(19) Deutsches Patent- und Markenamt

(10) **DE 10 2023 126 519 A1** 2025.04.03

(12) # Offenlegungsschrift

(21) Aktenzeichen: **10 2023 126 519.9**
(22) Anmeldetag: **28.09.2023**
(43) Offenlegungstag: **03.04.2025**

(51) Int Cl.: ***G06V 20/58*** (2022.01)
G06V 10/70 (2022.01)
G06V 10/44 (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

(56) Ermittelter Stand der Technik:

CHEN, Zhenpeng ; LIU, Qianfei ; LIAN, Chenfan: PointLaneNet: Efficient end-to-end CNNs for accurate real-time lane detection. In: 2019 IEEE intelligent vehicles symposium (IV), Paris, France, June 9-12, 2019. S. 2563-2568. – ISBN 978-1-7281-0560-4

QU, Zhan [u.a.]: Focus on local: Detecting lane marker from bottom up via key point. In: IEEE ; CVF: Conference on Computer Vision and Pattern Recognition (CVPR) - 20-25 June 2021 - Nashville, TN, USA, 2021, S. 14117-14125. - ISBN 978-1-6654-4509-2

WANG, Jinsheng [et al.]: A keypoint-based global association network for lane detection. In: 2012 IEEE/CVF conference on computer vision and pattern recognition (CVPR). IEEE, 2022. S. 1382-1391. – ISBN 978-1-6654-6946-3

WANG, Ruihao [et al.]: BEV-LaneDet: a simple and effective 3D lane detection baseline. 2023-03-11. URL: https://arxiv.org/pdf/2210.06006 [abgerufen am 20.09.2024]

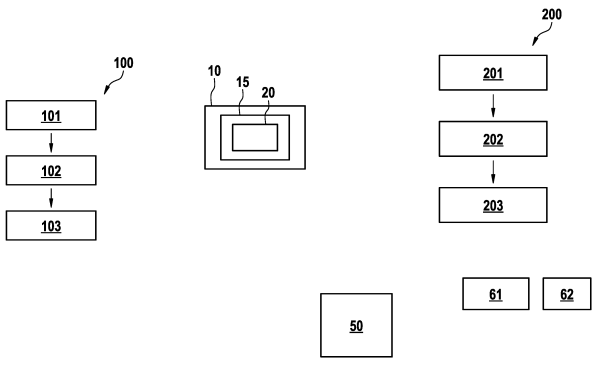
XU, Shenghua [et al.]: RCLane: relay chain prediction for lane detection. In: Computer vision – ECCV 2022 : 17th European conference Tel Aviv, Israel, October 23-27, 2022, Proceedings, Part 38. Cham : Springer, 2022 S. 461-477. – ISBN 978-3-031-19838-0

Rechercheantrag gemäß § 43 PatG ist gestellt.

Die folgenden Angaben sind den vom Anmelder eingereichten Unterlagen entnommen.

(54) Bezeichnung: **Trainingsverfahren für ein Maschinenlern-Modell und Verfahren zur Erkennung von Navigationsobjekten**

(57) Zusammenfassung: Die Erfindung betrifft ein Trainingsverfahren (100) für ein Maschinenlern-Modell (50) zur Erkennung von Navigationsobjekten (42), um die erkannten Navigationsobjekte (42) für eine Orientierung bei einer Navigation eines zumindest teilweise autonomen Roboters (61) oder Fahrzeuges (62) bereitzustellen, umfassend die nachfolgenden Schritte:
- Bereitstellen (101) von wenigstens einer Szenen-Repräsentation (40) wenigstens einer Verkehrsszene (41),
- Bereitstellen (102) von vorgegebenen Linienrepräsentationen (503) zumindest einer Auswahl von Navigationsobjekten (42), die in der wenigstens einen Verkehrsszene (41) vorgesehen sind,
- Trainieren (103) des Maschinenlern-Modells (50) auf Basis der wenigstens einen bereitgestellten Szenen-Repräsentation (40) und der bereitgestellten vorgegebenen Linienrepräsentationen (503) zur Ausgabe von Linienrepräsentationen (503) derjenigen Navigationsobjekte (42), die in der wenigstens einen Verkehrsszene (41) erkannt werden, wobei die Linienrepräsentationen (503) die Navigationsobjekte (42) jeweils in einer diskretisierten Punktedarstellung (303,313) mit einer Menge von Punkten (302) repräsentieren.



Beschreibung

[0001] Die vorliegende Erfindung betrifft ein Trainingsverfahren gemäß der im Oberbegriff des Anspruchs 1 näher definierten Art. Ferner bezieht sich die Erfindung auf ein Verfahren, ein Computerprogramm, eine Vorrichtung, ein Speichermedium sowie ein Maschinenlern-Modell.

Stand der Technik

[0002] Um sich im Rahmen des autonomen Fahrens in der Umgebung, beispielsweise auf Autobahnen, effektiv zu orientieren, ist es häufig erforderlich, die relevanten Navigationsobjekte zu identifizieren. Solche Objekte könnten Fahrbahnmarkierungen sein, die das Verständnis der Verkehrsregeln ermöglichen und die autonome Fahrzeugnavigation erleichtern, um die Spur zu halten. Eine weitere bedeutende Komponente sind Bordsteine, da sie den befahrbaren Bereich begrenzen. Auch Ampelmasten stellen wichtige Objekte dar, die zur Fahrzeuglokalisierung genutzt werden können. Es ist häufig von großer Bedeutung, diese Objekte in dreidimensionaler Form zu erkennen und die Kompatibilität mit unterschiedlichen Auflösungen zu gewährleisten, da dies den technischen Aufwand und die damit verbundenen Kosten für den Einsatz von Fahrzeugen oder Robotern mit Kameras verschiedener Auflösungen reduzieren kann.

[0003] Eine Methode aus dem Stand der Technik wurde in der Veröffentlichung „Zhan Qu et al. Focus on Local: Detection Lane Marker from Bottom Up via Key Point“ vorgestellt. Diese Methode ermöglicht die Erkennung von Fahrbahnmarkierungen im Bildraum. Trotz der Unterstützung für Mehrfachauflösungen kann das Verfahren nur vertikale Fahrbahnmarkierungen erkennen und liefert lediglich 2D-Pixel-Informationen über die Position von Fahrbahnmarkierungen. Der darin offenbarte Dekodierungsmechanismus geht außerdem davon aus, dass alle erkannten Fahrbahnmarkierungen das gesamte Bild durchqueren. Dies ist jedoch nicht der Fall, wenn nicht zusammenhängende Fahrbahnmarkierungen an Straßenkreuzungen vorhanden sind.

Offenbarung der Erfindung

[0004] Gegenstand der Erfindung ist ein Trainingsverfahren mit den Merkmalen des Anspruchs 1, ein Verfahren mit den Merkmalen des Anspruchs 5, ein Computerprogramm mit den Merkmalen des Anspruchs 8, eine Vorrichtung mit den Merkmalen des Anspruchs 9 sowie ein computerlesbares Speichermedium mit den Merkmalen des Anspruchs 10 und 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 Trainingsverfahren beschrieben sind, selbstverständlich auch im Zusammenhang mit dem erfindungsgemäßen Verfahren, dem erfindungsgemäßen Computerprogramm, der erfindungsgemäßen Vorrichtung sowie dem erfindungsgemäßen computerlesbaren Speichermedium und dem erfindungsgemäßen Maschinenlern-Modell, und jeweils umgekehrt, so dass bezüglich der Offenbarung zu den einzelnen Erfindungsaspekten stets wechselseitig Bezug genommen wird bzw. werden kann.

[0005] Gegenstand der Erfindung ist insbesondere ein Trainingsverfahren für ein Maschinenlern-Modell zur Erkennung von Navigationsobjekten, um die erkannten Navigationsobjekte für eine Orientierung bei einer Navigation eines zumindest teilweise autonomen Roboters oder Fahrzeuges bereitzustellen. Hierzu kann bspw. eine Navigationsvorrichtung zum autonomen Fahren bei dem Roboter bzw. Fahrzeug vorgesehen sein, um anhand der erkannten Navigationsobjekte die Fortbewegung des Roboters bzw. Fahrzeuges automatisch zu steuern und bspw. eine Fahrbahnerkennung durchzuführen. Das Trainingsverfahren kann die nachfolgenden Schritte umfassen, welche bevorzugt nacheinander und/oder wiederholt ausgeführt werden:

- Bereitstellen von wenigstens einer Szenen-Repräsentation wenigstens einer Verkehrsszene, wobei die jeweilige Szenen-Repräsentation bspw. als ein Kamerabild ausgeführt sein kann. Es können hierbei auch eine Vielzahl von Szenen-Repräsentationen vorgesehen sein, um ein zuverlässiges Training zu ermöglichen.

- Bereitstellen von vorgegebenen Linienrepräsentationen zumindest einer Auswahl von Navigationsobjekten, die in der wenigstens einen Verkehrsszene vorgesehen sind. Die Linienrepräsentationen können hierzu bspw. manuell durch eine Sichtung der Szenen-Repräsentation vorbereitet worden sein. Auf diese Weise können Trainingsdaten für das Training bereitgestellt werden.

- Trainieren des Maschinenlern-Modells auf Basis dieser Trainingsdaten, d.h. der wenigstens einen bereitgestellten Szenen-Repräsentation und der bereitgestellten vorgegebenen Linienrepräsentationen, zur Ausgabe von Linienrepräsentationen derjenigen Navigationsobjekte, die in der wenigstens einen

Verkehrsszene erkannt werden. Mit anderen Worten wird das Maschinenlern-Modell dafür trainiert, die Navigationsobjekte in der wenigstens einen bereitgestellten Szenen-Repräsentation zu erkennen und in der Form von Linienrepräsentationen auszugeben.

[0006] Die Linienrepräsentationen können dadurch gekennzeichnet sein, dass sie die Navigationsobjekte jeweils in einer diskretisierten Punktedarstellung mit einer Menge von Punkten repräsentieren. Dabei können die Linienrepräsentationen auch jeweils weitere Parameter umfassen, welche z. B. eine explizite Darstellung von Start- und Endpunkten der Menge von Punkten und/oder eine Angabe einer Reihenfolge zur Verbindung der Punkte bereitstellen. Dies ermöglicht es, das jeweilige Navigationsobjekt in der Form einer Linie zu reproduzieren.

[0007] Eine weitere Verbesserung kann erzielt werden, wenn der jeweilige Punkt durch zweidimensionale Koordinaten (anstelle der herkömmlichen Definition nur durch eine Koordinate) definiert wird. Dies ermöglicht die Darstellung der Navigationsobjekte als Linienobjekte beliebiger Form und eine vollständige generische Darstellung der Linie für mehrere Auflösungen.

[0008] Die Erfindung kann den Vorteil haben, dass die korrekte Erkennung und Dekodierung verschiedener Formen von Navigationsobjekten und insbesondere Fahrbahnmarkierungen ermöglicht werden. Damit können die Beschränkungen des Standes der Technik für die Erkennung von nur vertikalen Fahrbahnmarkierungen überwunden und auch die Erkennung von Fahrspuren in 3D - d. h. im dreidimensionalen Raum - und deren Klassifizierung nach ihrem Typ ermöglicht werden (z. B. können Fahrbahnmarkierungen Klassenbezeichnungen wie durchgezogene Fahrspur, Haltelinie oder gestrichelte Fahrspur usw. zugwiesen werden). Es können ferner ggf. nicht nur Fahrbahnmarkierungen, sondern auch weitere Navigationsobjekte wie Bordsteine und/oder Masten, insbesondere Ampelmasten, in 3D erkannt werden und die erforderlichen gewünschten Merkmale für das Objekt von Interesse bereitstellen werden (z. B. kann für Bordsteine zusätzlich die Höhe der Bordsteine bereitgestellt werden).

[0009] Die Erfindung weist mehrere einzigartige Hauptaspekte auf. So stellt sie einen Ansatz für die durchgängige Ausbildung der Erkennung von Navigationsobjekten - insbesondere in der Form von linienartigen Objekten mit beliebigen Formen - bereit. Es kann ferner die Möglichkeit bereitgestellt werden, die Erkennung von Navigationsobjekten, kurz Objekte bezeichnet, direkt in 3D zu erhalten, was die Komplexität von Linienerkennungsansätzen zur Wiederherstellung von 3D-Informationen reduziert. Es kann weiter eine Möglichkeit bereitgestellt werden, die semantische Bedeutung der Objekte zu klassifizieren. Dies erlaubt es, ein einziges neuronales Netzwerk anzuwenden, um mehrere verschiedene Arten von Objekten mit Linienformen zu erkennen. Diese Arten von Objekten können z.B. Fahrbahnmarkierungen, Bordsteine, Masten und dergleichen umfassen.

[0010] Durch die Erfindung kann auch der Aufwand und damit die Kosten für eingebettete Hardware reduziert werden, da der Ansatz sehr effizient ist und gleichzeitig die Funktionalität der aufwendigeren Bildverarbeitungsansätze (semantische Segmentierung und/oder Tiefenabschätzung) beibehält. Darüber hinaus kann vorteilhafterweise jede Linienform der Objekte in 3D mit einem einzigen Ansatz erkannt werden. Zudem ist es möglich, dass jedem erkannten Objekt eine semantische Bedeutung gegeben wird, die für die Navigation und die Lokalisierung von Robotern und/oder Fahrzeugen wichtig ist und hierfür verwendet werden kann. Auch ist es möglich, dass verschiedene Skalierungsvarianten verwendet werden (z.B. nur Fahrbahnmarkierungserkennung, nur Masten oder Bordsteinerkennung oder jede mögliche Kombination dieser Objekterkennungen). Darüber hinaus kann die erfindungsgemäße Lösung ggf. mit verschiedenen Kameraauflösungen verwendet werden, während nur auf eine einzige Auflösung trainiert wird. Das macht es zu einer sehr effizienten Lösung insbesondere für autonome Systeme zum autonomen Fahren.

[0011] Vorzugsweise kann vorgesehen sein, dass die Linienrepräsentationen jeweils wenigstens einen der folgenden Parameter umfassen:

- Eine Konfidenz- und/oder Heatmap zur Schätzung einer Pixelposition der Punkte,

- Eine Klassifikationsmaske zur semantischen Klassifikation des jeweiligen repräsentierten Navigationsobjekts,

- Einen Offset-Parameter einer Zelle einer Heatmap mit einer Offset-Information, um einen Offset zu einer Pixelposition der Punkte der Punktedarstellung anzugeben,

- Eine Tiefeninformation für den jeweiligen Punkt der Punktedarstellung, welche für eine Entfernung des repräsentierten Navigationsobjekts zum Fahrzeug und/oder Roboter spezifisch ist,

- Ein Klassenlabel für den jeweiligen Punkt der Punktedarstellung, um den Punkt semantisch zu klassifizieren.

[0012] Im Gegensatz zu bekannten Lösungen, die nicht in der Lage sind, beliebige linienförmige Objekte zu erkennen, kann die Erfindung diese Einschränkung damit durch die Einführung spezieller Parametrisierungs- und Dekodierungsschemata überwinden und die Erkennung in 3D um eine semantische Klassifizierung erweitern, die es ermöglicht, dasselbe Maschinenlern-Modell und vorzugsweise CNN (engl. Convolutional Neural Network) für die Erkennung verschiedener Navigationsobjekte, vorzugsweise linienförmiger Objekte wie Bordsteinkanten, Pfosten und Fahrbahnmarkierungen, zu verwenden. Dabei können die bereitgestellten vorgegebenen Linienrepräsentationen als Ground-Truth vorgegeben werden und/oder eine Parametrisierung der Linienrepräsentationen, insbesondere eine Festlegung der Parameter, für das Trainieren durch die Ground-Truth vorgegeben werden. Alternativ oder zusätzlich kann die wenigstens eine bereitgestellte Szenen-Repräsentation als ein zumindest oder genau zweidimensionales Kamerabild ausgeführt sein.

[0013] Ferner kann im Rahmen der Erfindung vorgesehen sein, dass jeder der Punkte eine Offset-Information für die Angabe der Reihenfolge zur Verbindung der Punkte umfasst, welche einen Offset zu einem vorangehenden und einen nachfolgenden Punkt der Punktedarstellung angibt, um durch eine Verbindung der Punkte die Linie wiederherzustellen, welche das jeweilige Navigationsobjekt repräsentiert. Dies hat den Vorteil, dass die herkömmliche Beschränkung der Form der Navigationsobjekte überwunden wird.

[0014] Nach einer weiteren Möglichkeit kann vorgesehen sein, dass das Maschinenlern-Modell - vorzugsweise Ende-zu-Ende - trainiert wird, um die erkannten Navigationsobjekte repräsentiert in der diskretisierten Punktedarstellung im Ausgangstensor des Maschinenlern-Modells auszugeben. Dies ermöglicht eine verbesserte Repräsentation der Navigationsobjekte.

[0015] Ebenfalls Gegenstand der Erfindung ist ein Verfahren zur Erkennung von wenigstens einem Navigationsobjekt wenigstens einer Verkehrsszene, um das wenigstens eine erkannte Navigationsobjekt für eine Orientierung bei einer Navigation eines zumindest teilweise autonomen Roboters oder Fahrzeuges zu verwenden. Hierbei können die folgenden Schritte durchgeführt werden, vorzugsweise nacheinander und/oder wiederholt:

- Bereitstellen von wenigstens einer Szenen-Repräsentation der wenigstens einen Verkehrsszene, welche aus einer Kameraerfassung resultiert und/oder das wenigstens eine Navigationsobjekt abbildet, wobei die Erfassung vorzugsweise durch wenigstens eine Kamera des Fahrzeuges bzw. Roboters durchgeführt wird bzw. wurde,

- Durchführen einer Verarbeitung der wenigstens einen Szenen-Repräsentation durch ein Maschinenlern-Modell, vorzugsweise durch die Eingabe der Szenen-Repräsentation als Eingangstensor für das Maschinenlern-Modell, vorzugsweise in der Form eines CNN,

- Erhalten einer Linienrepräsentation des wenigstens einen Navigationsobjektes auf Basis der durchgeführten Verarbeitung, wobei durch die erhaltene Linienrepräsentation das jeweilige Navigationsobjekt in einer diskretisierten Punktedarstellung mit einer Menge von Punkten repräsentiert wird, wobei die Linienrepräsentation weitere Parameter umfassen kann, welche eine explizite Darstellung von Start- und Endpunkten der Menge von Punkten und eine Angabe einer Reihenfolge zur Verbindung der Punkte bereitstellen.

[0016] Es kann ferner der jeweilige Punkt durch zweidimensionale Koordinaten definiert sein.

[0017] Das erfindungsgemäße Verfahren bringt die gleichen Vorteile mit sich, wie sie ausführlich mit Bezug auf ein erfindungsgemäßes Trainingsverfahren beschrieben worden sind. Des Weiteren kann es vorgesehen sein, dass das Maschinenlern-Modell durch ein erfindungsgemäßes Trainingsverfahren trainiert wurde, vorzugsweise um die Linienrepräsentation durch einen Ausgangstensor des Maschinenlern-Modells nach der Verarbeitung der wenigstens einen Szenen-Repräsentation auszugeben. Die diskretisierte Ausführung der Punktedarstellung kann sich darauf beziehen, dass eine Linie durch mehrere einzelne Punkte repräsentiert wird. Das Maschinenlern-Modell kann vorteilhafterweise als ein CNN ausgebildet sein.

[0018] Vorteilhafterweise kann bei der Erfindung vorgesehen sein, dass wenigstens einer der nachfolgenden Schritte zur Dekodierung des wenigstens einen Navigationsobjektes anhand der erhaltenen Linienrepräsentation durchgeführt wird:

- Ermitteln einer geschätzten Pixelposition anhand wenigstens einer Konfidenz- oder Heatmap aus der erhaltenen Linienrepräsentation,

- Ermitteln einer im Vergleich zur geschätzten Pixelposition genaueren Pixelposition anhand wenigstens einer Offset-Information aus der erhaltenen Linienrepräsentation,

- Durchführen einer Reproduktion der Start- und Endpunkte der Punktedarstellung anhand der expliziten Darstellung der Start- und Endpunkte aus der erhaltenen Linienrepräsentation,

- Durchführen einer iterativen Reproduktion einer vollständigen Linie anhand der Angabe der Reihenfolge zur Verbindung der Punkte aus der erhaltenen Linienrepräsentation,

- Durchführen einer Reproduktion einer dreidimensionalen Position der Punkte anhand einer Tiefeninformation aus der erhaltenen Linienrepräsentation,

- Durchführen einer Reproduktion einer semantischen Klasse des Navigationsobjektes anhand eines Klassenlabels für den jeweiligen Punkt aus der erhaltenen Linien repräsentation.

[0019] Ebenfalls Gegenstand der Erfindung ist ein Maschinenlern-Modell, welches durch ein erfindungsgemäßes Trainingsverfahren trainiert wurde.

[0020] Ebenfalls Gegenstand der Erfindung ist ein Computerprogramm, insbesondere Computerprogrammprodukt, umfassend Befehle, die bei der Ausführung des Computerprogrammes durch einen Computer diesen veranlassen, wenigstens eines der erfindungsgemäßen Verfahren auszuführen. Damit bringt das erfindungsgemäße Computerprogramm die gleichen Vorteile mit sich, wie sie ausführlich mit Bezug auf ein erfindungsgemäßes Verfahren beschrieben worden sind.

[0021] Ebenfalls Gegenstand der Erfindung ist eine Vorrichtung zur Datenverarbeitung, die eingerichtet ist, wenigstens eines der erfindungsgemäßen 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.

[0022] 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, wenigstens eines der erfindungsgemäßen 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.

[0023] Darüber hinaus kann das jeweilige erfindungsgemäße (Trainings-) Verfahren auch als ein computerimplementiertes Verfahren ausgeführt sein.

[0024] 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 sowie eines Computerprogramms gemäß Ausführungsbeispielen der Erfindung.

Fig. 2 eine schematische Darstellung eines Maschinenlern-Modells gemäß Ausführungsbeispielen der Erfindung.

Fig. 3 eine schematische Darstellung eines Ausgangstensors des Maschinenlern-Modells gemäß Ausführungsbeispielen der Erfindung.

Fig. 4 eine schematische Darstellung einer Vertrauens- und Heatmap Feature Map.

Fig. 5 Ein Beispiel für die Darstellung von Start- und Endpunkten.

Fig. 6 Eine Visualisierung von verschiedenen Linienformen.

Fig. 7 Eine beispielhafte Dekodierung der 3D-Linienobjekte mit ihren Klassenbezeichnungen.

Fig. 8 Ein Beispiel für eine 3D Rekonstruktion einer Linie.

Fig. 9 Beispielhafte Rohausgaben des neuronalen Netzes bzw. Convolutional Neural Network (CNN) vor der Dekodierung.

[0025] In **Fig.** 1 sind ein Verfahren 100, eine Vorrichtung 10, ein Speichermedium 15 sowie ein Computerprogramm 20 und ein Maschinenlern-Modell 50 gemäß Ausführungsbeispielen der Erfindung schematisch dargestellt.

[0026] **Fig.** 1 veranschaulicht gemäß Ausführungsbeispielen der Erfindung ein Trainingsverfahren 100 für ein Maschinenlern-Modell 50 zur Erkennung von Navigationsobjekten 42, um die erkannten Navigationsobjekte 42 für eine Orientierung bei einer Navigation eines zumindest teilweise autonomen Roboters 61 oder Fahrzeuges 62 bereitzustellen. Hierbei kann gemäß einem ersten Trainingsschritt 101 ein Bereitstellen von wenigstens einer Szenen-Repräsentation 40 wenigstens einer Verkehrsszene 41 erfolgen. Weiter kann gemäß einem zweiten Trainingsschritt 102 ein Bereitstellen von vorgegebenen Linienrepräsentationen 503 zumindest einer Auswahl von Navigationsobjekten 42 erfolgen, wobei die Navigationsobjekte 42 in der wenigstens einen Verkehrsszene 41 vorgesehen sind. Gemäß einem dritten Trainingsschritt 103 kann das Training des Maschinenlern-Modells 50 auf Basis der wenigstens einen bereitgestellten Szenen-Repräsentation 40 und der bereitgestellten vorgegebenen Linienrepräsentationen 503 zur Ausgabe von Linienrepräsentationen 503 derjenigen Navigationsobjekte 42, die in der wenigstens einen Verkehrsszene 41 erkannt werden, erfolgen. Dabei können die Linienrepräsentationen 503 die Navigationsobjekte 42 jeweils in einer diskretisierten Punktedarstellung 303,313 mit einer Menge von Punkten 302 repräsentieren. Des Weiteren können die Linienrepräsentationen 503 jeweils weitere Parameter umfassen, welche eine explizite Darstellung von Start- und Endpunkten 501 der Menge von Punkten 302 und eine Angabe einer Reihenfolge zur Verbindung der Punkte 302 bereitstellen, um das jeweilige Navigationsobjekt 42 in der Form einer Linie 301 zu reproduzieren, wobei der jeweilige Punkt 302 durch zweidimensionale Koordinaten x,y definiert wird.

[0027] Ebenfalls ist in **Fig.** 1 ein Verfahren 200 zur Erkennung von wenigstens einem Navigationsobjekt 42 wenigstens einer Verkehrsszene 41 dargestellt, um das wenigstens eine erkannte Navigationsobjekt 42 für eine Orientierung bei einer Navigation eines zumindest teilweise autonomen Roboters 61 oder Fahrzeuges 62 zu verwenden. Hierbei ist gemäß einem ersten Verfahrensschritt 201 ein Bereitstellen von wenigstens einer Szenen-Repräsentation 40 der wenigstens einen Verkehrsszene 41 vorgesehen. Weiter ist gemäß einem zweiten Verfahrensschritt 202 ein Durchführen einer Verarbeitung der wenigstens einen Szenen-Repräsentation 40 durch ein Maschinenlern-Modell 50 möglich. Gemäß einem dritten Verfahrensschritt 203 kann eine Linienrepräsentation 503 des wenigstens einen Navigationsobjektes 42 auf Basis der durchgeführten Verarbeitung erhalten werden, wodurch das jeweilige Navigationsobjekt 42 in einer diskretisierten Punktedarstellung 303,313 mit einer Menge von Punkten 302 repräsentiert wird.

[0028] In selbstfahrenden Fahrzeugsystemen und/oder in der Robotik ist oft eine vollständige 3D-Darstellung notwendig, um eine zuverlässige und sichere Navigation von linienförmigen Objekten (Fahrbahnmarkierungen, Bordsteinkanten, Pfosten) zu ermöglichen. Ausführungsbeispiele der Erfindung ermöglichen diese Funktionalität für ein kamerabasiertes System und vorzugsweise auf eingebetteten Edge Devices.

[0029] Ein Überblick des vorgeschlagenen Ansatzes gemäß Ausführungsvarianten der Erfindung ist in **Fig.** 2 dargestellt. Beispielhaft wird hier ein CNN 50 genutzt, d. h. ein Convolutional Neural Network, welches mehrere Schichten von Faltungsoperationen, Pooling und Aktivierungsfunktionen aufweisen kann. Das CNN kann dabei ein Eingangsbild 40 erhalten und 3D-Spurmarkierungen, 3D-Pfosten und 3D-Bordsteine (s. Linien 201, 202, 203) ausgeben. Es kann auch die detaillierte semantische Klasse des erkannten Objekts ausgeben werden (z.B. kann es im Falle von Fahrbahnmarkierungen den Typ der Fahrbahn ausgeben, wie durchgezogene Fahrbahn, gestrichelte Fahrbahn, Haltelinie usw.).

[0030] In **Fig.** 2 ist schematisch dargestellt, dass das neuronale Netz 50 ein Eingangsbild 50 erhält und die linienförmige Objekterkennung und deren Position in 3D-Koordinaten ausgeben kann. Die Mastenerkennung ist dabei über die Linien 201, die Bordsteinerkennung über die Linien 202 und die Fahrbahnmarkierungserkennung über die Linien 203 veranschaulicht. Der Ansatz kann in verschiedenen Skalierungsvarianten (z.B. nur Erkennung von Fahrbahnmarkierungen, nur Erkennung von Masten, nur Erkennung von Bordsteinen) oder in jeder möglichen Kombination der Objekterkennungen implementiert werden.

[0031] Anhand der nachfolgend angegebenen Schritte werden weitere Einzelheiten von Ausführungsbeispielen der Erfindung näher beschrieben. Zunächst kann gemäß einem ersten Schritt ein Training eines voll gefalteten CNN erfolgen. Hierzu kann jedes linienförmige Objekt als eine Menge von Punkten dargestellt werden (siehe **Fig.** 3) mit einer binären Klassifizierungsmaske für Konfidenz 401 und/oder Heatmap 402 (s.

Fig. 4) und einem zusätzlichen Offset (Δx, Δy) innerhalb der Heatmap-Zelle, um eine Subpixelgenauigkeit zu erreichen. Jeder Punkt kann zusätzlich Informationen über die Tiefe und die Klassenbezeichnung erhalten. Diese Informationen können in der Kanaldimension des Ausgangstensors des neuronalen Netzes hinzugefügt werden. Weiter kann jeder Punkt in der Zeile die Offsets zum Vorherigen (Δx, Δy) und zum Nächsten (Δx, Δy) sowie zum vorherigen Punkt in der Zeile und zum nächsten Punkt in der Zeile liefern. Dies kann vorgesehen sein, um alle Punkte in der richtigen Reihenfolge zu verbinden und die vollständige Linie wiederherzustellen. Darüber hinaus kann der Endpunkt und der Startpunkt eine Verbindung zu sich selbst haben, um zu signalisieren, dass die Linie hier beginnt/endet (s. **Fig.** 5, wobei die Verbindungen durch Pfeile dargestellt sind). Die beschriebene Parametrisierung erlaubt es, die Linien als Objekte beliebiger Form darzustellen (**Fig.** 6). In einem weitere Schritt kann in der Inferenzzeit für den resultierenden Ausgangstensor ein spezieller Dekodierungsalgorithmus vorgesehen werden, um alle Fahrspuren in 3D zu dekodieren (**Fig.** 7, **Fig.** 8).

[0032] In **Fig.** 3 ist beispielhaft ein linienförmiges Objekt im Ausgangstensor des CNNs dargestellt. Die Linie 301 kann als eine Menge von verbundenen Punkten 302 dargestellt werden. Jede Gitterzelle im Bild 303 entspricht einer Pixelposition im Ausgangstensor. Es kann vorgesehen sein, dass die Punkte als binäre Klassifizierung erkannt werden. Die Zellen 304, in denen sich Punkte 302 befinden, können als 1,0 klassifiziert werden, ansonsten als 0,0 (übrige Zellen im Bild 303). Der Ausgabetensor ist in der Regel kleiner als die ursprüngliche Auflösung, so dass zur Wiederherstellung der feinkörnigen Punktposition für jede Zelle, in die der Linienpunkt fällt, auch Δx- und Δy-Offsets bereitgestellt werden, um die Subpixelgenauigkeit der Punktposition wiederherzustellen.

[0033] Im Bild 313 in **Fig.** 3 ist die Punktverbindung als Offset zwischen dem aktuellen Punkt und Previous (Δx, Δy) und dem aktuellen Punkt und Next(Δx, Δy) dargestellt. Jeder Punkt 312 im Bild 313 hat diesen Versatz zu seinen Nachbarn. Dies wird verwendet, um angesichts der Reihenfolge der Verbindungen zwischen den Punkten eine vollständige Linie zu erhalten. Jeder Punkt 312 kann zusätzlich Tiefeninformationen und eine Klassenbezeichnung in der Kanaldimension des Ausgangstensors erhalten. Dies wird verwendet, um die Positionen der Punkte in 3D und ihre semantische Bedeutung wiederherzustellen.

[0034] Es kann vorgesehen sein, dass jede Linie als Punktesatz diskretisiert und ein räumliches Pixelpositionsgitter als Koordinatengitter verwendet wird, um sie zu lokalisieren. Der Lokalisierungsmechanismus kann als binäres Klassifikationsproblem implementiert werden (s. **Fig.** 3). Jede Gitterzelle kann ein Pixel des Ausgangstensors sein. Die Gitterzellen (Pixel), die Linienpunkte enthalten, können als Klasse - 1 klassifiziert werden, andernfalls - 0. Die binäre Klassifizierung kann als Konfidenzkarte 401 oder Heatmap 402 implementiert werden (siehe **Fig.** 4). In **Fig.** 4 sind dabei zwei Beispiele für die binäre Klassifizierung der Punkte im Ausgangstensor dargestellt. In der Darstellung 401 ist die Konfidenzkarte zu sehen. Pixel, in denen sich Punkte befinden, werden als 1,0 im Bild dargestellt, andernfalls als 0. In der Darstellung 402 ist die Heatmap erkennbar. Jeder Punkt wird als 1,0 dargestellt und weiter wird eine kleine Gaußsche Unschärfe in der Nähe seiner Spitze vorgesehen, andernfalls wird der Punkt als 0,0 dargestellt. Diese Darstellung ermöglicht es, die Menge an Informationen für das Netzwerktraining für die positiven Beispiele zu erhöhen und kann in einigen Fällen robuster sein als der Vertrauensansatz.

[0035] Im Falle der Konfidenzkarte kann der binäre Kreuzentropieverlust wie folgt eingesetzt werden:

$$L_{confidence} = \frac{-1}{N}\sum_{i=1}^{N} y_i \log(p_i) + (1-y_i)\,log\,(1-p_i) \qquad (1)$$

[0036] Hier steht y_i für Grundwahrheitsbezeichnungen, p_i für die Vorhersage des CNN und N für die Anzahl der Proben. Im Falle der Heatmap-Darstellung kann eine Strafe angewendet werden, die den fokalen Verlust reduziert:

$$L_{healtmap} = \frac{-1}{N}\sum_{i=1}^{N} Y_i\,(1-P_i)\,log\,(P_i) + (1-Y_i)^{\beta}\,P_i^{\alpha}\,log\,(1-P_i) \qquad (2)$$

[0037] Hier steht Y_i für die Grundwahrheits-Heatmap, P_i für die vorhergesagte Heatmap, β und α sind weitere Parameter (es kann bspw. α =2, β = 4 eingestellt werden) und N die Anzahl der Proben. Die Y_i Heatmap für die binäre Grundwahrheitsmaske y_i kann mit einem Gauß-Kernel erstellt werden. Wenn sich zwei Gauß-Kerne der gleichen Punkte überschneiden, kann das elementweise Maximum genommen werden.

[0038] Um den aktuellen Punkt mit Subpixel-Genauigkeit zu berechnen, kann folgendes angewendet werden:

$$P_{current} = p + [\Delta x_{current} \ \Delta y_{current}] \qquad (3)$$

[0039] Dabei ist $p_{current}$ - die Position des aktuellen Punktes mit Subpixelgenauigkeit, p - die grobe Position des Punktes aus der Vertrauens-/Heatmap (vtl. die Zellen in **Fig.** 3), $\Delta x_{current}$ - Versatz zur Zelle in x-Richtung (vgl. **Fig.** 3), $\Delta y_{current}$ - Versatz zur Zelle in y-Richtung (vgl. ebenfalls **Fig.** 3).

[0040] Zur Berechnung des nächsten und des vorherigen Punktes kann diese Methode angewendet werden:

$$p_{next} = p_{current} + [\Delta x_{next} \ \Delta y_{next}] \qquad (4)$$

sowie

$$p_{previous} = p_{current} + [\Delta x_{next} \ \Delta y_{next}] \qquad (5)$$

[0041] Dabei sind $p_{next,}$ $p_{previous}$ - nächster und vorheriger Punkt, $[\Delta x_{next} \ \Delta y_{next}]$ - Offsets vom aktuellen Punkt zum nächsten Punkt. Weiter sind $[\Delta x_{previous} \ \Delta y_{previous}]$ Offsets zwischen aktuellem Punkt und vorherigem Punkt. Als Verlustfunktion für die Offsets kann der L1-Verlust verwendet werden:

$$L_{offsets} = \frac{1}{N}\left(\sum_{i=1}^{N} \left| p_{current}(i) - p_{current_{gt}} \right| + \left| p_{next} - p_{next_{gt}} \right| + \left| p_{previous} - p_{previous_{gt}} \right| \right) \qquad (6)$$

[0042] Hier sind $p_{current\_gt}$, p_{next_gt}, $p_{previous\_gt}$ die Standorte der Bodenwahrnehmungspunkte, die Verlustwerte geteilt durch die Anzahl der Punkte N, um sie zu normalisieren.

[0043] Für die Berechnung der Tiefeninformation gibt es zwei Möglichkeiten: Es kann die Tiefe d als Disparität dargestellt werden:

$$disparity = 1 / d \qquad (7)$$

[0044] Durch diese Parametrisierung wird die Bedeutung der nächstgelegenen Punkte verstärkt. Es kann ein l1 Verlust angewendet werden:

$$L_{depth} = \sum_{i=1}^{N} \left\| disparity_{gt}(i) - disparity_{pd}(i) \right\|_1 * M_{gt}(i) \qquad (8)$$

[0045] Hier ist $disparity_{gt}$ die Grundwahrheit, $disparity_{pd}$ die CNN-Vorhersage der Disparität und $M_{gt}(i)$ die Vertrauens-Grundwahrheits-Maske.

[0046] Die zweite Möglichkeit liegt darin, die logarithmische Tiefe auf folgende Weise direkt vorherzusagen:

$$L_{depth} = \sum_{i=1}^{N} \left\| log\left(d_{gt}\right)(i) - log\left(d_{pd}\right)(i) \right\|_1 * M_{gt}(i) \qquad (9)$$

[0047] Dabei ist d_{gt}- die wahre Tiefe und log (d_{pd}) die CNN-Vorhersage der log-Tiefe

$$L_{classes} = \sum_{i=1}^{n} categorical\_cross\_entropy\left(class_{gt}(i), class_{pd}(i)\right) * M_{gt}(i) \qquad (10)$$

[0048] Dabei ist $class_{gt}$ die Grundwahrheit $class_{pd}$ CNN-Vorhersage der Klassenbezeichnungen. Der Gesamtverlust kann wie folgt berechnet werden:

$$L = w1 * L_{classes} + w2 * L_{offsets} + w3 * L_{confidence/heatmap} + w4 * L_{depth} \qquad (11)$$

[0049] Es kann in der Inferenzzeit für den resultierenden Ausgangstensor ein Dekodieralgorithmus vorgesehen sein, um alle Spuren in 3D zu dekodieren (vgl. **Fig.** 7 und 8). In **Fig.** 7 ist der Algorithmus mit weiteren Einzelheiten dargestellt. Gemäß einem ersten Schritt 701 erfolgt hierbei die Wiederherstellung der groben Pixelposition aus der Konfidenz- oder Heatmap. Im Falle der Konfidenz kann eine Sigmoid-Aktivierung angewandt werden, um die Logit-Werte zwischen 0 und 1 zu skalieren, und wenn die Wahrscheinlichkeit größer oder gleich dem definierten Schwellenwert ist (z. B. 0,5), kann diese Zelle als Standort des Lane-Punktes

betrachtet werden, andernfalls nicht. Im Falle einer Heatmap können die Aktivierungen sigmoid-skaliert und dann eine 3x3 Max-Pooling-Operation angewandt werden und die ursprüngliche Heatmap mit der maxpooled Heatmap verglichen werden. Die Werte, die kleiner als die „maxpooled“ Heatmap sind, können unterdrückt werden, so dass man nur die Spitzen der Heatmap erhält, bei denen sich die Punkte der Spur befinden. Dann können für jeden Peak nur die Werte ermittelt werden, die größer oder gleich dem vorgegebenen Schwellenwert (z. B. 0,5) sind. Das Ergebnis von Schritt 1 kann eine Menge von Schlüsselpunkten (p1, p2,..., pn) sein, die die grobe Lage der Punkte aller Linien angeben.

[0050] Gemäß einem zweiten Schritt 702 kann die Gewinnung der verfeinerten Position jedes Schlüsselpunktes $p_{current}$ und seiner benachbarten nächsten p_{next} und $p_{previous}$ Punkte durch Hinzufügen entsprechender Offsets zu den Schlüsselpunkten aus dem ersten Schritt 701 erfolgen. Gemäß einem dritten Schritt 703 kann eine Datenassoziation zwischen dem aktuellen Keypoint und seinen Nachbarpunkten konstruiert werden. Für jeden Keypoint können die nächsten p_{next} und $p_{previous}$ Punkte berechnet werden. Diese berechneten Punkte können verwendet werden, um die nächsten Punkte in der $p_{current}$-Punktmenge zu finden. Es kann ferner eine Datenassoziationsschwelle in (Δx, Δy) eingeführt werden. Wenn der Abstand innerhalb des Datenassoziationsradius liegt, kann diese Datenassoziation als gültig betrachtet werden, andernfalls nicht.

[0051] Gemäß einem vierten Schritt 704 kann die Wiederherstellung der Start- und Endpunkte vorgesehen sein. Start- und Endpunkte können die $p_{current}$ Punkte, bei denen einer der Datenassoziations-Deltas entweder zum nächsten oder zum vorherigen Punkt gleich Null ist (Assoziation zu sich selbst).

[0052] Gemäß einem fünften Schritt 705 können für jeden Startpunkt iterativ die gesamte Fahrspur entsprechend der Assoziationsbeziehung rekonstruiert werden, wobei alle mit diesem Startpunkt assoziierten Punkte und die folgenden Punkte mit derselben Linie verbunden und als globale Kurve betrachtet werden. Die Kurvenrekonstruktion kann beendet werden, wenn entweder der Endpunkt erreicht wird oder wenn eine falsche Zuordnung erfolgt.

[0053] Gemäß einem sechsten Schritt 706 kann zur Wiederherstellung der 3D-Position für jede Kurve die Tiefe und die Kameraeigenschaft verwendet werden, um die 2D-Punkte der Kurve in den 3D-Raum zu projizieren. Die Tiefe kann im Falle der Disparitätsparametrisierung wie folgt wiederhergestellt werden:

$$d = 1 / disparity \qquad (12)$$

und im Falle einer logarithmischen Parametrisierung:

$$depth(i) = \frac{1.0}{\frac{disparity(i)}{depth_{min}} + \frac{1.0}{depth_{max}}} \qquad (13)$$

[0054] Dabei sind $depth_{min}$ und $depth_{max}$ skalare Werte des erwarteten minimalen und maximalen Tiefenbereichs in Metern und i der gegebene Punktindex g.

[0055] Im siebten Schritt 707 kann, um die semantische Klasse zu ermitteln, argmax auf den Klassentensor angewandt werden und damit die Bezeichnung für jeden Punkt der Kurve erhalten werden:

$$class(i) = argmax_c\left(softmax_c\left(label\_tensor(i)\right)\right) \qquad (14)$$

[0056] Hierbei sind $argmax_c$ und $softmax_c$ argmax und softmax Operationen über Kanaldimensionen.

[0057] In **Fig.** 8 ist eine beispielhafte 3D-Rekonstruktion einer Linie 801 und der Pole 802 dargestellt. Die Kamera ist mit 810 gekennzeichnet. Darüber hinaus sind die Label „Dashed Line“ 805 und „Solid Line“ 806 angegeben.

[0058] In **Fig.** 9 sind beispielhafte Rohausgaben 900 des neuronalen Netzes/Convolutional Neural Network (CNN) 50 vor der Dekodierung dargestellt. Das CNN 50 kann dabei den Tensor 900 der Form [8+Anzahl_Klassen, Ausgabe_Höhe, Ausgabe_Breite] ausgeben. Der erste Kanal 901 entspricht bspw. der Konfidenz oder Heatmap, die nächsten beiden 902 aktuellen (Δx, Δy) Offsets für den aktuellen Punkt, die nächsten beiden 903 Offsets für den nächsten Punkt (Δx, Δy) und die nächsten beiden 904 Offsets für den

vorherigen Punkt (Δx, Δy). Ein Kanal 905 kann der Tiefe und der letzte Kanal 906 der Anzahl der Klassenlabels für die Klassifizierung entsprechen.

[0059] 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.

Patentansprüche

1. Trainingsverfahren (100) für ein Maschinenlern-Modell (50) zur Erkennung von Navigationsobjekten (42), um die erkannten Navigationsobjekte (42) für eine Orientierung bei einer Navigation eines zumindest teilweise autonomen Roboters (61) oder Fahrzeuges (62) bereitzustellen, umfassend die nachfolgenden Schritte:
- Bereitstellen (101) von wenigstens einer Szenen-Repräsentation (40) wenigstens einer Verkehrsszene (41),
- Bereitstellen (102) von vorgegebenen Linienrepräsentationen (503) zumindest einer Auswahl von Navigationsobjekten (42), die in der wenigstens einen Verkehrsszene (41) vorgesehen sind,
- Trainieren (103) des Maschinenlern-Modells (50) auf Basis der wenigstens einen bereitgestellten Szenen-Repräsentation (40) und der bereitgestellten vorgegebenen Linienrepräsentationen (503) zur Ausgabe von Linienrepräsentationen (503) derjenigen Navigationsobjekte (42), die in der wenigstens einen Verkehrsszene (41) erkannt werden,
wobei die Linienrepräsentationen (503) die Navigationsobjekte (42) jeweils in einer diskretisierten Punktedarstellung (303,313) mit einer Menge von Punkten (302) repräsentieren,
dadurch gekennzeichnet, dass die Linienrepräsentationen (503) jeweils weitere Parameter umfassen, welche eine explizite Darstellung von Start- und Endpunkten (501) der Menge von Punkten (302) und eine Angabe einer Reihenfolge zur Verbindung der Punkte (302) bereitstellen, um das jeweilige Navigationsobjekt (42) in der Form einer Linie (301) zu reproduzieren, wobei der jeweilige Punkt (302) durch zweidimensionale Koordinaten (x,y) definiert wird.

2. Trainingsverfahren (100) nach Anspruch 1,
dadurch gekennzeichnet,
dass die Linienrepräsentationen (503) ferner jeweils wenigstens einen der folgenden Parameter umfassen:
- Eine Konfidenz- und/oder Heatmap zur Schätzung einer Pixelposition der Punkte (302),
- Eine Klassifikationsmaske zur semantischen Klassifikation des jeweiligen repräsentierten Navigationsobjekts (42),
- Einen Offset-Parameter einer Zelle einer Heatmap mit einer Offset-Information, um einen Offset zu einer Pixelposition der Punkte (302) der Punktedarstellung (303,313) anzugeben,
- Eine Tiefeninformation für den jeweiligen Punkt (302) der Punktedarstellung (303,313), welche für eine Entfernung des repräsentierten Navigationsobjekts (42) zum Fahrzeug (62) und/oder Roboter (61) spezifisch ist,
- Ein Klassenlabel für den jeweiligen Punkt (302) der Punktedarstellung (303,313), um den Punkt (302) semantisch zu klassifizieren,
wobei die bereitgestellten vorgegebenen Linienrepräsentationen (503) als Ground-Truth vorgegeben werden und/oder eine Parametrisierung der Linienrepräsentationen (503), insbesondere eine Festlegung der Parameter, für das Trainieren (103) durch die Ground-Truth vorgegeben wird, wobei die wenigstens eine bereitgestellte Szenen-Repräsentation (40) als ein zumindest oder genau zweidimensionales Kamerabild ausgeführt ist.

3. Trainingsverfahren (100) nach Anspruch 1 oder 2, **dadurch gekennzeichnet**, dass jeder der Punkte (302) eine Offset-Information für die Angabe der Reihenfolge zur Verbindung der Punkte (302) umfasst, welche einen Offset zu einem vorangehenden und einen nachfolgenden Punkt (302) der Punktedarstellung (303,313) angibt, um durch eine Verbindung der Punkte (302) (302) die Linie (301) wiederherzustellen, welche das jeweilige Navigationsobjekt (42) repräsentiert.

4. Trainingsverfahren (100) nach einem der vorhergehenden Ansprüche, **dadurch gekennzeichnet**, dass das Maschinenlern-Modell (50) Ende-zu-Ende trainiert wird, um die erkannten Navigationsobjekte (42) repräsentiert in der diskretisierten Punktedarstellung (303,313) im Ausgangstensor des Maschinenlern-Modells (50) auszugeben.

5. Verfahren (200) zur Erkennung von wenigstens einem Navigationsobjekt (42) wenigstens einer Verkehrsszene (41), um das wenigstens eine erkannte Navigationsobjekt (42) für eine Orientierung bei einer Navigation eines zumindest teilweise autonomen Roboters (61) oder Fahrzeuges (62) zu verwenden, umfassend die nachfolgenden Schritte:
- Bereitstellen (201) von wenigstens einer Szenen-Repräsentation (40) der wenigstens einen Verkehrsszene (41), welche aus einer Kameraerfassung resultiert und das wenigstens eine Navigationsobjekt (42) abbildet,
- Durchführen (202) einer Verarbeitung der wenigstens einen Szenen-Repräsentation (40) durch ein Maschinenlern-Modell (50),
- Erhalten (203) einer Linienrepräsentation (503) des wenigstens einen Navigationsobjektes (42) auf Basis der durchgeführten Verarbeitung, durch welche das jeweilige Navigationsobjekt (42) in einer diskretisierten Punktedarstellung (303,313) mit einer Menge von Punkten (302) repräsentiert wird, **dadurch gekennzeichnet**, dass die Linienrepräsentation (503) weitere Parameter umfasst, welche eine explizite Darstellung von Start- und Endpunkten (501) der Menge von Punkten (302) und eine Angabe einer Reihenfolge zur Verbindung der Punkte (302) bereitstellen, wobei der jeweilige Punkt (302) durch zweidimensionale Koordinaten (x,y) definiert ist.

6. Verfahren (200) nach Anspruch 5, **dadurch gekennzeichnet**, dass die nachfolgenden Schritte zur Dekodierung des wenigstens einen Navigationsobjektes (42) anhand der erhaltenen Linienrepräsentation (503) durchgeführt werden:
- Ermitteln einer geschätzten Pixelposition anhand wenigstens einer Konfidenz- oder Heatmap aus der erhaltenen Linienrepräsentation (503),
- Ermitteln einer im Vergleich zur geschätzten Pixelposition genaueren Pixelposition anhand wenigstens einer Offset-Information aus der erhaltenen Linienrepräsentation (503),
- Durchführen einer Reproduktion der Start- und Endpunkte (501) der Punktedarstellung (303,313) anhand der expliziten Darstellung der Start- und Endpunkte (501) aus der erhaltenen Linienrepräsentation (503),
- Durchführen einer iterativen Reproduktion einer vollständigen Linie (301) anhand der Angabe der Reihenfolge zur Verbindung der Punkte (302) aus der erhaltenen Linienrepräsentation (503),
- Durchführen einer Reproduktion einer dreidimensionalen Position der Punkte anhand einer Tiefeninformation aus der erhaltenen Linienrepräsentation (503),
- Durchführen einer Reproduktion einer semantischen Klasse des Navigationsobjektes (42) anhand eines Klassenlabels für den jeweiligen Punkt aus der erhaltenen Linienrepräsentation (503).

7. Verfahren (200) nach Anspruch 5 oder 6, **dadurch gekennzeichnet**, dass das Maschinenlern-Modell (50) durch ein Trainingsverfahren (100) nach einem der Ansprüche 1 bis 4 trainiert wurde, um die Linienrepräsentation (503) durch einen Ausgangstensor des Maschinenlern-Modells (50) nach der Verarbeitung der wenigstens einen Szenen-Repräsentation (40) auszugeben.

8. Computerprogramm (20), umfassend Befehle, die bei der Ausführung des Computerprogramms (20) durch einen Computer (10) diesen veranlassen, das Trainingsverfahren (100) nach einem der Ansprüche 1 bis 4 und/oder das Verfahren (200) nach einem der Ansprüche 5 bis 7 auszuführen.

9. Vorrichtung (10) zur Datenverarbeitung, die eingerichtet ist, das Trainingsverfahren (100) nach einem der Ansprüche 1 bis 4 und/oder das Verfahren (200) nach einem der Ansprüche 5 bis 7 auszuführen.

10. Computerlesbares Speichermedium (15), umfassend Befehle, die bei der Ausführung durch einen Computer (10) diesen veranlassen, die Schritte des Trainingsverfahrens (100) nach einem der Ansprüche 1 bis 4 und/oder des Verfahrens (200) nach einem der Ansprüche 5 bis 7 auszuführen.

11. Maschinenlern-Modell (50), trainiert durch ein Trainingsverfahren (100) nach einem der Ansprüche 1 bis 4.

Es folgen 6 Seiten Zeichnungen

Anhängende Zeichnungen

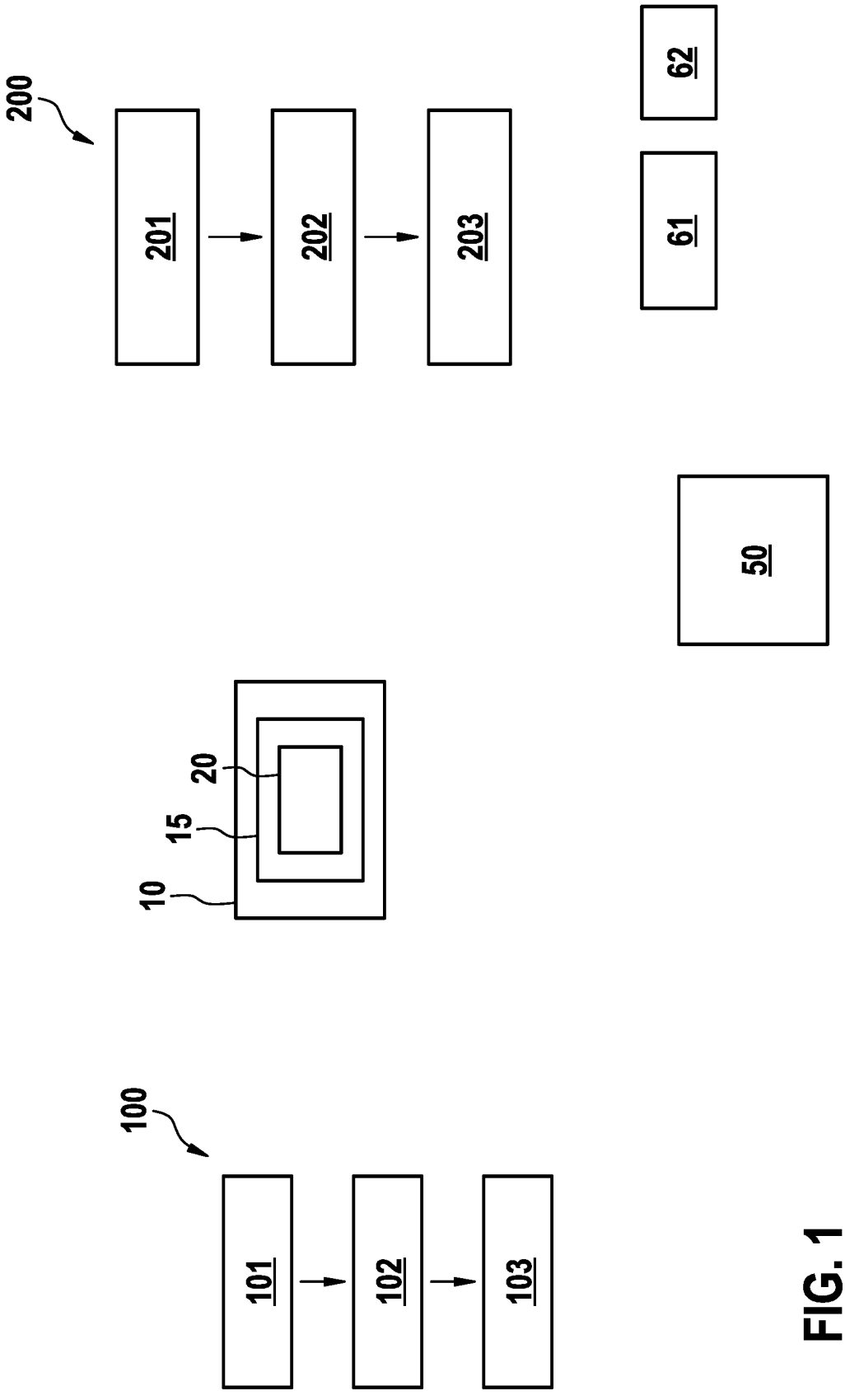


FIG. 1

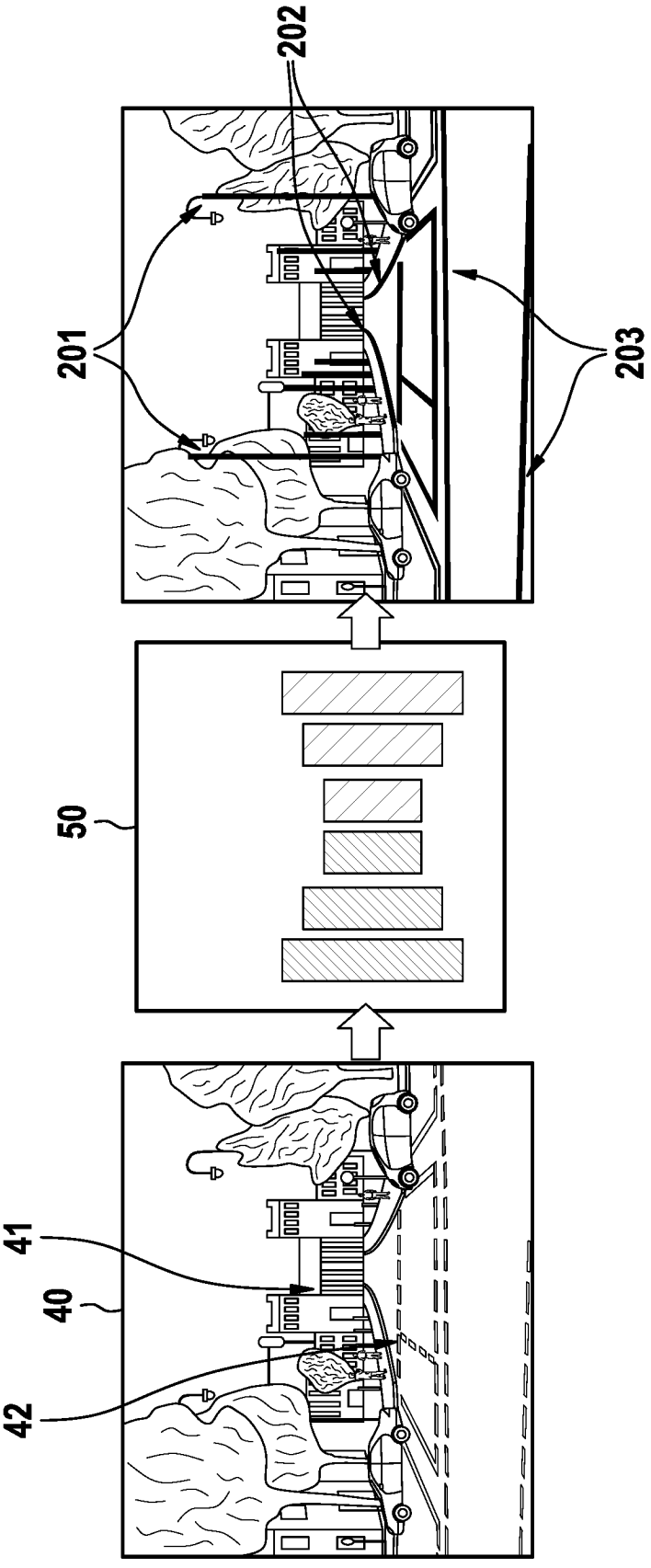


FIG. 2

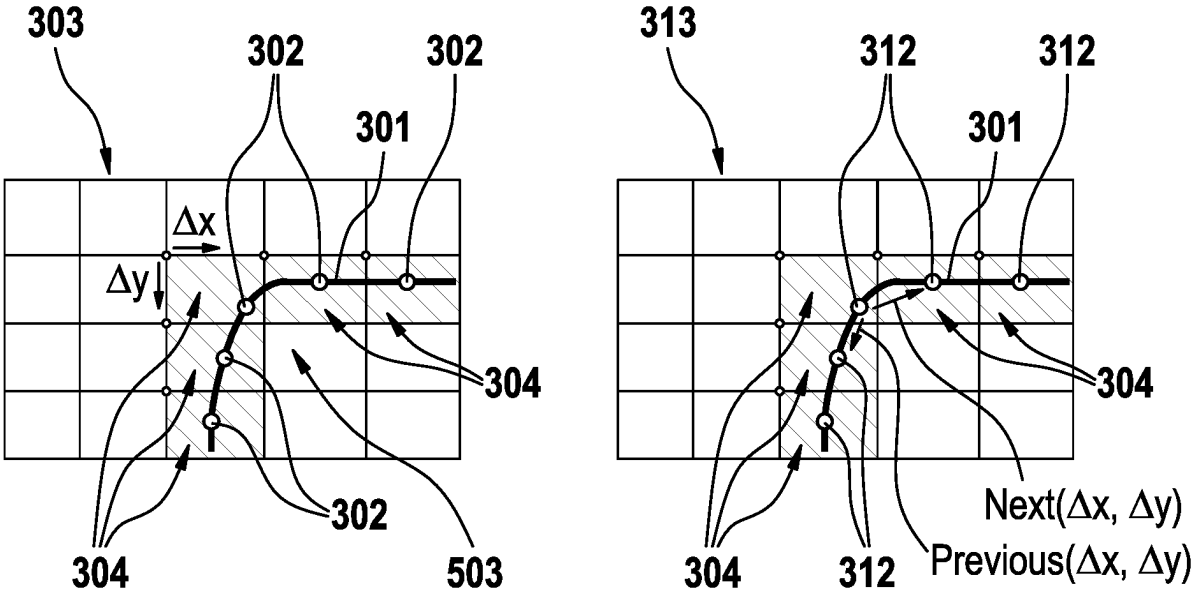

303
302
302
301
Δx
Δy
304
302
304
503
313
312
312
301
304
304
312
Next(Δx, Δy)
Previous(Δx, Δy)

FIG. 3

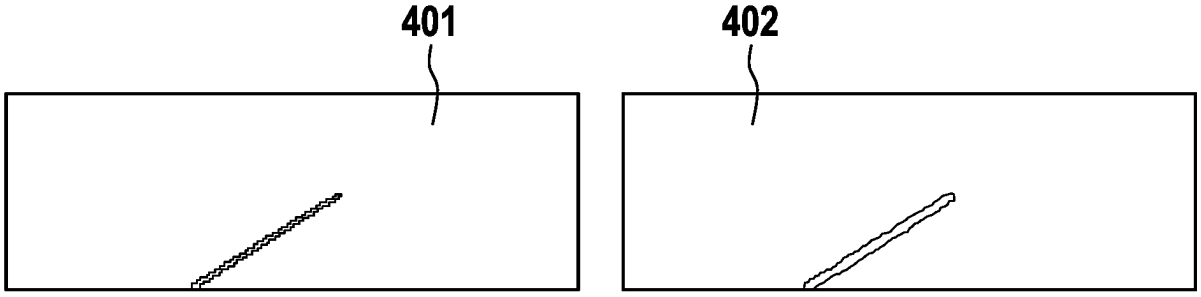

401
402

FIG. 4

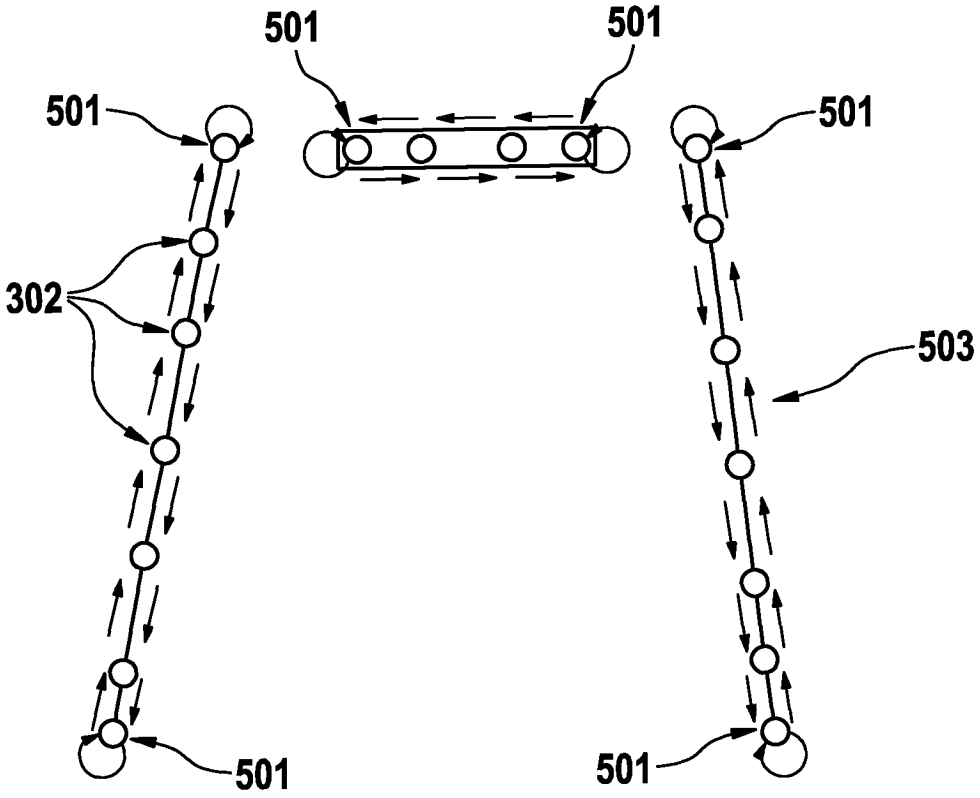


FIG. 5

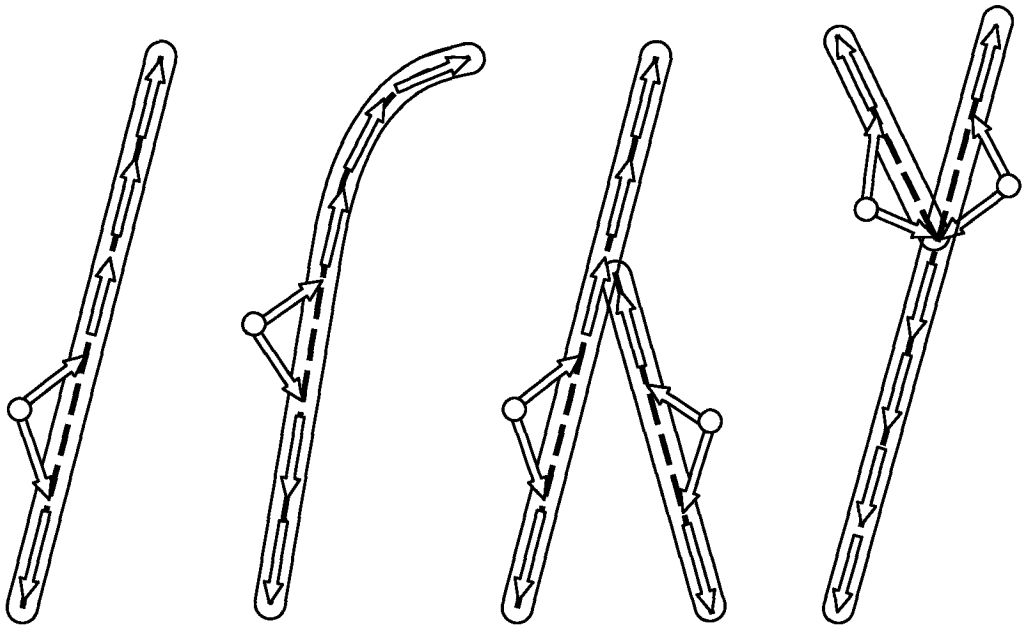


FIG. 6

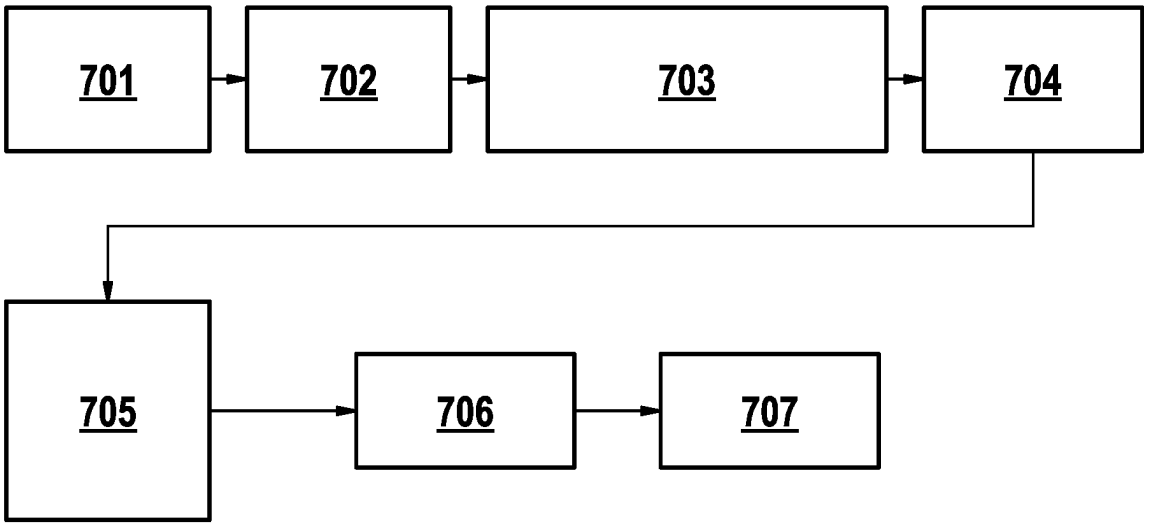

701
702
703
704
705
706
707

FIG. 7

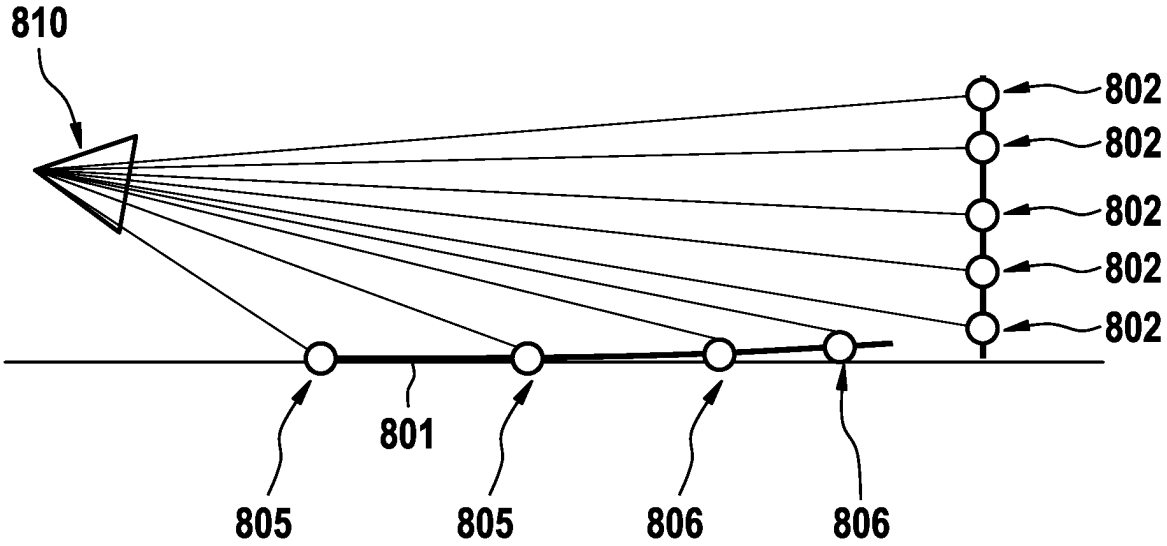

810
802
802
802
802
802
801
805
805
806
806

FIG. 8

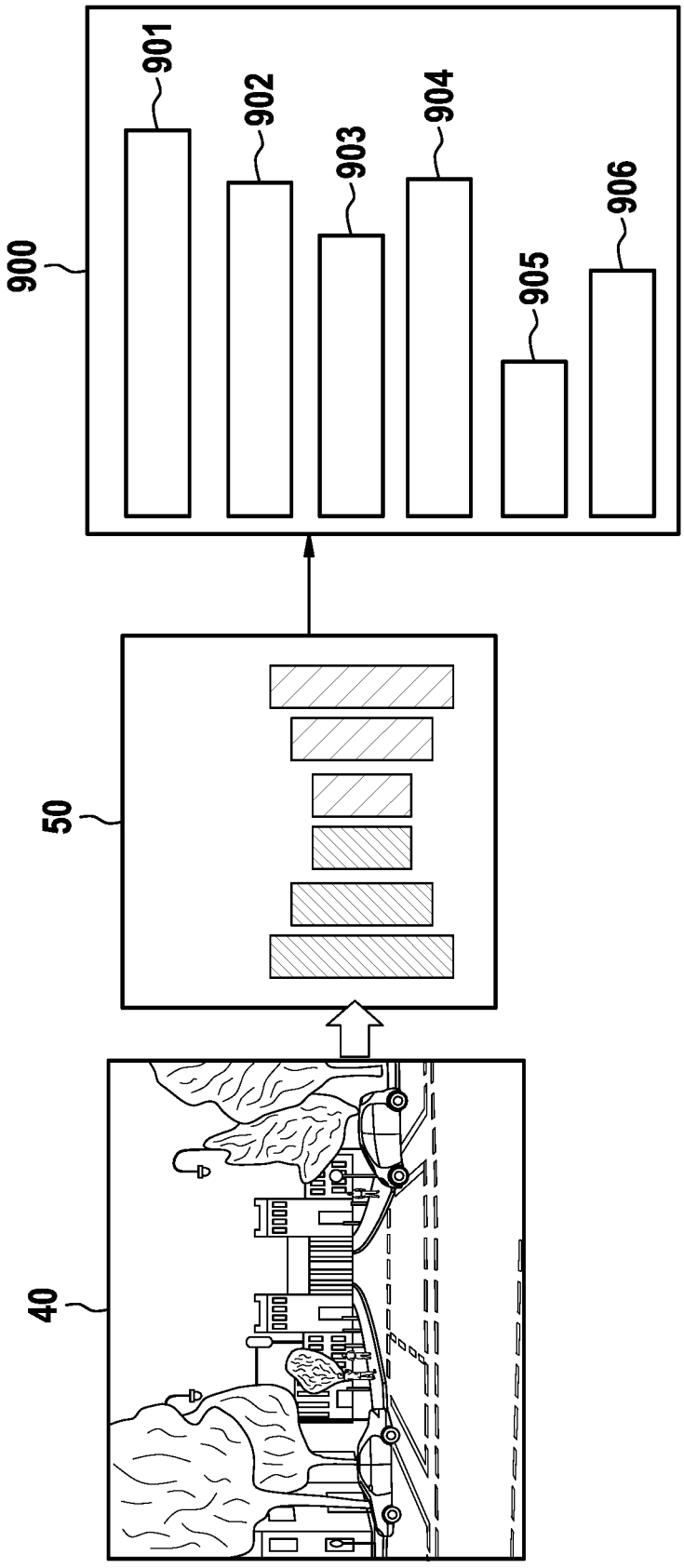


FIG. 9

