtet werden, wenn die Veröffentlichung mit einer oder mehreren Veröffentlichungen dieser Kategorie in Verbindung gebracht wird und diese Verbindung für einen Fachmann naheliegend ist "&" Veröffentlichung, die Mitglied derselben Patentfamilie ist
Datum des Abschlusses der internationalen Recherche		Absendedatum des internationalen Recherchenberichts
31. Mai 2024		19/06/2024
Name und Postanschrift der Internationalen Recherchenbehörde Europäisches Patentamt, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016		Bevollmächtigter Bediensteter Gardella, Simone

C. (Fortsetzung) ALS WESENTLICH ANGESEHENE UNTERLAGEN		
Kategorie*	Bezeichnung der Veröffentlichung, soweit erforderlich unter Angabe der in Betracht kommenden Teile	Betr. Anspruch Nr.
A	<p>LI PEIJIN ET AL: "A Q-learning Control Method for a Soft Robotic Arm Utilizing Training Data from a Rough Simulator", 2021 IEEE INTERNATIONAL CONFERENCE ON ROBOTICS AND BIOMIMETICS (ROBIO), IEEE, 27. Dezember 2021 (2021-12-27), Seiten 839-845, XP034106495, DOI: 10.1109/ROBIO54168.2021.9739524 * I. Introduction * * II. Methods *</p> <p style="text-align: center;">-----</p>	1,6-10, 14-16
A	<p>US 2005/075752 A1 (BAN KAZUNORI [JP] ET AL) 7. April 2005 (2005-04-07) Absatz [0004] - Absatz [0010] Absatz [0043] - Absatz [0046] Abbildung 6</p> <p style="text-align: center;">-----</p>	1-5,9, 10,12,13
A	<p>US 2021/276188 A1 (TANG HAORAN [US] ET AL) 9. September 2021 (2021-09-09) das ganze Dokument</p> <p style="text-align: center;">-----</p>	1,5,9, 10,13
A	<p>RYO MATSUMURA ET AL: "Learning Based Industrial Bin-picking Trained with Approximate Physics Simulator", ARXIV.ORG, CORNELL UNIVERSITY LIBRARY, 201 OLIN LIBRARY CORNELL UNIVERSITY ITHACA, NY 14853, 23. Mai 2018 (2018-05-23), XP080881337, * I. Introduction *</p> <p style="text-align: center;">-----</p>	1,9,10

Feld Nr. II Bemerkungen zu den Ansprüchen, die sich als nicht recherchierbar erwiesen haben (Fortsetzung von Punkt 2 auf Blatt 1)

Gemäß Artikel 17(2)a) wurde aus folgenden Gründen für bestimmte Ansprüche kein internationaler Recherchenbericht erstellt:

1. Ansprüche Nr. **1-8 (teilweise)**
weil sie sich auf Gegenstände beziehen, zu deren Recherche diese Behörde nicht verpflichtet ist, nämlich
Verfahren für rein gedankliche Tätigkeiten - vgl. Regel 39.1 iii).
2. Ansprüche Nr.
weil sie sich auf Teile der internationalen Anmeldung beziehen, die den vorgeschriebenen Anforderungen so wenig entsprechen, dass eine sinnvolle internationale Recherche nicht durchgeführt werden kann, nämlich
3. Ansprüche Nr.
weil es sich dabei um abhängige Ansprüche handelt, die nicht entsprechend Satz 2 und 3 der Regel 6.4 a) abgefasst sind.

Feld Nr. III Bemerkungen bei mangelnder Einheitlichkeit der Erfindung (Fortsetzung von Punkt 3 auf Blatt 1)

Diese Internationale Recherchenbehörde hat festgestellt, dass diese internationale Anmeldung mehrere Erfindungen enthält:

1. Da der Anmelder alle erforderlichen zusätzlichen Recherchegebühren rechtzeitig entrichtet hat, erstreckt sich dieser internationale Recherchenbericht auf alle recherchierbaren Ansprüche.
2. Da für alle recherchierbaren Ansprüche die Recherche ohne einen Arbeitsaufwand durchgeführt werden konnte, der zusätzliche Recherchegebühr gerechtfertigt hätte, hat die Behörde nicht zur Zahlung solcher Gebühren aufgefordert.
3. Da der Anmelder nur einige der erforderlichen zusätzlichen Recherchegebühren rechtzeitig entrichtet hat, erstreckt sich dieser internationale Recherchenbericht nur auf die Ansprüche, für die Gebühren entrichtet worden sind, nämlich auf die Ansprüche Nr.
4. Der Anmelder hat die erforderlichen zusätzlichen Recherchegebühren nicht rechtzeitig entrichtet. Dieser internationale Recherchenbericht beschränkt sich daher auf die in den Ansprüchen zuerst erwähnte Erfindung;; diese ist in folgenden Ansprüchen erfasst:

Bemerkungen hinsichtlich eines Widerspruchs

- Der Anmelder hat die zusätzlichen Recherchegebühren unter Widerspruch entrichtet und die gegebenenfalls erforderliche Widerspruchsgebühr gezahlt.
- Die zusätzlichen Recherchegebühren wurden vom Anmelder unter Widerspruch gezahlt, jedoch wurde die entsprechende Widerspruchsgebühr nicht innerhalb der in der Aufforderung angegebenen Frist entrichtet.
- Die Zahlung der zusätzlichen Recherchegebühren erfolgte ohne Widerspruch.

INTERNATIONALER RECHERCHENBERICHT

Angaben zu Veröffentlichungen, die zur selben Patentfamilie gehören

Internationales Aktenzeichen

PCT/EP2024/054821

Im Recherchenbericht angeführtes Patentdokument	Datum der Veröffentlichung	Mitglied(er) der Patentfamilie	Datum der Veröffentlichung
US 2018089589 A1	29-03-2018	CN 107866809 A	03-04-2018
		DE 102017008836 A1	29-03-2018
		JP 6514171 B2	15-05-2019
		JP 2018051664 A	05-04-2018
		US 2018089589 A1	29-03-2018
US 2023040623 A1	09-02-2023	JP 7398830 B2	15-12-2023
		JP 2023024296 A	16-02-2023
		KR 102346900 B1	04-01-2022
		US 2023040623 A1	09-02-2023
US 2005075752 A1	07-04-2005	EP 1522911 A2	13-04-2005
		JP 2005111607 A	28-04-2005
		US 2005075752 A1	07-04-2005
US 2021276188 A1	09-09-2021	EP 4115340 A1	11-01-2023
		US 2021276188 A1	09-09-2021
		WO 2021178872 A1	10-09-2021