



(10) **DE 10 2023 113 925 A1** 2024.11.28

(12) **Offenlegungsschrift**

(21) Aktenzeichen: **10 2023 113 925.8**

(22) Anmeldetag: **26.05.2023**

(43) Offenlegungstag: **28.11.2024**

(51) Int Cl.: **G06V 20/58** (2022.01)

G06V 10/764 (2022.01)

(71) Anmelder:

**CARIAD SE, 38440 Wolfsburg, DE; Robert Bosch
Gesellschaft mit beschränkter Haftung, 70469
Stuttgart, DE**

(74) Vertreter:

**Bösherz Goebel Partnerschaft von
Patentanwälten mbB, 40670 Meerbusch, DE**

(72) Erfinder:

**Tananaev, Denis, 71067 Sindelfingen, DE;
Backfisch, Niels, 30519 Hannover, DE**

(56) Ermittelter Stand der Technik:

DE	10 2007 049 706	A1
DE	10 2009 048 699	A1
US	2023 / 0 351 769	A1

**Soumik Sarkar, „Using Deep Convolutional
Networks for Occlusion Edge Detection in RGB-D
Frames arXiv:1412.7007v3 [cs.CV] 8 Jul 2015**

Prüfungsantrag gemäß § 44 PatG ist gestellt.

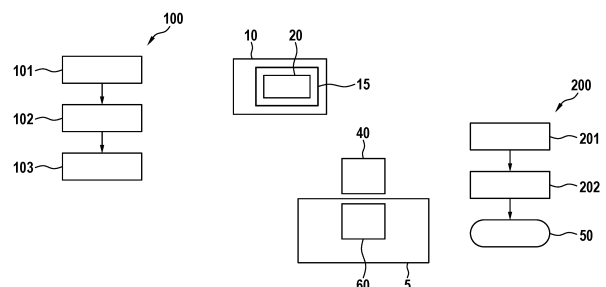
Rechercheantrag gemäß § 43 PatG ist gestellt.

Die folgenden Angaben sind den vom Anmelder eingereichten Unterlagen entnommen.

(54) Bezeichnung: **Verfahren zur Detektion von wenigstens einem Hindernis bei einem automatisierten und/oder zumindest teilautonomen Fahrsystem**

(57) Zusammenfassung: Die Erfindung betrifft ein Verfahren (100) zur Detektion von wenigstens einem Hindernis bei einem automatisierten und/oder zumindest teilautonomen Fahrsystem (60), umfassend die nachfolgenden Schritte:

- Bereitstellen (101) von Bilddaten, wobei die Bilddaten für eine Aufzeichnung einer Umgebung des Fahrsystems (60) spezifisch sind,
- Durchführen (102) einer Auswertung der bereitgestellten Bilddaten, wobei die Auswertung auf Basis einer Anwendung eines Maschinenlern-Modells (50) erfolgt, durch welches eine Verdeckungskennzeichnung zu wenigstens einer Verdeckung der Umgebung ermittelt wird,
- Durchführen (103) der Detektion des wenigstens eines Hindernisses auf Basis der ermittelten Verdeckungskennzeichnung.



Beschreibung

[0001] Die vorliegende Erfindung betrifft ein Verfahren zur Detektion von wenigstens einem Hindernis bei einem automatisierten und/oder zumindest teilautonomen Fahrsystem. Ferner bezieht sich die Erfindung auf ein Trainingsverfahren, ein Computerprogramm, eine Vorrichtung, ein Computerlesbares Speichermedium sowie ein Maschinenlern-Modell.

Stand der Technik

[0002] Eine der wichtigsten Herausforderungen bei der Wahrnehmung (Perzeption) für autonome mobile Roboter oder Fahrerassistenzsysteme besteht darin, gefährliche Objekte zuverlässig zu erkennen. Damit soll eine sichere Navigation in einer 3D-Umgebung ermöglicht werden.

[0003] Herkömmliche lernbasierte Objekterkennungsalgorithmen auf der Grundlage von konvolutionellen neuronalen Netzen (engl. convolutional neural networks oder kurz CNN bezeichnet) können oft keine allgemeine Repräsentation von gefährlichen Objekten erlernen, wenn keine ausreichend große Menge an vom Menschen annotierten Beispielen aller möglichen Varianten der gefährlichen Objekte vorhanden sind.

[0004] Da sich die manuelle Beschriftung sämtlicher möglicher generischer Objekte praktisch kaum realisieren lässt, werden häufig für die Erkennung gefährlicher Objekte Algorithmen eingesetzt, die auf Heuristiken und deterministischen Formulierungen basieren. Ein Beispiel für einen solchen Algorithmus stellt P. Pinggera, U. Franke and R. Mester, „High-performance long range obstacle detection using stereo vision,“ 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Hamburg, Germany, 2015, pp. 1308-1313, doi: 10.1109/IROS.2015.7353537 dar. Dieser Ansatz kann zwar einige unerwartete Hindernisse erkennen, lässt sich aber oft nicht auf verschiedene Szenarien verallgemeinern und setzt außerdem eine Stereokamera voraus. In vielen Robotersystemen kommt jedoch eine Monokamera zum Einsatz.

Offenbarung der Erfindung

[0005] Gegenstand der Erfindung ist ein Verfahren mit den Merkmalen des Anspruchs 1, ein Trainingsverfahren mit den Merkmalen des Anspruchs 7, ein Computerprogramm mit den Merkmalen des Anspruchs 8, eine Vorrichtung mit den Merkmalen des Anspruchs 9, ein computerlesbares Speichermedium mit den Merkmalen des Anspruchs 10 sowie ein Maschinenlern-Modell mit den Merkmalen des Anspruchs 11. Weitere Merkmale und Details der Erfindung ergeben sich aus den jeweiligen Unteransprüchen, der Beschreibung und den Zeichnungen. Dabei gelten Merkmale und Details, die im Zusammenhang mit dem erfindungsgemäßen Verfahren beschrieben sind, selbstverständlich auch im Zusammenhang mit dem erfindungsgemäßen Trainingsverfahren, dem erfindungsgemäßen Computerprogramm, der erfindungsgemäßen Vorrichtung, dem erfindungsgemäßen computerlesbaren Speichermedium sowie dem erfindungsgemäßen Maschinenlern-Modell, und jeweils umgekehrt, so dass bezüglich der Offenbarung zu den einzelnen Erfindungsaspekten stets wechselseitig Bezug genommen wird bzw. werden kann.

[0006] Gegenstand der Erfindung ist insbesondere ein Verfahren zur Detektion von wenigstens einem Hindernis bei einem automatisierten und/oder zumindest teilautonomen Fahrsystem, umfassend die nachfolgenden Schritte, welche vorzugsweise nacheinander und/oder wiederholt ausgeführt werden:

- Bereitstellen von Bilddaten, wobei die Bilddaten für eine Aufzeichnung einer Umgebung des Fahrsystems spezifisch sind,
- Durchführen einer Auswertung der bereitgestellten Bilddaten, wobei die Auswertung auf Basis einer Anwendung eines Maschinenlern-Modells erfolgen kann, durch welches eine Verdeckungskennzeichnung zu wenigstens einer Verdeckung der Umgebung ermittelt wird,
- Durchführen der Detektion des wenigstens eines Hindernisses, insbesondere eines gefährlichen Objektes, auf Basis der ermittelten Verdeckungskennzeichnung.

[0007] Die Erfindung kann den Vorteil haben, dass Einschränkungen von herkömmlichen lernbasierten Ansätzen überwunden werden. Dies kann konkret die mangelnde Verfügbarkeit von gelabelten Daten betreffen. Dabei kann erfindungsgemäß anstelle einer unmittelbaren Klassifizierung von gefährlichen Objekten anhand der Bilddaten zunächst - als Zwischenschritt - die Verdeckungskennzeichnung mittels Maschinenlernen ermittelt werden. Darüber hinaus kann es die Erfindung ermöglichen, einen qualitativ hochwertigen generischen Detektor für gefährliche Objekte zu trainieren, welcher die Detektion auf Basis der ermittelten Verdeckungskennzeichnung durchführt. Wie nachfolgend noch näher beschrieben wird, kann hierzu

selbstüberwachtes Training mit bestimmten überwachten Elementen verwendet werden. Auch kann es möglich sein, den erfindungsgemäßen Ansatz nicht nur in Stereokameras, sondern auch in Monokamerasystemen zuverlässig anzuwenden. In anderen Worten können die Bilddaten Einzelbilder anstelle von Bildsequenzen umfassen, sodass auf eine Bewegungsinformation für die Anwendung des Maschinenlern-Modells verzichtet werden kann.

[0008] Außerdem ist es von Vorteil, wenn ein Training des Maschinenlern-Modells darauf basiert, dass auf Basis einer Bewegung bei einer Kameraaufzeichnung ein Verdeckungsbereich ermittelt wird. Hierzu kann ein optischer Fluss bei einer aus der Kameraaufzeichnung resultierenden Bildersequenz geschätzt werden. Sodann kann anhand des geschätzten optischen Flusses das Maschinenlern-Modell zur Ermittlung der Verdeckungskennzeichnung trainiert werden, insbesondere zur Ermittlung der Verdeckungskennzeichnung nur anhand Bilddaten in der Form eines Einzelbildes trainiert wird. Das Training kann bevorzugt in der Form eines selbstüberwachten Trainings durchgeführt werden. Das Maschinenlern-Modell kann ferner als ein CNN ausgeführt sein. Im Training kann das Maschinenlern-Modell als Eingabe eine Bildsequenz umfassend zwei (oder mehr) aufeinanderfolgende Bilder aus einem monokularen Video erhalten und auf dieser Basis eine Schätzung des optischen Flusses liefern. Als Verlustfunktion kann die photometrische Fehlerminimierung verwendet werden. Der Verdeckungsbereich kann den räumlichen Bereich bezeichnen, welcher in der Umgebung durch das Hindernis verdeckt wird.

[0009] Eine Besonderheit bei einem Training des Maschinenlern-Modells ist insbesondere, dass sich zunutze gemacht werden kann, dass der optische Fluss nie alle Pixel erfassen kann, wenn sich die Kamera bewegt. Dieser Aspekt kann entweder direkt genutzt werden, um die Verdeckungskennzeichnung in der Form einer Hinderniskarte (engl. obstacle map) zu erzeugen, oder indirekt, um die Verdeckungskennzeichnung in der Form einer anderen Art von Hinderniskarte gemäß einer im Folgenden beschriebenen „weiteren Verlustoption“ zu erzeugen. Die dadurch ermittelte Verdeckungskennzeichnung kann auch als Hindernispunktwolke ausgeführt sein. Eine weitere Besonderheit der Erfindung kann es sein, dass das selbstüberwachte Training auf großen Mengen von ungelabelten Daten zurückgreifen kann. Ferner können optional Bounding Boxes aus der Hindernispunktwolke erstellt werden und in einer weiteren Phase ein Falsch-Positiv-Reduktions-Klassifikator auf den erkannten Objektkandidaten laufen gelassen werden.

[0010] Weiter ist im Rahmen der Erfindung denkbar, dass die Bilddaten (insbesondere im Inferenzmodus) wenigstens ein oder genau ein Einzelbild umfassen, welches aus einer Aufzeichnung mittels einer monokularen oder Stereo-Kamera resultiert. Mit anderen Worten: Genau ein Einzelbild kann nach einem Training, wie oben beschrieben, im Inferenzmodus verwendbar sein. Bevorzugt werden die für das Maschinenlern-Modell als Eingabe verwendeten Bilddaten zur Ermittlung der Verdeckungskennzeichnung auf das Einzelbild beschränkt. Mit anderen Worten benötigt das Maschinenlern-Modell als Eingabe keine Bewegungsinformation, um die Verdeckungskennzeichnung zu ermitteln, vorzugsweise zu erzeugen.

[0011] Optional kann es vorgesehen sein, dass die Verdeckungskennzeichnung für die wenigstens eine Verdeckung spezifisch ist, und/oder als eine Verdeckungskarte ausgeführt ist, welche wenigstens einen oder mehrere durch wenigstens ein Objekt in der Umgebung verdeckten Bereich in den Bilddaten kennzeichnet. Die Verdeckungskarte kann dabei auch als eine Verdeckungsmaske ausgeführt sein, welche bspw. für einzelne Elemente wie Pixel der Bilddaten angibt, vorzugsweise binär angibt, ob diese durch wenigstens ein Objekt verdeckt werden.

[0012] Bevorzugt kann im Rahmen der Erfindung vorgesehen sein, dass die Detektion (des) wenigstens eines Hindernisses eine Bewertung der Verdeckungskennzeichnung, vorzugsweise mittels eines Klassifikators, bevorzugt eines mittels Maschinellen trainierten Klassifikators, umfasst, bei welchem eine Einstufung eines der jeweiligen Verdeckung zugeordneten und in den Bilddaten anhand der Verdeckungskennzeichnung detektierten Objekts als ein gefährliches Objekt vorgenommen wird. Das gefährliche Objekt kann insbesondere eine abgefallene Ladung eines Lastkraftfahrzeuges umfassen. Ferner kann das gefährliche Objekt auch als ein Objekt bezeichnet werden, welches potentiell für ein bewegtes Fahrzeug und/oder einen bewegten Roboter mit dem Fahrsystem gefährlich sein kann. Der Klassifikator kann sich darauf beschränken, festzustellen, ob es sich bei dem Objekt um ein gefährliches Objekt handelt. Auf eine Einstufung in weitere Klassen kann somit verzichtet werden, der Klassifikator kann somit klassenagnostisch ausgeführt sein.

[0013] Ferner ist es optional vorgesehen, dass auf Basis der Auswertung und/oder der Detektion eine zumindest teilautonome Steuerung eines Ego-Fahrzeuges und/oder Roboters mit dem Fahrsystem, vorzugsweise durch ein Bewegungsplanungssystem, durchgeführt wird. Das Fahrsystem kann bspw. als ein Fahrerassistenzsystem oder ein autonomes Fahrsystem ausgeführt sein, z. B. für den Einsatz bei autonomen mobilen

Robotern oder autonomen Fahrzeugen. Dabei kann ein Wahrnehmungssystem eine Darstellung der 3D-Umgebung liefern, und diese Darstellung als Eingabe für ein Bewegungsplanungssystem dienen, das dann entscheidet, wie das Ego-Fahrzeug manövriert werden soll.

[0014] Ebenfalls Gegenstand der Erfindung ist ein Trainingsverfahren zum Trainieren eines Maschinenlern-Modells, umfassend:

- Bereitstellen von Trainingsdaten, wobei die Trainingsdaten wenigstens eine Bildersequenz umfassen, die eine Umgebung eines Fahrsystems während einer Fahrt repräsentieren können, wobei die Trainingsdaten ferner Annotationsdaten umfassen können, insbesondere als Ground-Truth, die eine Verdeckungskennzeichnung angeben, welche wenigstens eine Verdeckung der Umgebung während der Fahrt repräsentiert,
- Durchführen eines Trainings des Maschinenlern-Modells anhand der Trainingsdaten, bei welchem ein optischer Fluss bei der Bildersequenz berücksichtigt wird, um die Verdeckungskennzeichnung zu prädiizieren.

[0015] Damit bringt das erfindungsgemäße Trainingsverfahren die gleichen Vorteile mit sich, wie sie ausführlich mit Bezug auf ein erfindungsgemäßes Verfahren zur Detektion von wenigstens einem Hindernis beschrieben sind. Das bei dem erfindungsgemäßen Verfahren zur Detektion von wenigstens einem Hindernis angewandte Maschinenlern-Modell kann dabei bevorzugt aus dem erfindungsgemäßen Trainingsverfahren resultieren. Ebenfalls Gegenstand der Erfindung ist das Maschinenlern-Modell, welches durch das erfindungsgemäße Trainingsverfahren erhalten wird.

[0016] Ferner ist es bei dem Trainingsverfahren denkbar, dass auf Basis des geschätzten optischen Flusses und der Verdeckungskennzeichnung in der Form einer Verdeckungskarte, welche (die) wenigstens eine Verdeckung angibt, und vorzugsweise einer Kalibrierungsmatrix der Kamera eine essenzielle Matrix berechnet wird. Es kann als weiterer Schritt eine 3D-Punkt-Triangulation und/oder Tiefenschätzung vorgenommen werden. Anhand der relativen Transformationen zwischen zwei Bildern der Bildsequenz kann die Triangulation angewendet werden, um 3D-Punkte für jede Punktkorrespondenz aus dem optischen Fluss zu erhalten.

[0017] Ebenfalls Gegenstand der Erfindung ist ein Computerprogramm, insbesondere Computerprogrammprodukt, umfassend Befehle, die bei der Ausführung des Computerprogrammes durch einen Computer diesen veranlassen, das erfindungsgemäße Verfahren auszuführen. Damit bringt das erfindungsgemäße Computerprogramm die gleichen Vorteile mit sich, wie sie ausführlich mit Bezug auf ein erfindungsgemäßes Verfahren beschrieben worden sind.

[0018] Ebenfalls Gegenstand der Erfindung ist eine Vorrichtung zur Datenverarbeitung, die eingerichtet ist, das erfindungsgemäße Verfahren auszuführen. Als die Vorrichtung kann bspw. ein Computer vorgesehen sein, welcher das erfindungsgemäße Computerprogramm ausführt. Der Computer kann wenigstens einen Prozessor zur Ausführung des Computerprogramms aufweisen. Auch kann ein nicht-flüchtiger Datenspeicher vorgesehen sein, in welchem das Computerprogramm hinterlegt und von welchem das Computerprogramm durch den Prozessor zur Ausführung ausgelesen werden kann.

[0019] Ebenfalls Gegenstand der Erfindung kann ein computerlesbares Speichermedium sein, welches das erfindungsgemäße Computerprogramm aufweist und/oder Befehle umfasst, die bei der Ausführung durch einen Computer diesen veranlassen, das erfindungsgemäße Verfahren auszuführen. Das Speichermedium ist bspw. als ein Datenspeicher wie eine Festplatte und/oder ein nicht-flüchtiger Speicher und/oder eine Speicherkarte ausgebildet. Das Speichermedium kann z. B. in den Computer integriert sein.

[0020] Darüber hinaus kann das erfindungsgemäße Verfahren auch als ein computerimplementiertes Verfahren ausgeführt sein.

[0021] Weitere Vorteile, Merkmale und Einzelheiten der Erfindung ergeben sich aus der nachfolgenden Beschreibung, in der unter Bezugnahme auf die Zeichnungen Ausführungsbeispiele der Erfindung im Einzelnen beschrieben sind. Dabei können die in den Ansprüchen und in der Beschreibung erwähnten Merkmale jeweils einzeln für sich oder in beliebiger Kombination erfindungswesentlich sein. Es zeigen:

Fig. 1 eine schematische Visualisierung eines Verfahrens, einer Vorrichtung, eines Speichermediums, eines Maschinenlern-Modells, eines Trainingsverfahrens sowie eines Computerprogramms gemäß Ausführungsbeispielen der Erfindung.

Fig. 2 eine weitere schematische Darstellung zur Visualisierung eines Trainingsverfahrens gemäß Ausführungsbeispielen der Erfindung.

Fig. 3 eine Visualisierung einer Ermittlung der Verdeckungskennzeichnung mit bewegter monokularer Kamera.

Fig. 4 eine weitere Visualisierung einer Ermittlung der Verdeckungskennzeichnung mit bewegter monokularer Kamera.

[0022] In **Fig. 1** sind ein Verfahren 100, eine Vorrichtung 10, ein Speichermedium 15 sowie ein Computerprogramm 20 gemäß Ausführungsbeispielen der Erfindung schematisch dargestellt. Dabei ist gezeigt, dass bei dem Verfahren 100 zur Detektion von wenigstens einem Hindernis bei einem automatisierten und/oder zumindest teilautonomen Fahrsystem 60 gemäß einem ersten Verfahrensschritt 101 zunächst Bilddaten bereitgestellt werden können. Die Bilddaten können für eine Aufzeichnung einer Umgebung des Fahrsystems 60 spezifisch sein. Sodann kann gemäß einem zweiten Verfahrensschritt 102 eine Auswertung der bereitgestellten Bilddaten erfolgen, wobei die Auswertung auf Basis einer Anwendung eines Maschinenlern-Modells 50 erfolgt, durch welches eine Verdeckungskennzeichnung zu wenigstens einer Verdeckung der Umgebung ermittelt wird. In einem dritten Verfahrensschritt 103 kann daraufhin eine Detektion des wenigstens eines Hindernisses auf Basis der ermittelten Verdeckungskennzeichnung erfolgen. Darüber hinaus kann auf Basis der Auswertung und/oder der Detektion eine zumindest teilautonome Steuerung eines Ego-Fahrzeuges 5 und/oder Roboters 5 mit dem Fahrsystem 60, vorzugsweise durch ein Bewegungsplanungssystem, durchgeführt werden.

[0023] In den **Fig. 2 bis 4** wird weiter veranschaulicht, dass ein Training des Maschinenlern-Modells 50 darauf basieren kann, dass auf Basis einer Bewegung 303 bei einer Kameraaufzeichnung ein Verdeckungsbereich 304,401 ermittelt wird, wobei hierzu vorzugsweise ein optischer Fluss bei einer aus der Kameraaufzeichnung resultierenden Bildersequenz geschätzt wird und anhand des geschätzten optischen Flusses das Maschinenlern-Modell 50 zur Ermittlung der Verdeckungskennzeichnung trainiert wird, vorzugsweise im Gesamtverlust als L_o bezeichnet, wobei bevorzugt das Training in der Form eines selbstüberwachten Trainings durchgeführt wird. Auch ist es möglich, dass ein Verdeckungsbereich durch eine Berechnung der bestmöglichen den Pixel beschreibenden Normalen bestimmt wird (repräsentiert durch L_{repr}). Letztere Verdeckungskennzeichnung kann auch aus Stereobildern gewonnen werden. Die Verdeckungskennzeichnung kann dabei für die wenigstens eine Verdeckung spezifisch sein, und vorzugsweise als eine Verdeckungskarte ausgeführt sein, welche wenigstens einen oder mehrere durch wenigstens ein Objekt 70 in der Umgebung verdeckten Bereich 304,401 in den Bilddaten kennzeichnet.

[0024] In **Fig. 1** ist ferner ein Trainingsverfahren 200 zum Trainieren eines Maschinenlern-Modells 50 veranschaulicht, bei welchem gemäß einem ersten Trainingsschritt 201 die Trainingsdaten bereitgestellt werden und gemäß einem zweiten Trainingsschritt 202 das Training des Maschinenlern-Modells 50 erfolgt, um die Verdeckungskennzeichnung zu präzisieren. Auf diese Weise kann das trainierte Maschinenlern-Modell 50 erhalten werden.

[0025] Ausführungsbeispiele der Erfindung können den Vorteil haben, dass ein trainierbarer Algorithmus ohne die Notwendigkeit einer großen Menge an gelabelten Daten bereitgestellt wird. Während es relativ einfach ist, große Mengen ungelabelter Daten zu sammeln, ist die Verarbeitung und Markierung dieser Daten für die Verwendung in überwachten Algorithmen wie der CNN-basierten Objekterkennung sehr kostspielig oder im Falle einer unbekannten Menge von Objekten, wie z. B. gefährlichen Objekten, nicht oder kaum durchführbar. Es kann dabei gemäß Ausführungsbeispielen der Erfindung ein Algorithmus bereitgestellt werden, welcher auch für Monokamera-Setups geeignet ist. Im Gegensatz zu deterministischen Algorithmen können dabei die lernbasierten Algorithmen gemäß Ausführungsbeispielen der Erfindung anpassungsfähig sein. Das bedeutet, dass sie trainiert werden können, um Problemfälle durch Hinzufügen von Daten zu lösen. Diese Anpassung und das Trainieren von schwierigen Situationen, die vor der ersten Veröffentlichung und Anwendung des Algorithmus nicht bekannt sind, sind hingegen mit nicht lernbasierten Ansätzen nicht möglich. Ausführungsbeispiele der Erfindung können dabei nicht nur für Stereo, sondern auch für Monokameras geeignet sein.

[0026] Ausführungsbeispiele können eine Detektion von gefährlichen Objekten ermöglichen, um die Navigation autonomer Systeme zu ermöglichen. Ferner kann die Erstellung von HD-Karten ermöglicht werden.

[0027] In fortschrittlichen Fahrerassistenzsystemen oder autonomen Fahrsystemen liefert das Wahrnehmungssystem eine Darstellung der 3D-Umgebung, und diese Darstellung dient als Eingabe für ein Bewe-

gungsplanungssystem, das dann entscheidet, wie das Ego-Fahrzeug manövriert werden soll. Eine Schlüsseltechnologie des Wahrnehmungssystems besteht darin, zu erkennen, wohin das Fahrzeug fahren kann und wie die Umgebung um das Auto herum aussieht. Es sind herkömmliche Computer-Vision-Techniken bekannt, die häufig nicht robust genug sein können, da sie nicht wie maschinelle Lerntechniken lernen können. Im Gegensatz dazu liefern lernbasierte Methoden hervorragende Ergebnisse, erfordern aber eine hohe Anzahl von Labels, also ein manuelles Annotieren der Daten. Ausführungsbeispiele der Erfindung nutzen qualitativ hochwertige lernbasierte Ansätze und können das Problem des Labelling mit selbstüberwachtem Vortraining lösen, wodurch die erforderliche Menge an Annotationen der Daten erheblich reduziert werden kann. Selbstüberwachtes Training kann sich dabei auf ein Trainingsverfahren im Zusammenhang mit Maschinlernen bzw. künstlicher Intelligenz beziehen, bei dem das Modell aus ungelabelten Daten lernt, indem es seine eigenen Vorhersagen mit den tatsächlichen Ergebnissen vergleicht und daraus lernt, ohne auf manuell annotierte Daten angewiesen zu sein. Bei einem halbüberwachten Training wird hingegen das Modell sowohl mit gelabelten Daten als auch mit ungelabelten Daten trainiert, um eine verbesserte Leistung und Generalisierungsfähigkeit zu erreichen.

[0028] Ein halbüberwachter (engl. semi-supervised) generische Algorithmus zur Hinderniserkennung kann gemäß **Fig. 2** wie folgt ausgeführt sein. Es kann zunächst ein selbstüberwachtes Training 201 eines CNN 202 durchgeführt werden. Das Training kann dabei zumindest eine der folgenden Schritte umfassen:

- ein selbstüberwachtes CNN-Training für den optischen Fluss,
- eine Berechnung einer Verdeckungsmaske (engl. occlusion mask) aus dem berechneten optischen Fluss,
- eine Schätzung einer Essential Matrix (oder auch „essentielle Matrix“ bezeichnet),
- eine 3D-Punkt-Triangulation und Tiefenschätzung sowie
- ein Training eines Einzelbild-Verdeckung CNN.

[0029] Die Essential Matrix kann dabei eine Matrix sein, die aus der Korrespondenz von Bildpunkten zwischen zwei Kamerabildern berechnet wird. Sie beschreibt auf diese Weise die Beziehung zwischen den Kamerapositionen und ermöglicht es, die Position eines Objekts in einem dreidimensionalen Raum zu rekonstruieren. Die Berechnung der Essential Matrix erfolgt bspw. durch den Einsatz von Algorithmen wie RAN-SAC (Random Sample Consensus) oder der Einführung von Constraints (z.B. Epipolargeometrie).

[0030] In **Fig. 2** sind bspw. ein Kamerabild 203, insbesondere im RGB-Farbmodell, und eine Verdeckungs-
maske 204 visualisiert. Anschließend kann eine überwachte Feinabstimmung 210 des Algorithmus 220, der auf Machine-Learning- oder Deep-Learning-Techniken basiert, oder eines DBScan-basiertes Clusterings 221 erfolgen. Sodann kann eine Reduzierung 222 der falsch-positiven Ergebnisse durch einen überwachten Klassifikator vorgesehen sein, der jeden Hinderniskandidaten verarbeitet.

[0031] In einer ersten Phase kann ein selbstüberwachtes Vortraining des CNN vorgesehen sein. Diese Phase ermöglicht es, den Mangel an gelabelten Daten in Bezug auf alle möglichen gefährlichen Objekte zu überwinden. Das Funktionsprinzip soll nachfolgend verdeutlicht werden. Jedes erhöhte Objekt führt zu Verdeckungen, siehe **Fig. 3**. Es kann daher ein CNN 202 zur Erkennung von Verdeckungen auf selbstüberwachte Weise trainiert werden (vgl. **Fig. 2**), da die Verdeckungen qualitativ hochwertige Merkmale sowie Interessengebiete (engl. regions of interest) liefern, in denen sich gefährliche Objekte befinden können. Mit einer Folge von aufeinanderfolgenden Bildern einer monokularen Kamera 40 oder Stereokamera 40 kann ein CNN trainiert werden, indem der photometrische Verlust zwischen den Bildern minimiert wird, um einen optischen Fluss zu erhalten, und den Verdeckungsverlust maximieren, um einen Verdeckungsdetektor zu erhalten. Die Eingabe für diese erste Phase sind bspw. Videosequenzen von einer monokularen Kamera 40 und die Kalibrierungsmatrix K . In der Seitenansicht gemäß **Fig. 3** sind beispielhaft zwei zeitlich aufeinanderfolgende Kamerapositionen 301, 302 dargestellt, welche sich durch die Bewegung 303 des Fahrsystems 60 verändern. Ferner ist ein Hindernis 70 dargestellt und visualisiert, dass aufgrund der Bewegung 303 ein oberer Bereich 304 der Verdeckung durch das Hindernis 70 ermittelt werden kann. In **Fig. 4** ist eine Draufsicht auf das Hindernis 70 gezeigt, wobei ebenfalls die beiden Kamerapositionen 301, 302 visualisiert sind. Ferner ist gezeigt, dass sich aufgrund der Bewegung 303 des Fahrsystems 60 auch ein Seitenbereich 401 der Verdeckung (engl. Okklusion) ermitteln lässt.

[0032] Es kann bei Ausführungsbeispielen der Erfindung ein selbstüberwachtes Training des optischen Flusses vorgesehen sein. In anderen Worten kann in einem ersten Schritt das Training des selbstüberwachten optischen Fluss-CNNs erfolgen. Das CNN kann hierzu als Eingabe zwei (oder mehr) aufeinanderfolgende

Bilder aus einem monokularen Video erhalten und auf dieser Basis eine Schätzung des optischen Flusses liefern. Als Verlustfunktion kann die photometrische Fehlerminimierung verwendet werden:

$$L_p = \sum_{t'} 0 \cdot pe(I_t, I_{t' \rightarrow t})$$

[0033] Die elementweise Multiplikation wird durch dargestellt, 0 ist die Verdeckungsmaske, $I_{t' \rightarrow t}$ = Inverse-Warp(opticalflow $t \rightarrow t'$, I_t) ist das verzerrte Bild vom Quellbild I_t zum Zielbild $I_{t'}$ unter Verwendung des optischen Flusses. Der photometrische Verlust ist:

$$pe(I_t, I_{t' \rightarrow t}) = \frac{a}{2}(1 - SSIM(I_t, I_{t' \rightarrow t})) + (1 - a)\|I_t - I_{t' \rightarrow t}\|_1$$

[0034] Dabei ist SSIM die Strukturähnlichkeit. Ebenfalls kann ein kantenbewusster Glättungsfluss angewandt werden:

$$L_{smooth} = |\partial_x opticalflow| e^{-|\partial_x I|} + |\partial_y opticalflow| e^{-|\partial_y I|}$$

[0035] Der Glättungsverlust sorgt für eine Glättung des optischen Flusses in homogenen Bereichen des Bildes und ermöglicht Änderungen des Flusses an den Rändern. Der Gesamtverlust beträgt:

$$L = w_1 * L_p + w_2 * L_{smooth}$$

[0036] Dabei sind w_1 , w_2 - die Gewichte der Verlustkomponenten. Die opticalflow $t \rightarrow t'$ ist der optische Fluss vom Zielbild $I_{t'}$ zum Quellbild I_t . Um eine Verdeckungsmaske zu berechnen, kann ferner ein CNN verwendet werden, um den entgegengesetzten optischen Fluss vom Quellbild zum Ziel opticalflow $t' \rightarrow t$ wie folgt zu berechnen:

$$V(x, y) = \sum_{i=1}^W \sum_{j=1}^H \max\left(0, 1 - \left|x - opticalflow_{t' \rightarrow t}^x(i, j)\right|\right) \\ * \max\left(0, 1 - \left|y - opticalflow_{t' \rightarrow t}^y(i, j)\right|\right)$$

[0037] Dabei ist $V(x, y)$ eine Bereichskarte am Ort (x, y) des Bildes mit der Höhe H und der Breite W , und opticalflow $t' \rightarrow t$ $\times$ opticalflow $t' \rightarrow t$ y sind die horizontalen bzw. vertikalen Komponenten des optischen Flusses. Eine Verdeckungskarte, auch Verdeckungsmaske bezeichnet, kann durch Schwellenwertbildung wie folgt ermittelt werden:

$$O = \min(1, V(x, y))$$

[0038] Hier hat die Verdeckungskarte weiche Werte zwischen 0 und 1, wobei 0 bedeutet, dass das Bild verdeckt ist und 1, dass es nicht verdeckt ist.

[0039] Anschließend kann die essentielle Matrix geschätzt werden. Mit den optischen Fluss- und Verdeckungsmasken aus den vorherigen Schritten sowie der Kalibrierungsmatrix K der Kamera kann die essentielle Matrix E geschätzt und die relative Rotation R und Translation B zwischen den Bildern mit Hilfe des Algorithmus zur Zerlegung der essentiellen Matrix ermitteln. Die essentielle Matrix beschreibt die Beziehungen zwischen den Pixeln zweier Bilder bei gegebener Koplanaritätsbedingung wie folgt:

$${}^k x^T E {}^k x'' = 0$$

[0040] So, dass: ${}^k x^T = x^T (K')^{-T}$

x^T ist der k -te transponierte Pixelpositionsvektor aus dem ersten Bild, $(K')^{-T}$ die inverse und transponierte Kalibrierungsmatrix des ersten Bildes.

[0041] So, dass: ${}^k x'' = (K'')^{-1} x''$

x'' ist der k -te Pixelpositionsvektor aus dem zweiten Bild, $(K'')^{-1}$ ist die inverse Kalibrierungsmatrix des zweiten Bildes. Um die wesentliche Matrix E zu schätzen, kann ein 5-Punkte-Algorithmus verwendet werden, der darin besteht, 5 korrespondierende Punkte auf zwei Bildern zu ermitteln und das sich daraus ergebende ein-

geschränkte Optimierungsproblem in Form einer Formulierung der kleinsten Quadrate zu lösen. Da der optische Fluss viele Ausreißer enthalten kann, können zunächst alle verdeckten Punkte maskiert und RANSAC für die Behandlung von Ausreißern verwendet werden.

[0042] Es kann als weiterer Schritt eine 3D-Punkt-Triangulation und Tiefenschätzung vorgenommen werden. Angesichts der relativen Transformationen zwischen zwei Bildern kann die Triangulation angewendet werden, um 3D-Punkte für jede Punktkorrespondenz aus dem optischen Fluss zu erhalten. Es kann die Triangulation mit einer Zwei-Vektor-Schnittpunktlösung in geschlossener Form initialisiert und dann der Fehler bei der Reprojektion mit Hilfe der Methode der kleinsten Quadrate minimiert werden.

[0043] Angesichts der relativen Transformationen zwischen den Bildern, der Kalibrierung, der Verdeckungsmasken und der triangulierten Tiefe kann das Einzelbild-Verdeckungs-CNN wie folgt trainiert werden: Das Netz kann als Eingabe Einzelbilder erhalten oder auf einem Stereobild arbeiten und sodann eine binäre Maske für verdeckte Objekte, ein Vektorfeld von Normalen $\vec{n}$ der Ebene, die jeder Punkt mit den umgebenden Punkten in einer kleinen Umgebung Ω bildet, und eine Tiefenschätzung ausgeben. Die Vertrauensmaske kann auf überwachte Weise trainiert werden, indem der binäre Kreuzentropieverlust und die Verdeckungsmaske aus dem optischen Fluss als Ground Truth verwendet werden.

$$L_o = \text{binaryCrossEntropy}(\text{Prediction}, 0)$$

[0044] Hier ist prediction die Vorhersage des CNN im Bereich $[0, 1]$, wobei 1 keine Verdeckung bedeutet und 0 verdeckt ist. 0 beschreibt die Verdeckungsmaske aus dem optischen Fluss, wobei 1 nicht verdeckt und 0 verdeckt ist. Die Tiefe kann auch für verdeckte Objekte unter Verwendung des L1-Verlustes gelernt werden.

$$L_d = \|d - \hat{d}\|_1$$

[0045] Dabei ist d die vorhergesagte Disparität und $\hat{d}$ die wahre Disparität. Die Tiefe kann als $\text{Depth} = 1,0 / d$ berechnet werden. Die Oberflächennormale für erhöhte Objekte kann berechnet werden, indem zunächst die Homographie berechnet wird:

$$H_i = K \left(R - \frac{1}{g} \vec{b} \vec{n}_i^T \right) K^{-1}$$

[0046] Dabei ist H die Homographie, K die Kalibrierungsmatrix, g der Skalierungsfaktor, $\vec{b}$ der Translationsvektor und $\vec{n}_i$ die Vektornormale der Oberflächenebene am Ort $i \in \text{Pos}$, wobei sich Pos auf alle räumlichen Orte in dem vom CNN erzeugten Vektorfeld bezieht. Es kann die Position einer Ebene an der Position i mit $\Theta_i = (\vec{n}_i, d_i)$ identifiziert werden, wobei d_i die Tiefe an der Position i ist. Es gibt bspw. zwei Möglichkeiten, eine Verlustfunktion zu definieren. Die erste Option zielt darauf ab, die Oberflächennormale $\vec{n}$ direkt zu regressieren, wobei vernachlässigt wird, ob es sich um ein Hindernis oder die Straßenebene handelt. In diesem Fall kann der geglättete L1-Verlust verwendet werden:

$$L_{repr} = \sum_{i \in \text{Pos}} \sum_{j \in \Omega} \| \text{HomographyWarp}(I(j), \Theta_i) - I(j) \|_1$$

[0047] Hier verzerrt HomographyWarp den Teil des Bildes mit Homographie, und I ist das Originalbild.

[0048] Anhand des Winkels der berechneten Normalen $\vec{n}$ kann nun in jedem Punkt eine Schätzung errechnet werden, ob an diesem Punkt ein gefährliches Objekt steht. Dies geschieht insbesondere dann, wenn der Winkel einen bestimmten Winkel g übersteigt. Auf diese Weise kann eine Hinderniskarte bzw. Punktwolke erzeugt werden.

[0049] Die zweite Option ist komplizierter, bietet aber den Vorteil, durch Hypothesentests eine zusätzliche Hindernispunktwolke zu erzeugen. Bei dieser Option gibt das CNN zwei Vektorfelder aus: $\vec{n}_f$, das die Straßenebene/den freien Raum darstellt, und $\vec{n}_o$, das die Oberfläche eines Objekts darstellt. Was die Verlustberechnung kompliziert macht, ist die Tatsache, dass eine Ground Truth oder eine Entscheidung darüber verfügbar sein muss, welcher der „richtige“ Normalenvektor ist. Dies liegt daran, dass der Verlust nur für den Beitrag des „richtigen“ Normalenvektors addiert werden soll. Um dies auf unüberwachte Weise zu tun, können dem Label „Straßenebene/Freiraum“ f zu, wenn

$$\sum_{j \in \Omega} \left\| \text{HomographyWarp}\left(I(j), \Theta_i^f\right) - I(j) \right\|_1$$

$$\geq \sum_{j \in \Omega} \left\| \text{HomographyWarp}\left(I(j), \Theta_i^o\right) - I(j) \right\|_1$$

[0050] Mit dieser Zuordnung kann die weitere (zweite) Verlustoption formuliert werden:

$$L_{repr} = \sum_{i \in \text{Pos}} \sum_{j \in \Omega} 1_{i \text{ has label } f} \left\| \text{HomographyWarp}\left(I(j), \Theta_i^f\right) - I(j) \right\|_1 +$$

$$1_{i \text{ has label } o} \left\| \text{HomographyWarp}\left(I(j), \Theta_i^o\right) - I(j) \right\|_1$$

[0051] Für diese weitere (zweite) Verlustoption kann eine Hindernispunktwolke auf der Grundlage eines Verhältnis-Hypothesentests erstellt werden:

$$\frac{\sum_{j \in \Omega} \left\| \text{HomographyWarp}\left(I(j), \Theta_i^o\right) - I(j)(\Theta) \right\|_1}{\sum_{j \in \Omega} \left\| \text{HomographyWarp}\left(I(j), \Theta_i^f\right) - I(j)(\Theta) \right\|_1} \geq \gamma$$

[0052] Dabei ist γ ein Schwellenwert, der anhand eines Validierungsdatensatzes kalibriert wird und beschreibt, ob der Punkt i als Objekt/Hindernis betrachtet wird. Der Gesamtverlust kann definiert sein als:

$$L = w_1 L_o + w_2 L_d + w_3 L_{repr}$$

[0053] Für den zweiten Schritt kann somit eine überwachte Option und eine gemischte Lösung mit überwachten Elementen erhalten werden.

[0054] Die in **Fig. 2** dargestellte erste Lösung 210 verwendet insbesondere eine überwachte Feinabstimmung des resultierenden CNN zur Verdeckungserkennung aus der ersten Phase. Um eine Zuordnung von selbstüberwachten Merkmalen zu den einzelnen 2D-Boxen der Objekte zu erlernen, kann auf überwachte Weise ein maschineller Lernklassifikator trainiert werden, bei dem es sich um eine SVM, eine logistische Regression oder einen CNN-basierten Head handeln kann, um eine direkte Zuordnung zwischen Verdeckungen und 2D-Objekten im Bild zu lernen. Das CNN kann klassenunabhängig trainiert werden, d. h. es gibt nur die Klasse „Hindernis“.

[0055] Die in **Fig. 2** ebenfalls dargestellte zweite Lösung 211 verwendet insbesondere die in der ersten Phase erzeugten Verdeckungspunkte und führt ein Clustering auf der Grundlage von DBScan mit einem R-Tree durch. Die äußersten Punkte der Cluster definieren Bounding Boxes von Hindernisobjekten. Diese Nachbearbeitung kann die gleiche sein wie in der Veröffentlichung P. Pinggera, U. Franke and R. Mester, „High-performance long range obstacle detection using stereo vision,“ 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Hamburg, Germany, 2015, pp. 1308-1313, doi: 10.1109/IROS.2015.7353537. Auf der Bounding-Box-Region können kleine Klassifikator laufen, z. B. ein CNN oder ein voll vernetztes neuronales Netz als Mechanismus zur Reduzierung von Fehlalarmen. Dieser Klassifikator kann vollständig unter Aufsicht trainiert werden und mehr Klassen als nur „Hindernis“ haben.

[0056] Die voranstehende Erläuterung der Ausführungsformen beschreibt die vorliegende Erfindung ausschließlich im Rahmen von Beispielen. Selbstverständlich können einzelne Merkmale der Ausführungsformen, sofern technisch sinnvoll, frei miteinander kombiniert werden, ohne den Rahmen der vorliegenden Erfindung zu verlassen.

ZITATE ENTHALTEN IN DER BESCHREIBUNG

Diese Liste der vom Anmelder aufgeführten Dokumente wurde automatisiert erzeugt und ist ausschließlich zur besseren Information des Lesers aufgenommen. Die Liste ist nicht Bestandteil der deutschen Patent- bzw. Gebrauchsmusteranmeldung. Das DPMA übernimmt keinerlei Haftung für etwaige Fehler oder Auslassungen.

Zitierte Nicht-Patentliteratur

- P. Pinggera, U. Franke and R. Mester, „High-performance long range obstacle detection using stereo vision,“ 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Hamburg, Germany, 2015, pp. 1308-1313, doi: 10.1109/IROS.2015.7353537 [0004, 0055]

Patentansprüche

1. Verfahren (100) zur Detektion von wenigstens einem Hindernis bei einem automatisierten und/oder zumindest teilautonomen Fahrsystem (60), umfassend die nachfolgenden Schritte:

- Bereitstellen (101) von Bilddaten, wobei die Bilddaten für eine Aufzeichnung einer Umgebung des Fahrsystems (60) spezifisch sind,
- Durchführen (102) einer Auswertung der bereitgestellten Bilddaten, wobei die Auswertung auf Basis einer Anwendung eines Maschinenlern-Modells (50) erfolgt, durch welches eine Verdeckungskennzeichnung zu wenigstens einer Verdeckung der Umgebung ermittelt wird,
- Durchführen (103) der Detektion des wenigstens eines Hindernisses auf Basis der ermittelten Verdeckungskennzeichnung.

2. Verfahren (100) nach Anspruch 1, **dadurch gekennzeichnet**, dass ein Training des Maschinenlern-Modells (50) darauf basiert, dass auf Basis einer Bewegung (303) bei einer Kameraaufzeichnung ein Verdeckungsbereich (304,401) ermittelt wird, wobei hierzu vorzugsweise ein optischer Fluss bei einer aus der Kameraaufzeichnung resultierenden Bildersequenz geschätzt wird und anhand des geschätzten optischen Flusses das Maschinenlern-Modell (50) zur Ermittlung der Verdeckungskennzeichnung trainiert wird, wobei bevorzugt das Training in der Form eines selbstüberwachten Trainings durchgeführt wird.

3. Verfahren (100) nach einem der vorhergehenden Ansprüche, **dadurch gekennzeichnet**, dass die Bilddaten wenigstens ein oder genau ein Einzelbild, insbesondere im Inferenzmodus, umfassen, welches aus einer Aufzeichnung mittels einer monokularen oder Stereo-Kamera resultiert, wobei bevorzugt die für das Maschinenlern-Modell (50) als Eingabe verwendeten Bilddaten zur Ermittlung der Verdeckungskennzeichnung auf das Einzelbild beschränkt werden.

4. Verfahren (100) nach einem der vorhergehenden Ansprüche, **dadurch gekennzeichnet**, dass die Verdeckungskennzeichnung für die wenigstens eine Verdeckung spezifisch ist, und vorzugsweise als eine Verdeckungskarte ausgeführt ist, welche wenigstens einen oder mehrere durch wenigstens ein Objekt in der Umgebung verdeckten Bereich in den Bilddaten kennzeichnet.

5. Verfahren (100) nach einem der vorhergehenden Ansprüche, **dadurch gekennzeichnet**, dass die Detektion des wenigstens eines Hindernisses eine Bewertung der Verdeckungskennzeichnung, vorzugsweise mittels eines Klassifikators, umfasst, bei welchem eine Einstufung eines der jeweiligen Verdeckung zugeordneten und in den Bilddaten anhand der Verdeckungskennzeichnung detektierten Objekts als ein gefährliches Objekt vorgenommen wird, wobei das gefährliche Objekt insbesondere eine abgefallene Ladung eines Lastkraftfahrzeuges umfasst.

6. Verfahren (100) nach einem der vorhergehenden Ansprüche, **dadurch gekennzeichnet**, dass auf Basis der Auswertung und/oder der Detektion eine zumindest teilautonome Steuerung eines Ego-Fahrzeuges (5) und/oder Roboters (5) mit dem Fahrsystem (60), vorzugsweise durch ein Bewegungsplanungssystem, durchgeführt wird.

7. Trainingsverfahren (200) zum Trainieren eines Maschinenlern-Modells (50), umfassend:

- Bereitstellen (201) von Trainingsdaten, wobei die Trainingsdaten wenigstens eine Bildersequenz umfassen, die eine Umgebung eines Fahrsystems (60) während einer Fahrt repräsentieren, wobei die Trainingsdaten ferner Annotationsdaten umfassen, die eine Verdeckungskennzeichnung angeben, welche wenigstens eine Verdeckung der Umgebung während der Fahrt repräsentiert,
- Durchführen (202) eines Trainings des Maschinenlern-Modells (50) anhand der Trainingsdaten, bei welchem ein optischer Fluss bei der Bildersequenz berücksichtigt wird, um die Verdeckungskennzeichnung zu präzisieren.

8. Computerprogramm (20), umfassend Befehle, die bei der Ausführung des Computerprogramms (20) durch einen Computer (10) diesen veranlassen, das Verfahren (100) nach einem der vorhergehenden Ansprüche auszuführen.

9. Vorrichtung (10) zur Datenverarbeitung, die eingerichtet ist, das Verfahren (100) nach einem der Ansprüche 1 bis 7 auszuführen.

10. Computerlesbares Speichermedium (15), umfassend Befehle, die bei der Ausführung durch einen Computer (10) diesen veranlassen, die Schritte des Verfahrens (100) nach einem der Ansprüche 1 bis 7 auszuführen.

11. Maschinlern-Modell (50), welches durch ein Trainingsverfahren (200) nach Anspruch 7 erhalten wird.

Es folgen 3 Seiten Zeichnungen

Anhängende Zeichnungen

Fig. 1

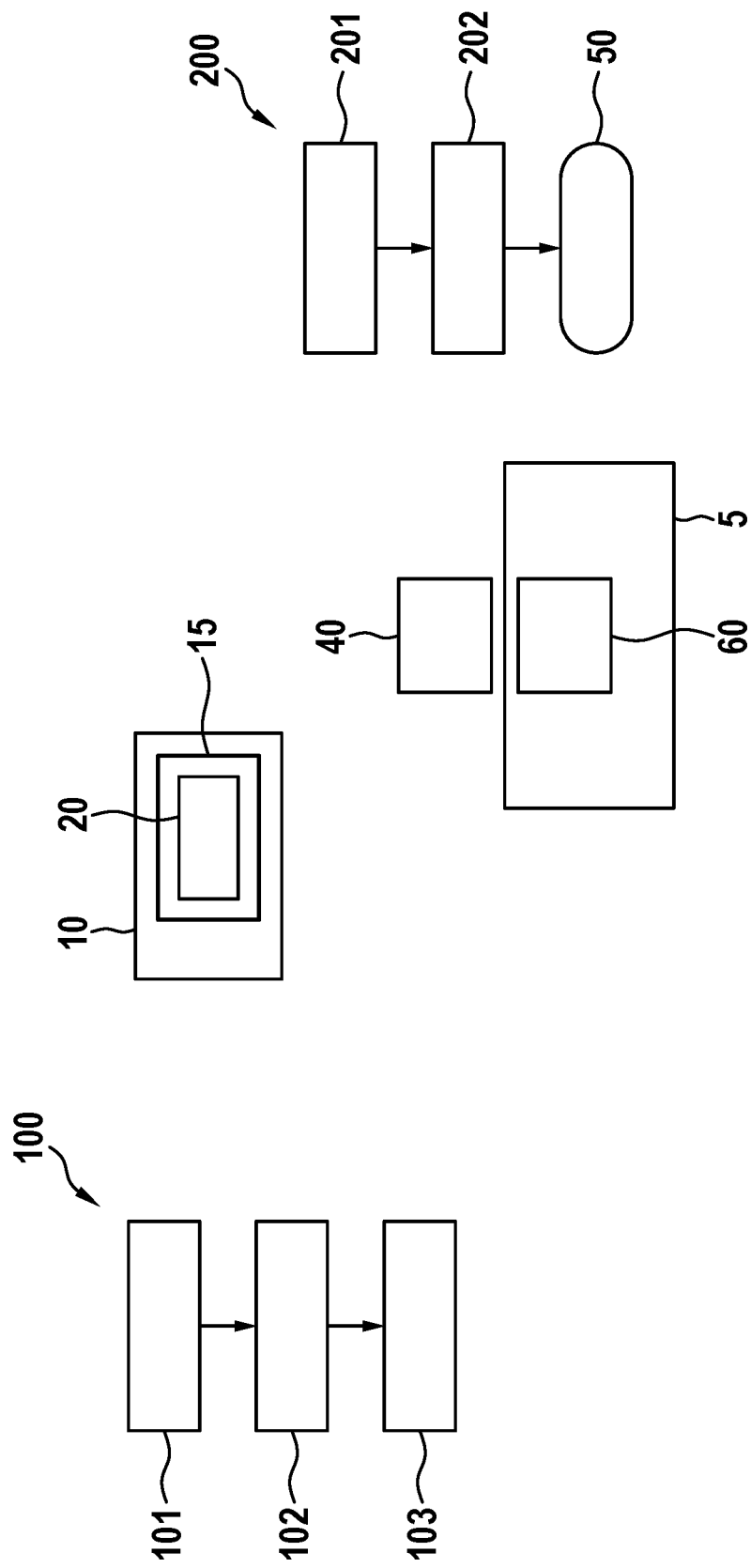


Fig. 2

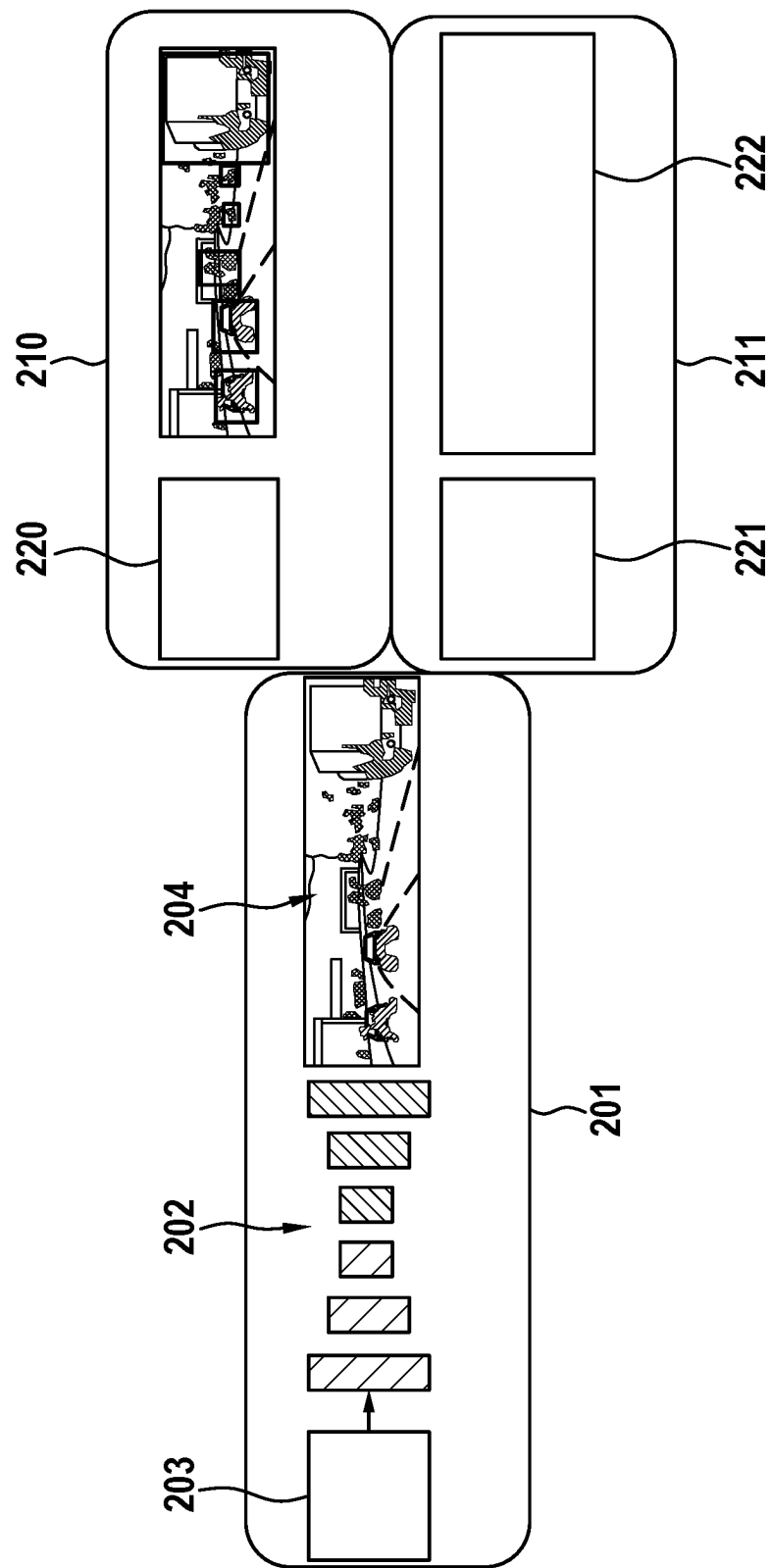


Fig. 3

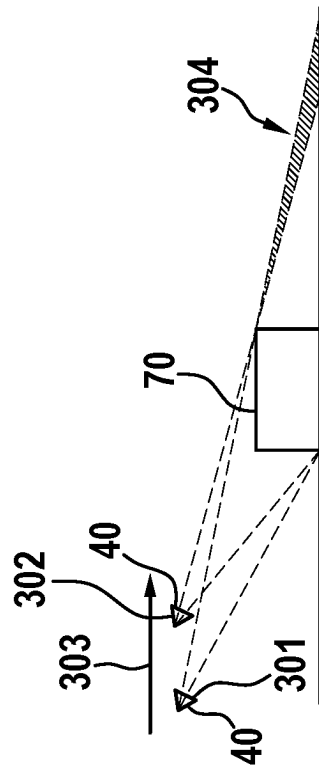


Fig. 4

