



(10) **DE 10 2017 215 079 A1** 2019.02.28

(12)

Offenlegungsschrift

(21) Aktenzeichen: **10 2017 215 079.3**

(22) Anmeldetag: **29.08.2017**

(43) Offenlegungstag: **28.02.2019**

(51) Int Cl.: **G08G 1/017** (2006.01)

(71) Anmelder:

OSRAM GmbH, 80807 München, DE

(72) Erfinder:

**Amin, Sikandar, 85375 Neufahrn, DE; Galasso,
Fabio, 85748 Garching, DE; Kaestle, Herbert,
83278 Traunstein, DE**

(56) Ermittelter Stand der Technik:

US	2012 / 0 106 794	A1
US	2015 / 0 279 021	A1
WO	2016/ 166 950	A1

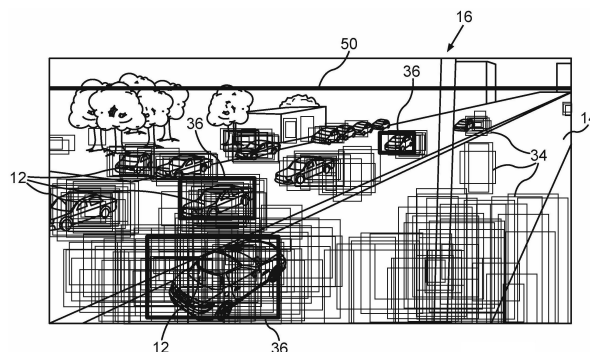
Rechercheantrag gemäß § 43 PatG ist gestellt.

Die folgenden Angaben sind den vom Anmelder eingereichten Unterlagen entnommen.

(54) Bezeichnung: **Erfassen von Verkehrsteilnehmern auf einem Verkehrsweg**

(57) Zusammenfassung: Die Erfindung betrifft ein Verfahren (10) zum Erfassen von Verkehrsteilnehmern (12) auf einem Verkehrsweg (14) in einer Abbildung, umfassend:

- Erzeugen (42) einer Vielzahl von Bereichsvorschlägen (18) für mögliche Objekte, die in der Abbildung (16) aufgezeichnet sind, durch Anwenden eines Bereichsvorschlagengenerators;
- Bereitstellen von Objekterfassung (72) für alle Bereichsvorschläge (18), um den Verkehrsweg (14) und/oder die Verkehrsteilnehmer (12) durch Klassifizieren unter Berücksichtigung eines vordefinierten Vertrauensniveaus zu erfassen;
- Ausgeben von Erfassungsdaten, die durch die Objekterfassung empfangen werden; und
- Bereitstellen eines Filters (48) für die Bereichsvorschläge (18) vor dem Schritt des Bereitstellens von Objekterfassung, wobei das Filtern basierend auf jeweiligen Filterdaten ausgeführt wird, die basierend auf einer Relevanz der Bereichsvorschläge (18) in Zusammenhang mit den Verkehrsteilnehmern (12) und/oder dem Verkehrsweg (14) geschätzt werden.



Beschreibung

[0001] Die Erfindung betrifft ein Verfahren zum Erfassen von Verkehrsteilnehmern auf einem Verkehrsweg in einer Abbildung, die von einer Kamera, die die Abbildung aufnimmt, bereitgestellt wird, wobei das Verfahren Erzeugen einer Vielzahl von Bereichsvorschlägen für mögliche Objekte, die in der Abbildung aufgezeichnet sind, indem ein Bereichsvorschlaggenerator auf die Abbildung angewandt wird, Bereitstellen von Objekterfassung für alle Bereichsvorschläge, indem ein Objekterfassungsgerät auf alle Bereichsvorschläge angewandt wird, um den Verkehrsweg und/oder die Verkehrsteilnehmer durch Klassifizieren unter Berücksichtigen eines vorbestimmten Vertrauensniveaus zu erfassen, und Ausgeben von Erfassungsdaten, die von der Objekterfassung für den erfassten Verkehrsweg und/oder die erfassten Verkehrsteilnehmer empfangen werden, umfasst. Die Erfindung betrifft auch eine Vorrichtung zum Erfassen von Verkehrsteilnehmern auf einem Verkehrsweg in einer Abbildung, wobei die Vorrichtung mindestens eine Kamera, die die Abbildung des Verkehrswegs aufnimmt, und ein Gerät umfasst, das konfiguriert ist, um eine Vielzahl von Bereichsvorschlägen für mögliche Objekte, die in der Abbildung aufgezeichnet sind, zu erzeugen, indem ein Bereichsvorschlaggenerator auf die Abbildung angewandt wird, Objekterfassung für alle Bereichsvorschläge bereitzustellen, um den Verkehrsweg und/oder die Verkehrsteilnehmer durch Klassifizieren unter Berücksichtigung eines vorbestimmten Vertrauensniveaus zu erfassen, und Erfassungsdaten auszugeben, die von der Objekterfassung für den erfassten Verkehrsweg und/oder die erfassten Verkehrsteilnehmer empfangen werden. Außerdem betrifft die Erfindung auch ein Verfahren zum Bereitstellen von Verkehrsführung, die das Erfassen von Verkehrsteilnehmern auf einem Verkehrsweg in einer Sequenz aufeinanderfolgender Abbildungen umfasst, die von einer Kamera bereitgestellt werden, die die Sequenz von Abbildungen des Verkehrswegs aufnimmt, Bestimmen einer verwendeten Kapazität des Verkehrswegs, der durch die Verkehrsteilnehmer bereitgestellt wird, Identifizieren individueller tatsächlicher Geschwindigkeit und/oder Position jedes der Verkehrsteilnehmer, um jeweilige individuelle tatsächliche Bahnen zu bestimmen, Bestimmen mindestens einer jeweiligen Nennbahn für jeden der Verkehrsteilnehmer, und Kommunizieren der bestimmten Nennbahnen zu den jeweiligen Verkehrsteilnehmern. Ferner betrifft die Erfindung auch ein Verkehrsleitsystem, um Verkehrsführung bereitzustellen, das mindestens eine Vorrichtung zum Erfassen von Verkehrsteilnehmern auf einem Verkehrsweg in einer Sequenz aufeinanderfolgender Abbildungen, und ein Gerät umfasst, das konfiguriert ist, um Verkehrsteilnehmer auf dem Verkehrsweg in der Sequenz aufeinanderfolgender Abbildungen zu erfassen, eine verwendete Kapazität des Verkehrswegs zu bestimmen, der durch die Verkehrsteilnehmer bereitgestellt wird, individuelle tatsächliche Geschwindigkeit und/oder Position jedes der Verkehrsteilnehmer zu identifizieren, um jeweilige individuelle tatsächliche Bahnen zu bestimmen, mindestens eine jeweilige Nennbahn für jeden der Verkehrsteilnehmer zu bestimmen, und die bestimmten Nennbahnen zu den jeweiligen Verkehrsteilnehmern zu kommunizieren. Schließlich betrifft die Erfindung auch ein Computerprogrammprodukt, das ein Programm für eine Verarbeitungsvorrichtung aufweist.

[0002] Aktuelle Verfahren und Vorrichtungen werden gewöhnlich an auf Kamera basierende Überwachung und Steuerung von Verkehrsknotenpunkten angewandt. Das kann eine automatische Bestimmung und Zuweisung jeweils eines adaptiven Zeitfensters und/oder einer Spur eines Verkehrswegs oder eines vollständigen Wegs jeweils zu jedem der Fahrzeuge auf dem Verkehrsweg, als Verkehrsteilnehmer aufweisen.

[0003] Bisher wurden Überwachung und Steuerung von Verkehrsknotenpunkten gewöhnlich von herkömmlichen Steuersystemen mit statischem Licht, vom Fahrzeug betätigten Ampelsteuersystemen, jeweiligem Kreisverkehr, Extrapuren, die während Stoßzeiten bereitgestellt werden, und dergleichen bereitgestellt. Die Erfindung betrifft folglich das Verwenden von Daten mindestens einer Kamera, insbesondere von Videodaten in einer Umgebung im Freien zur Verkehrsüberwachung.

[0004] Das Wachstum und das Ausmaß von Fahrzeugen machen das Verkehrsmanagement laufend schwieriger. Existierende und herkömmliche Verkehrsmanagementsteuersysteme basieren auf Timingmechanismen, wie zum Beispiel Ampeln, die üblicherweise gleiche Zeitfenster für jede Kreuzungsspur bereitstellen, was zu inhärenter Ineffizienz durch nicht gleichförmigen Verkehrsfluss führt. Außerdem können während Stoßzeiten Kreisverkehre und Extrapuren bereitgestellt werden, die typischerweise eine größere Landfläche erfordern oder das Abreißen benachbarter Strukturen für ihren Bau bedingen.

[0005] Ein effizienterer Ansatz basiert auf dem Bereitstellen adaptiver Zeitfenster für jede Spur der Straße basierend auf Echtzeit-Verkehrsdichtedaten durch fortschrittliche Kommunikationssysteme eines entsprechenden Satzes von Sensortechnologie zum Beobachten des Verkehrsszenarios. Autonome Fahrzeuge oder selbstfahrende Automobile werden zunehmend machbarer, und die Interkommunikation zwischen mehreren Fahrzeugen bzw. Automobilen und lokalen Kreuzungssteuerungsvorrichtungen erlauben automatische Wechselwirkung.

[0006] Obwohl aktuelle Verfahren zur Fahrzeugkoordination für das Fahren auf offener Straße zum Arbeiten mit menschlichen Fahrern konzipiert sind, erlaubt das Konzept autonomer Fahrzeuge das Anwenden von Gesamtstraßensteuervorrichtungen, die direkt mit den einzelnen Fahrzeugen interagieren und das Identifizieren des effizientesten und sichersten Navigationswegs sowohl bei Stadtverkehrsszenarien als auch bei Schwerknotenknotenpunkten erlauben.

[0007] In dem Fall von von Menschen gefahrenen Automobilen, können die Betriebsstraße und Kreuzungssteuervorrichtung mindestens Navigationsinformationen über optimale Spur- und Streifenauswahl gemeinsam mit dienlichsten Geschwindigkeitsempfehlungen bereitstellen.

[0008] Das herkömmliche Verkehrssteuersystem basiert auf Stopptafeln, Ampeln oder dem Einsetzen eingelassener Schleifendetektoren zur Fahrzeugerkennung und automatischer Geschwindigkeitskontrolldisplays, um die Fahrer zu informieren. Für die Verkehrsflusssteuerung wenden traditionelle Systeme Kreuzungssteuermechanismen an, die gewöhnlich gleiche oder vorprogrammierte Zeitfenster für jede Kreuzungsspur zuweisen, wobei das zeitliche Umschaltmuster zum Einrichten gleichförmiger Fahrzeugströme für von Menschen gefahrene nicht automatisierte Automobile beschränkt ist.

[0009] Angesichts des Stands der Technik verbleibt eine Notwendigkeit, die Verkehrssteuerung zu verbessern, um den Verkehrsfluss zu verbessern und Staus auf dem Verkehrsweg zu verringern.

[0010] Um den oben erwähnten Gegenstand zu meistern, schlägt die Erfindung Verfahren, Vorrichtungen und Computerprogrammprodukte gemäß den unabhängigen Ansprüchen vor.

[0011] Weitere Verbesserungen können durch Merkmale der abhängigen Ansprüche erzielt werden.

[0012] Im Hinblick auf ein generisches Verfahren zum Erfassen von Verkehrsteilnehmern auf einem Verkehrsweg, lehrt die Erfindung insbesondere das Bereitstellen der Bereichsvorschläge vor dem Schritt des Bereitstellens von Objekterfassung, wobei Filtern basierend auf jeweiligen Filterdaten ausgeführt wird, die basierend auf einer Relevanz der Bereichsvorschläge hinsichtlich der Verkehrsteilnehmer und/oder des Verkehrswegs geschätzt werden.

[0013] In Zusammenhang mit einer jeweiligen generischen Vorrichtung zum Erfassen von Verkehrsteilnehmern auf einem Verkehrsweg in einer Sequenz aufeinanderfolgender Abbildungen, lehrt die Erfindung insbesondere, dass das Gerät ferner konfiguriert ist, um ein Filtern für die Bereichsvorschläge bereitzustellen, bevor Objekterfassung bereitgestellt wird, wobei das Filtern basierend auf jeweiligen Filterdaten ausgeführt wird, die basierend auf einer Relevanz der Bereichsvorschläge in Zusammenhang mit den Verkehrsteilnehmern und/oder dem Verkehrsweg geschätzt werden.

[0014] Im Hinblick auf ein generisches Verfahren zum Bereitstellen von Verkehrsführung, lehrt die Erfindung insbesondere das Erfassen von Verkehrsteilnehmern auf dem Verkehrsweg durch Anwenden eines erfindungsgemäßen Verfahrens zum Erfassen von Verkehrsteilnehmern auf einem Verkehrsweg.

[0015] Die Erfindung lehrt auch insbesondere für ein Verkehrsleitsystem der generischen Art, dass die Vorrichtung zum Erfassen von Verkehrsteilnehmern erfindungsgemäß konfiguriert ist.

[0016] Die Erfindung basiert auf der Feststellung, dass richtige Erfassung von Verkehrsteilnehmern auf einem Verkehrsweg ein gutes Konzept sein kann, um die allgemeine Fahrzeugwartezeit durch auf Kamera basierende Verkehrsüberwachung verbunden mit jeweils autonomen und automatisierten Automobilen oder Fahrzeugen zu verringern. Das erlaubt das Optimieren des Verkehrsflusses und der Fahrbahneffizienz bzw. Verkehrsweffizienz, insbesondere in Anbetracht des Konzepts autonomer Verkehrsteilnehmer, wie zum Beispiel Fahrzeuge, Automobile, LKWs und dergleichen. Autonome Fahrzeuge sind mit gegenseitiger Kommunikationsinfrastruktur ausgestattet, die Mittel zur Fahrzeugkoordination und Verkehrsflusssteuerung ohne die Einschränkungen für das signifikante Beschleunigen oder Verlangsamen oder unnützes Anhalten bereitstellt.

[0017] Die Erfindung erlaubt es Verkehrsteilnehmern bzw. Fahrzeugen insbesondere, mit einem zentralen Kreuzungssteuersystem verbunden zu sein, das Videoinfrastruktur zum Prüfen der angerechneten Bahnen für die einzelnen Fahrzeuge einsetzt. Die Videoinfrastruktur kann durch eine oder mehrere Kameras bereitgestellt werden. Das erlaubt es, herkömmliche Ampeln durch die Kreuzungssteuervorrichtung zu ersetzen.

[0018] Die Kreuzungssteuervorrichtung, gelegentlich auch Kreuzungsmanagement-Steuervorrichtung genannt, hat sowohl die Fähigkeit zur Erfassung als auch zur Kommunikation und Steuerung, und sie kann zusätzlich mit bevorzugt jedem der sich nähernden Verkehrsteilnehmer bzw. Fahrzeuge verknüpft sein, um ihre einzelnen Bahnverläufe zu koordinieren, was eine adaptivere und intelligendere Verkehrsflusssteuerung ermöglicht.

[0019] Außerdem erlaubt ein auf Zeitfenster basierender Kreuzungssteuermechanismus, der angemessene zeitliche Auflösung aufweist, zusätzlich das Koordinieren eines bevorzugt gleichmäßigen Fahrzeugflusses, wobei das Anhalten im Leerlauf oder das vollständige Stoppen des Fahrzeugs an Ampeln geglättet oder vermieden werden kann, bzw. restliche vorwegnehmende Bewegungen, wie zum Beispiel Anpassen einer Geschwindigkeit eines bestimmten Fahrzeugs an den Bahnverlauf eines vorbestimmten Fahrzeugs, um einen im Wesentlichen stationären Verkehr zu schaffen.

[0020] Die Erfindung resultiert darin, dass sich im Wesentlichen alle Fahrzeuge gleichzeitig und schneller bewegen, so dass die Kreuzungseffizienz signifikant erhöht werden kann. Mit autonomen und automatisierten Verkehrsteilnehmern bzw. Fahrzeugen, werden Ampeln zum Verwalten von Verkehrskreuzungen folglich entbehrlich. Das erfinderische Konzept kann auch an Zubringer zum Erreichen von Autobahnen, für das Zusammenführen von Fahrspuren im Allgemeinen, und dergleichen angewandt werden.

[0021] Bei einer bestimmten Ausführungsform kann das Erfassen von Verkehrsteilnehmern auf einem Verkehrsweg, insbesondere der automatische Fahrzeugerfassungsrahmen des Verkehr-Kreuzungssteuersystems, auf Deep Convolutional Neural Networks basieren. Das erlaubt es, ohne Weiteres in Betracht zu ziehen, dass Fahrzeuge gewöhnlich in Bildern bzw. Abbildungen, die durch die Kamera aufgenommen werden, aufgrund ihrer variablen Entfernung von der Kamera in unterschiedlichen Maßstäben erscheinen. Um die Verkehrsteilnehmer bzw. Fahrzeuge variierender Größen zu erfassen, muss die Vorrichtung jeweils nach Objekten in mehreren Maßstäben in den Bildern oder Abbildungen suchen. Die Suche nach mehreren Maßstäben zieht jedoch eine hohe Latenzzeit nach sich und könnte in niedrigerer Erfassungspräzision resultieren. In diesem Hinblick erlaubt es die Erfindung, Selbstinbetriebnahme auszuführen. Sie schätzt und verwendet automatisch Szenen-Layoutinformationen wie zum Beispiel hinsichtlich einer Bodenfläche oder dergleichen.

[0022] Außerdem erlaubt es die Erfindung auch, einen Maßstab der Bodenfläche ausgehend von einer Langzeitbeobachtung der erscheinenden Fahrzeuge mit unterschiedlichen Maßstäben zu schätzen. Das kann auf selbstkonsistenter Analyse oder auf einer Regression und dergleichen basieren. Zu diesem Zweck kann die Erfassung von Skalierungsinformationen und/oder Bodenflächeninformationen aus einer Langzeitverkehrsbeobachtung bereitgestellt werden, was ebenfalls helfen kann, die Erfassung zu beschleunigen. Die Skalierungsinformationen können zum Beispiel automatisch in dem RPN enthalten sein.

[0023] Das Beginnen des erfinderischen Verfahrens oder der erfinderischen Vorrichtung erfordert keine Kenntnis über die vorliegende Szene. Die Erfindung beginnt mit Erfassen von Objekten in allen Maßstäben. Nachdem eine ausreichende Anzahl von Objekten, nämlich von Verkehrsteilnehmern, bevorzugt Fahrzeugen, an mehreren unterschiedlichen Bereichen in dem Bild bzw. der Abbildung erfasst wurde, ist die Erfindung fähig, ein Layout der Bodenfläche der Szene zu schätzen. Das macht die Erfindung robuster und schneller, da die Anzahl von Maßstäben zum Suchen nach einem Gegenstand an unterschiedlichen Lagen in dem Bild bzw. der Abbildung, verringert werden kann.

[0024] Selbstinbetriebnahme kann folglich auch hilfreich sein, falls die Kameraposition neu eingestellt wird, zum Beispiel während regelmäßiger Wartung, Servicearbeit, die gewöhnlich von Zeit zu Zeit bereitgestellt wird, so dass es folglich nicht erforderlich ist, zusätzlichen Aufwand bereitzustellen, wie zum Beispiel technische Unterstützung oder Inbetriebnahme.

[0025] Im Allgemeinen kann der Begriff „Verkehrsteilnehmer“ auf einen beliebigen Benutzer, der am Verkehr teilnimmt, angewandt werden. Ein Verkehrsteilnehmer kann folglich ein Fahrzeug, ein Fahrrad, ein Fußgänger und dergleichen sein. In dieser Hinsicht entspricht der Begriff „Verkehrsweg“ einer beliebigen auf Land basierenden Fläche, die bereitgestellt wird, damit sich Verkehrsteilnehmer auf ihr bewegen können. Ein Verkehrsweg kann folglich eine Landstraße, eine Straße, eine Allee, eine Autobahn aber auch ein Bürgersteig und dergleichen sein. Zu bemerken ist jedoch, dass die Erfindung nicht auf Anwendungen im Freien beschränkt ist. Sie kann auch an Anwendungen in Innenräumen angewandt werden, zum Beispiel an Lagerhaltung in einem Lager, wo autonome Fahrzeuge Waren von und zu vordefinierten Lagerpositionen und dergleichen transportieren können.

[0026] In diesem Zusammenhang kann ein „Verkehrsweg“ durch eine vordefinierte Fläche für die autonomen Fahrzeuge, auf der die autonomen Fahrzeuge bewegt werden können, gebildet sein. Die vordefinierte Fläche kann ein bestimmter Abschnitt eines Bodens sein, der für Transport bestimmt ist, insbesondere innerhalb des Lagers.

[0027] Die Kamera wird verwendet, um Bilder bzw. Abbildungen in Zusammenhang mit einer gewissen vorbestimmten Sicht aufzunehmen. Die Kamera ist konfiguriert, um mehr als nur ein Bild bzw. nur eine Abbildung aufzunehmen, um insbesondere eine Sequenz aufeinanderfolgender Abbildungen derselben Ansicht aufzunehmen. In dieser Hinsicht kann die Kamera bevorzugt aus einer Videokamera bestehen.

[0028] Die Kamera stellt Bilddaten bzw. Abbildungsdaten bereit, die zu einer Bildverarbeitungsvorrichtung geliefert werden, die bevorzugt zur Echtzeit-Vorverarbeitung einer großen Menge von Bilddaten bzw. Abbildungsdaten fähig ist, um das Datenvolumen, das zu der Kreuzungssteuervorrichtung zu übertragen ist, zu verringern. Die Bildverarbeitungseinheit kann aus einer Onboard-Verarbeitung, Edge Computing und dergleichen gebildet sein.

[0029] In diesem Zusammenhang ist die Bildverarbeitungseinheit bevorzugt Teil der Kamera. Sie kann jedoch durch eine separate Einheit, die kommunikativ mit der Kamera verbunden ist, bereitgestellt werden. Echtzeit-Vorverarbeitung bedeutet Echtzeiterfassung und Überwachung von Verkehrsteilnehmern, wie zum Beispiel von Fahrzeugen, Fahrrädern, Fußgängern und dergleichen, darunter Bahnberechnung und -simulation. Die Kamera sowie die Bildverarbeitungseinheit können Teil einer Infrastruktureinheit sein, die einen Straßenlaternenmast, ein Gebäude und dergleichen umfassen kann. Die Infrastruktureinheit kann die Kamera umfassen. Sie kann zusätzlich einen FLIR, einen LPR, einen RADAR und dergleichen aufweisen. Die Infrastruktureinheit kann eine erfindungsgemäße Vorrichtung bilden.

[0030] Um die Erfindung auszuführen, kann ein Erfassungsrahmen bereitgestellt werden, der das Ausführen des erfinderischen Verfahrens ermöglicht und die erfinderische Vorrichtung bilden kann. In dieser Hinsicht kann der Erfassungsrahmen fähig sein, Szenen- und Straßenlayoutdaten zum Beispiel in Zusammenhang mit der Bodenfläche oder dergleichen als eine zusätzliche Kenntnis zu verwenden, um die Effizienz und Robustheit eines Erfassungsalgorithmus durch automatisches Beschneiden der Objektmaßstäbe zu verbessern. Das bedeutet, dass es die Kenntnis des Layouts der Bodenfläche erlaubt, dass die Kamera die internen Maßstäbe mit absoluten realen Maßstäben in Beziehung bringen kann.

[0031] Der Erfassungsrahmen erlaubt es daher, eine virtuelle dreidimensionale Szene der Objekte zu schaffen, die in den Bildern bzw. Abbildungen aufgezeichnet wurden. Insbesondere kann es die Größe des sichtbaren 2D-Objekts sein, die zur Objekterkennung hilfreich ist. Bevorzugt ist der Erfassungsrahmen fähig, Selbstinbetriebnahme auszuführen. Das bedeutet, dass der Erfassungsrahmen in einer anfänglichen Phase allmählich Daten über das Szenenlayout, Daten aus einer Langzeitbeobachtung durch Analysieren erscheinender Größen und Maßstäbe der erfassten Verkehrsteilnehmer und/oder des erfassten Verkehrswegs erfassen kann. Nachdem der Erfassungsrahmen eine ausreichende Anzahl von Verkehrsteilnehmern bevorzugt an mehreren unterschiedlichen Bereichen in dem Bild bzw. in der Abbildung erfasst hat, kann er fähig sein, das Layout der Bodenfläche der Szene zu schätzen. Das Layout der Bodenfläche kann folglich allmählich abgestimmt und im Laufe der Zeit verbessert werden. Das Neueinstellen der Kamera, insbesondere der Kameraposition, kann daher berücksichtigt und automatisch abgeglichen werden.

[0032] Der Erfassungsrahmen kann es einem Bediener auch erlauben, direkt bestimmte Daten in Zusammenhang mit der Szene einzugeben, insbesondere die Bodenfläche der Szene. Das kann verwirklicht werden, indem bestimmte Parameter sichtbarer Bodenfläche bereitgestellt werden, die vorab als Inbetriebnahme bereitgestellt werden können. In diesem Zusammenhang kann manuelle Inbetriebnahme bereitgestellt werden. Falls die Kenntnis über das Szenenlayout verfügbar ist, kann ein Suchraum für den Maßstab während der Mitverfolgung von Verkehrsteilnehmern verringert werden.

[0033] Erfindungsgemäß wird optional eine erste Abbildung der Sequenz aufeinanderfolgender Abbildungen ausgewählt und als die zu verarbeitende Abbildung definiert. Die erste Abbildung braucht nicht die erste Abbildung der Sequenz von Abbildungen zu sein. Im Allgemeinen ist es möglich, dass diese Abbildung eine beliebige der Sequenz von Abbildungen sein kann. Bevorzugt ist es jedoch die Abbildung, die einen frühesten Zeitstempel hat. Zum Zweck der Auswahl kann eine bestimmte Auswahlereinheit bereitgestellt werden, die die erste Abbildung aus einer Datenbank auswählt, in der die Sequenz von Abbildungen gespeichert ist, mindestens teilweise, zum Beispiel vorübergehend gespeichert ist. Es kann jedoch vorgesehen sein, dass eine Ab-

bildung, die von der Kamera geliefert wird, unmittelbar als die zu verarbeitende Abbildung definiert und dem erfinderischen Verfahren unterzogen wird.

[0034] Dann erzeugt ein Bereichsvorschlaggenerator eine Vielzahl von Bereichsvorschlägen, um mögliche Objekte, die in der Abbildung aufgezeichnet sind, zu verarbeiten. Der Bereichsvorschlaggenerator kann eine Recheneinheit sein, wie zum Beispiel ein Computer, der einen Mikroprozessor hat, eine digitale Signalverarbeitungseinheit (Digital Signal Processing - DSP) oder dergleichen, die von einem bestimmten Computerprogramm gesteuert wird, um die Bereichsvorschläge zu erzeugen.

[0035] Eine Objekterfassung wird bevorzugt für alle Bereichsvorschläge bereitgestellt, indem ein Objekterfassungsgerät auf alle Bereichsvorschläge angewandt wird, um den Verkehrsweg und/oder die Verkehrsteilnehmer zu erfassen, während ein vorbestimmtes Vertrauensniveau berücksichtigt wird. Die Objekterfassung führt zum Erzeugen von Erfassungsdaten für den erfassten Verkehrsweg und/oder die erfassten Verkehrsteilnehmer, die für weitere Verarbeitung oder Meldung ausgegeben werden.

[0036] Dann kann das Verfahren durch Auswählen einer anderen Abbildung der Sequenz aufeinanderfolgender Abbildungen als die Abbildung, die zu verarbeiten ist, fortgesetzt werden.

[0037] Gemäß einem Aspekt der Erfindung wird ein Filtern der Bereichsvorschläge bereitgestellt, bevor der Schritt des Bereitstellens von Objekterfassung ausgeführt wird, wobei das Filtern basierend auf jeweiligen Filterdaten ausgeführt wird, die basierend auf einer Relevanz der Bereichsvorschläge in Zusammenhang mit den Verkehrsteilnehmern und/oder dem Verkehrsweg geschätzt werden. Das erlaubt es, die Bodenfläche der Szene zu finden, basierend auf dieser Feststellung kann der Aufwand für die Objekterfassung verringert werden, weil die Anzahl von Bereichsvorschlägen verringert werden kann. Insbesondere falls die Bodenfläche bzw. der Verkehrsweg verfügbar ist, können die erzeugten Bereichsvorschläge derart ausgewählt werden, dass nur die Bereichsvorschläge, die in relevanten erfassten Objekten resultieren können, während der Objekterfassung berücksichtigt werden.

[0038] Das Filtern der Bereichsvorschläge resultiert daher in einer Verringerung der Anzahl aller Bereichsvorschläge, die von dem Bereichsvorschlaggenerator erzeugt werden. Das Filtern wird basierend auf jeweiligen Filterdaten ausgeführt, die basierend auf einer Relevanz der Bereichsvorschläge in Zusammenhang mit den Verkehrsteilnehmern und/oder dem Verkehrsweg, insbesondere der Bodenfläche der Szene, geschätzt werden.

[0039] Gemäß einer beispielhaften Verbesserung werden Maßstabinformationen vor dem Schritt des Ausführens von Objekterfassung bereitgestellt, wobei Maßstabinformationen aus Langzeitbeobachtung der erfassten Fahrzeuge auf der Bodenfläche erfasst werden. Insbesondere kann ein Bodenflächenskalieren aus der Selbstinbetriebnahmephase abgeleitet werden. Das erlaubt es, die Erfindung zu verbessern und den Aufwand insbesondere in Zusammenhang mit der Ausführung der Objekterfassung zu verringern.

[0040] Bevorzugt werden die Filterdaten basierend auf einer Szene geschätzt, die durch die zu verarbeitende Abbildung aufgenommen wurde, insbesondere der Verkehrsweg, der in der Szene enthalten ist. Das erlaubt es, das Skalieren derart bereitzustellen, dass die Objekterfassung verbessert werden kann. Insbesondere, falls angenommen wird, dass vorbestimmte Klassen von Objekten erfasst werden, kann das Skalieren zusätzlich berücksichtigt werden, um den Aufwand für Objekterfassung zu verringern. Im Laufe der Zeit können die Skalierungsdaten präziser werden.

[0041] Außerdem wird vorgeschlagen, dass die Filterdaten einen Beobachtungsbereich für ein Objekt, das von der Objekterfassung als Verkehrsteilnehmer eingestuft wird, bereitstellen. Der Beobachtungsbereich kann durch Identifizieren einer oder mehrerer Flächen innerhalb der Abbildung, die zu verarbeiten ist, abgeleitet werden, wo vordefinierte Objektklassen erscheinen bzw. nicht erscheinen können. Insbesondere falls der Verkehrsteilnehmer ein Fahrzeug, ein Fahrrad, ein Fußgänger ist, kann die Klassifikation, die durch die Objekterfassung bereitgestellt wird, verbessert werden. Die Erfindung kann es auch erlauben, dass Filterdaten in Abhängigkeit von den Daten der Objekterfassung aktualisiert werden. Es ist daher möglich, vorhandene Filterdaten durch weiteres Ausführen der Erfindung zu verbessern. Im Laufe der Zeit können die Filterdaten präziser werden.

[0042] Der Beobachtungsbereich kann zum Beispiel ein Horizont sein, der aus der zu verarbeitenden Abbildung abgeleitet wird. Gewöhnlich können Objekte, wie zum Beispiel Fahrzeuge, Fahrräder, Fußgänger und

dergleichen, nicht oberhalb des Horizonts erscheinen. Bereichsvorschläge oberhalb des Horizonts können folglich aus weiterer Verarbeitung ausgeschlossen werden.

[0043] Außerdem kann die Größe der Bereichsvorschläge durch Einsatz des Filterns präziser sein. Was die Fahrzeuge betrifft, kann die Größe durch Berücksichtigung der Szene und der Position, an der ein bestimmter Bereichsvorschlag erscheinen soll, ausgewählt werden.

[0044] Bevorzugt können Deep-Learning-Verfahren und/oder künstliche Intelligenz eingesetzt werden. Die Infrastruktureinheit kann zum Beispiel Deep-Learning-Verfahren und/oder künstliche Intelligenz gemeinsam mit dem erfinderischen Konzept einsetzen, um Verkehrsteilnehmererfassung auszuführen, wie zum Beispiel Fahrzeugerfassung, Fahrzeugidentifikation, Fahrzeugmitverfolgung und dergleichen.

[0045] Die Infrastruktureinheit kann ferner die Deep-Learning-Verfahren und/oder künstliche Intelligenz einsetzen, um aus Langzeitbeobachtungen eine Beziehung zwischen sichtbarer Objektskalierung in Zusammenhang mit der Lage des Objekts auf der Bodenfläche der Kreuzung insbesondere in einer bestimmten zu verarbeitenden Abbildung zu lernen.

[0046] Sie kann auch Langzeitbeobachtungen in Zusammenhang mit dem Bereich und der Fläche eines Felds lernen, das gültige Beobachtungen enthält, die durch die Horizontlinie oder dergleichen bereitgestellt werden können. Die Infrastruktureinheit kann auch das Deep-Learning-Verfahren und künstliche Intelligenz einsetzen und kann die gewonnenen Kenntnisse über Maßstäbe in Zusammenhang mit den erfassten Objekten unter Berücksichtigung einer Lage in der Abbildung, die zu verarbeiten ist, zum Optimieren der Leistungsparameter der Objekterkennungsvorgehensweise anwenden, wie zum Beispiel Latenz, Zuverlässigkeit und Entfernungsbereich, wobei Objekte über die Horizontlinie hinaus nicht mehr berücksichtigt werden müssen.

[0047] Ferner wird beispielhaft vorgeschlagen, dass mindestens einer der Verkehrsteilnehmer eine gewünschte Bahn anfragt, was das Bestimmen einer bestimmten Nennbahn für den Verkehrsteilnehmer und das Kommunizieren der bestimmten Nennbahn zu dem Verkehrsteilnehmer veranlasst. Das erlaubt es, bestimmte Wünsche des Verkehrsteilnehmers zu berücksichtigen. Die Infrastruktureinheit kann zum Beispiel die Anfrage prüfen und die Nennbahn durch Berücksichtigen von Bahnen aller Verkehrsteilnehmer bestimmen. Die Nennbahn kann dann dem Verkehrsteilnehmer bekannt gegeben werden, so dass der Verkehrsteilnehmer auf der Nennbahn fahren kann. Das kann auch in einem besseren Verkehrsfluss resultieren, da die Infrastruktureinheit die meisten, insbesondere alle Bahnen der gegenwärtigen Verkehrsteilnehmer berücksichtigen kann.

[0048] Beispielhaft wird auch vorgeschlagen, dass das Prüfen einer bestimmten Nennbahn durch Simulation für den Verkehrsteilnehmer enthalten ist. Insbesondere kann eine S/W-basierte Prüfung der Bahnen vor ihrem Versand bereitgestellt werden.

[0049] Bei einer anderen beispielhaften Verbesserung ist das Abfragen einer tatsächlichen Geschwindigkeit und/oder Position mindestens eines der Verkehrsteilnehmer enthalten, um seine tatsächliche Bahn zu bestimmen.

[0050] Bei einer weiteren beispielhaften Verbesserung wird in Abhängigkeit von der eingesetzten Kapazität des Verkehrswegs die Geschwindigkeit der Verkehrsteilnehmer automatisch angepasst.

[0051] Gemäß einer anderen Verbesserung ist das automatische Anpassen der Geschwindigkeit mindestens eines der Verkehrsteilnehmer enthalten, während sich der Verkehrsteilnehmer einem vorbestimmten Durchfahrtmuster für den Verkehrsweg nähert. Diese Verbesserung kann insbesondere autonomes Fahren, bevorzugt in Innenräumen sowie auch im Freien, betreffen.

[0052] Jede der Infrastruktureinheiten, insbesondere erfinderische Vorrichtungen, kann zur bidirektionalen Kommunikation der Verkehrsteilnehmer fähig sein, insbesondere erfasste Fahrzeuge, um Anfragen von den Verkehrsteilnehmern und jeweiligen Kreuzungssteuervorrichtungen und Versandbefehle zu den Verkehrsteilnehmern zu empfangen.

[0053] Die Infrastruktureinheit kann auch fähig sein, Anfragen und Befehlsdaten von den Verkehrsteilnehmern und den Kreuzungssteuervorrichtungen zu empfangen sowie Versand verarbeiteter Objektinformationen erneut zu senden.

[0054] Bevorzugt sind alle Infrastruktureinheiten bzw. erfinderischen Vorrichtungen bevorzugt miteinander verbunden und können eine Maschennetzwerktopologie für bidirektionale Kommunikation, Fahrzeuganfragen und Befehle von der Kreuzungssteuervorrichtung bilden, um Einzelfehlerstellen zu vermeiden und den Bereich der Kreuzungsfläche durch Anwenden von Auffrischungen und Wiederholungen zu erhöhen.

[0055] Bevorzugter kann die Infrastruktureinheit mit bevorzugt allen der sich nähernden Verkehrsteilnehmer, insbesondere mit Fahrzeugen, die den Kommunikationsbereich erreichen, kommunizieren und kann die Koordinaten angefragter Zielorte abfragen und die Anfragen zu der Kreuzungssteuervorrichtung weiterleiten.

[0056] Mindestens eine der Infrastruktureinheiten, die an der Kreuzung oder Zufahrt positioniert sind, kann auch ein zentrales Kreuzungssteuersystem bzw. eine Kreuzungssteuervorrichtung hosten, das/die Verkehrsdaten bevorzugt aus allen Infrastruktureinheiten sammeln kann, darunter Anfrage von den Verkehrsteilnehmern, insbesondere von den Fahrzeugen.

[0057] Die Kreuzungssteuervorrichtung kann mehrere selbst fahrende autonome oder automatisierte Verkehrsteilnehmer bzw. Fahrzeuge, die um Zugriff auf einen jeweiligen exklusiven Bahnverlauf auf dem Verkehrsweg anfragen, insbesondere eine oder mehrere Spuren zu der Kreuzung, koordinieren. Auf diese Art kann eine bestimmte Passage für einen bestimmten Verkehrsteilnehmer reserviert werden.

[0058] Ferner kann die Kreuzungssteuervorrichtung die Reservierungsanfrage von dem Verkehrsteilnehmer, insbesondere dem Fahrzeug, empfangen und kann einen Bahnverlauf berechnen, der unmittelbar in Bezug auf die tatsächlichen Verkehrsdaten, die insbesondere von externen Sensoren empfangen werden, gemeinsam mit den Wegdaten anderer Fahrzeuge vorab simuliert werden kann. Die Kreuzungssteuervorrichtung kann Anfragen von dem Verkehrsteilnehmer, insbesondere von Fahrzeugen, handhaben und ununterbrochene Verlaufsplanung und Steuerung bereitstellen.

[0059] Außerdem kann die Kreuzungssteuervorrichtung jedem der Verkehrsteilnehmer, insbesondere Fahrzeugen, ein spezifisches Zeitfenster oder eine Spur in dem vorherrschenden Durchfahrtmuster zuweisen.

[0060] Die Kommunikation zwischen der Kreuzungssteuervorrichtung und den Verkehrsteilnehmern bzw. Fahrzeugen kann durch die Infrastruktureinheiten erfolgen.

[0061] Außerdem kann die Kreuzungssteuervorrichtung eine beliebige präzise Lage und Geschwindigkeit bevorzugt jedes der einzelnen Verkehrsteilnehmer, insbesondere der Fahrzeuge, abfragen, und kann ihre individuellen Bahnen zum präzisen Steuern der Position jedes der Verkehrsteilnehmer zu einer bestimmten Zeit berechnen.

[0062] Außerdem kann die Kreuzungssteuervorrichtung sowohl einen herkömmlichen Koordinationsmechanismus in der Form einer deterministischen Zustandsmaschine sowie künstliche Intelligenz basierend auf einem Koordinationsmechanismus in der Form überwachter und auf Regeln basierender Maschinenlertechniken anwenden.

[0063] Die Kreuzungssteuervorrichtung kann über das Gewähren oder Zurückweisen der Anfrage in Abhängigkeit von dem vorherrschenden Verkehrsfluss und an die grundlegende Kreuzungssteuerpolitik ausgerichtet entscheiden.

[0064] Bevorzugt kann die Kreuzungssteuervorrichtung Onboard-Simulationsanalyse anwenden, um die Funktionalität vorgeschlagener Fahrzeugbahnen zu prüfen. Ununterbrochene und andauernde Kreuzungssimulation kann helfen, Zusammenstöße zu vermeiden.

[0065] Falls die Anfrage eines individuellen Verkehrsteilnehmers nicht gewährt werden kann, muss der Verkehrsteilnehmer verlangsamt werden, um auf eine spätere Reservierungsgenehmigung zu warten. Schlimmstenfalls kann der Verkehrsteilnehmer vollständig angehalten werden, bevor er Erlaubnis zum Einfahren in die Kreuzung erhält.

[0066] Bevorzugt, falls das Fahrzeug automatisch gefahren oder automatisch gesteuert wird, kann die Geschwindigkeit des einzelnen Fahrzeugs automatisch angepasst werden, während es sich einem Vierszenen-Durchfahrtmuster nähert, so dass das Fahrzeug in den projizierten Bahnverlauf zur richtigen Zeit und im richtigen Zeitfenster einfährt, während das Gesamtmuster beständig ununterbrochen ungestört fließen kann.

[0067] Bei Bedingungen mit schwerem Verkehr organisiert eine beliebige der Kreuzungssteuervorrichtungen, bevorzugt alle Kreuzungssteuervorrichtungen, bevorzugt die sich nähernden Fahrzeuge in virtuelle Chargen, statt den Gesamtverkehrsfluss auf der Basis der einzelnen Fahrzeuge zu koordinieren. Die Schwerverkehr-Kooperativ-Wahrheitskontrolle (Heavy Traffic Cooperative Truth Control) kann das Fahrzeug veranlassen, einem anderen zu folgen, um gemeinsame Beschleunigungsmanöver auf einem Feed-Forward-Steuerweg durch Fahrzeug-zu-Fahrzeug-Kommunikation zu teilen.

[0068] Bevorzugt ist die Kreuzungssteuervorrichtung bevorzugt mit mindestens einem auf Kamera basierenden Überwachungssensor (CCTV-Kamera) verbunden, um Videodaten des vorherrschenden Verkehrsszenarios zu steuern.

[0069] Bevorzugt kann die Kreuzungssteuervorrichtung ein Onboard-Bildverarbeitungssystem hosten, das fähig ist, die individuellen Verkehrsteilnehmer, insbesondere Fahrzeuge, auf der Basis des Erscheinens zu identifizieren, um vorübergehend einen individuellen Identitätsindex zu zeichnen, der voll automatisiertes Mitverfolgen und Neuidentifikation in dem Fall erlaubt, in dem das Fahrzeug in dem Blickfeld benachbarter Kamerasensoren stromabwärts des Verkehrswegs wieder erscheint.

[0070] Für diesen Zweck kann die Kreuzungssteuervorrichtung eine Onboard-Bildverarbeitung hosten, die fähig ist, die einzelnen Verkehrsteilnehmer, insbesondere Fahrzeuge, auf der Basis automatischer Nummernschilderkennung mittels optischer Zeichenerkennung (Optical Character Recognition - OCR) oder dergleichen zu identifizieren.

[0071] Außerdem kann die Kreuzungssteuervorrichtung ein Onboard-Bildverarbeitungssystem hosten, das fähig ist, Fußgänger, insbesondere Fußgänger außerhalb von Zebrastreifen zu identifizieren und mitzuverfolgen.

[0072] Das Onboard-Bildverarbeitungssystem kann auch konfiguriert sein, um Fußgänger zu identifizieren und mitzuverfolgen. Bevorzugt kann das Onboard-Bildverarbeitungssystem der Kreuzungssteuervorrichtung konfiguriert sein, um Fußgänger außerhalb des Zebrastreifens zu identifizieren und auf dem Verkehrsweg mitzuverfolgen, insbesondere, falls der Verkehrsweg eine Straße ist, oder auf der Kreuzung, um die Ausführung einer speziellen Zusammenstoßschutzvorgehensweise auszulösen, die den Verkehrsteilnehmern, insbesondere Fahrzeugen, die sich in unmittelbarer Nähe befinden, neue Verlaufs- und Wegdaten zuzuweisen.

[0073] Bevorzugt können alle Infrastruktureinheiten mit der Kreuzungssteuereinheit, die durch eine Kreuzungssteuervorrichtung bereitgestellt wird, kommunizieren.

[0074] Die Lehren der vorliegenden Erfindung können ohne Weiteres verstanden werden, und mindestens einige zusätzliche spezifische Einzelheiten ergeben sich bei der Betrachtung der folgenden ausführlichen Beschreibung mindestens einer beispielhaften Ausführungsform verbunden mit den begleitenden Zeichnungen, in welchen

Fig. 1 schematisch eine zu verarbeitende Abbildung zeigt, die eine Vielzahl von Bereichsvorschlägen für das Erfassen von Objekten sowie drei beispielhafte klassifizierte Objekte aufweist;

Fig. 2 in einer schematischen dreidimensionalen Skizze eine erfindungsgemäße Skalierfunktion für das Filtern von Vorschlägen unter Verwenden geschätzter skalierte Filter zeigt;

Fig. 3 eine schematische zweidimensionale Skizze zeigt, die eine Projektion der **Fig. 2** in einer Ebene ist, die verwendet wird, um die Skalierfunktionen zu schätzen;

Fig. 4 die Abbildung, die gemäß **Fig. 1** zu verarbeiten ist, zeigt, wobei die Bereichsvorschläge erfindungsgemäß gefiltert werden, und zusätzlich zeigt, dass Bereichsvorschläge kleiner werden, während sie sich dem Horizont nähern;

Fig. 5 ein schematisches Flussdiagramm zeigt, das ein Verfahren zum Erfassen von Verkehrsteilnehmern auf einem Verkehrsweg gemäß der Erfindung zeigt;

Fig. 6 eine schematische Skizzenskala verglichen mit Wiedererkennung für VGG-M-Netzwerke zeigt, wobei Bins durch gleichmäßig verteilte Muster bestimmt werden;

Fig. 7 eine schematische Okklusionsskizze im Vergleich zu Wiedererkennung für VGG-M-Netzwerke zeigt, wobei Bins durch gleichmäßig verteilte Muster bestimmt werden, und wobei ein maximales Okklusionsverhältnis auf 0,5 gestellt ist.

[0075] Die ausführlichen Ausführungsformen, die unten beschrieben sind, befassen sich damit, wie die Erfindung insbesondere autonomem Fahren unterzogen wird. Die Erfindung ist jedoch nicht auf Anwendungen im Freien beschränkt, sondern kann auch in Innenraumanwendungen angewandt werden, wie zum Beispiel bei Lagerhallenanwendungen, insbesondere im Hinblick auf das Lagern von Waren und dergleichen.

[0076] Autonomes Fahren bleibt immer noch eine große Herausforderung, bei der die Umgebung, wie sie von einer oder mehreren Kameras aufgenommen wird, die eingerichtet sind, um Bilder bzw. Abbildungen von Fahrzeugen als Verkehrsteilnehmeränderungen, die schnell und unerwartet auftreten können, beobachtet wird. Fahrzeuge können zum Beispiel am Straßenrand geparkt werden, diverse Initiativen und Ereignisse in der Stadtmitte können sich auf den Verkehr auswirken, und dergleichen. Insbesondere kann die Gegenwart von mehr Leuten in einer höheren Wahrscheinlichkeit, dass jemand die Straße überquert, resultieren.

[0077] In diesem Zusammenhang hat auf Infrastruktur basierendes Erfassen das Potenzial, den Einzelfahrzeug-Standpunkt zu ergänzen und die Einführung voll autonomer Fahrzeuge zu beschleunigen. Die Erfindung betrifft insbesondere das Erfassen und Erkennen von Verkehrsteilnehmern, wie zum Beispiel von Fahrzeugen, aus der Voraussicht von Überwachungskameras, die im Wesentlichen den Fahrzeugen eine Wahrnehmung über das Fahrzeug vor ihnen hinaus und hinter einer Ecke verleihen.

[0078] Die Kameras können ein Bauteil einer oder mehrerer erfindungsgemäßer Vorrichtungen sein. Es besteht ein großes Potenzial in der Nutzung der statischen Sicht einer Überwachungskamera, die eine bessere und schnellere Erfassung ergeben kann.

[0079] Die Erfindung betrifft insbesondere Bereichsvorschläge, die ein wichtiges Merkmal moderner Erfassungsalgorithmen bilden. In diesem Zusammenhang schlägt die Erfindung eine einfache Erweiterung des Faster R-CNN (Regional-based Convolutional Neural Network) vor und zeigt, dass das Einstufen von Vorschlägen in Zusammenhang mit einer Szenengeometrie in weniger falschen positiven Resultaten resultieren kann, indem Vorschläge in überladenen Bereichen verringert werden, wo jeweilige Algorithmen gewöhnlich für Fehler anfällig sind.

[0080] Außerdem ergibt die Erfindung weniger falsche negative Resultate, da sie Wiedererkennung durch Enthalten von mehr Vorschlägen, wo sie am nötigsten sind, zum Beispiel für kleine Fahrzeuge an der Entfernung erhöht. In Zusammenhang mit einer solchen Ausführungsform wird mit dem UA-DETRAC-Datensatz experimentiert, der auf dem Vanilla Faster R-CNN (VGG-16) um etwa mehr als 19 % verbessern kann. Diese Verbesserung kann weitgehend beim Umschalten auf ein Faster VGG-M-Netzwerk aufrechterhalten werden.

[0081] Erfindungsgemäß wird vorgeschlagen, 3D-Objektvorschläge zu erzeugen, indem Szenengeometrie unter Verwenden kalibrierter monokularer oder Stereokameraanordnung genutzt wird. Eine automatische Schätzung der ungefähren Szenengeometrie im Sinne eines maßstabgerechten Layouts wird vorgeschlagen.

[0082] Diese Informationen werden in einen Detektor eingebaut, um Bereichsvorschläge zu erzeugen. Der erfinderische Ansatz berücksichtigt, dass die Szene weitgehend statisch ist, aber gewöhnlich keine Kamera-kalibrierungsinformationen bzw. -daten benötigt.

[0083] Die Bedeutung von Hochauflösungsfunktionen wird in der Literatur semantischer Segmentierung ebenfalls angesprochen, wo das Ziel darin besteht, präzise Objektgrenzen und hochwertige Semantiken zu bestimmen. Für dichte Vorhersageaufgaben werden typischerweise Ent-Faltungen eingesetzt. Es wurde jedoch auch eine einfachere Alternative vorgeschlagen, nämlich das Entfernen von Max-Pooling-Schichten für dichte Funktionskarten, und das Einsetzen erweiterter Faltungen, um die Kontaktgröße so groß wie möglich zu halten, ohne die Anzahl von Parametern zu erhöhen. Das Entfernen von Pooling-Schichten kann auch das Erfassen von kleinformatigen Fußgänger betreffen. Sie berücksichtigen jedoch keine erweiterten Faltungen.

[0084] Die Erfindung erweitert Faster R-CNN in Zusammenhang mit Objekterfassung, um vorgeschlagene geometrische Vorschläge einzugliedern. Geometrische Vorschläge codieren das Szenenlayout einer statischen Kamera auf eine einfache und effektive Art. Im Allgemeinen kann das Faster R-CNN-Detektion in mindestens zwei Etappen arbeiten. Bei einer ersten Etappe wird ein volles Faltungs-Bereichsvorschlagsnetzwerk (Region Proposal Network - RPN) bereitgestellt, das das vollständige Bild bzw. die vollständige Abbildung, das/die zu verarbeiten ist, an dem Eingang nimmt und klassenagnostische Objektvorschläge erzeugen kann.

[0085] Die zweite Stufe basiert auf einem Klassifizierungsnetzwerk, das die eingehenden Vorschläge in gegebene Objektklassen einstuft. Die Faltungsschichten werden für beide Aufgaben, das heißt Erzeugen von Vorschlägen und deren Klassifizieren, gemeinsam verwendet.

Geometrische Erweiterung auf RPN

[0086] Gewöhnlich kann ein Bild bzw. eine Abbildung, das/die zu verarbeiten ist, potenziell einige wenige große Objekte und einige oder mehrere kleine Objekte enthalten. Das wird jedoch üblicherweise nicht in dem Faster R-CNN-Algorithmus insbesondere aufgrund des RPN, das eine gleiche Menge von Objekten über Maßstäbe hinweg vorschlägt, berücksichtigt. Eine Beziehung zwischen den Bereichsvorschlägen und der Szenengeometrie wird mittels einer Objekt-Maßstab-Schätzung wieder ermittelt. Zuerst werden die sichersten Objekte erfasst. Dann wird eine pixelweise Maßstabschätzung als ein Proxy für die tatsächliche Szenengeometrie geschätzt. Abschließend wurden die RPN-Vorschläge beschnitten.

Anfängliche Erfassungen

[0087] Die sichersten (mit hohen Resultaten) Verkehrsteilnehmererfassungen bzw. Fahrzeugerfassungen werden verwendet, um automatisch dieses Layout, das in **Fig. 1** gezeigt ist, zu schätzen. Über eine Ansicht anfänglicher Sekunden einer Videosequenz, zum Beispiel Daten 10 s zuverlässiger Erfassungen, reichen für eine wirklich zuverlässige Maßstabschätzung. Zu bemerken ist, dass die sichersten Erfassungen im Allgemeinen nicht okkludiert und nicht trunkiert sind.

Pixelweise Maßstabschätzung

[0088] Dann wird das Maßstablayout für das zu verarbeitende Bild bzw. die zu verarbeitende Abbildung geschätzt, das heißt eine Bildmaßstabfunktion, die den Maßstab des Objekts angesichts seiner Position in dem Bild beschreibt. In diesem Fall entspricht das der erwarteten Größe eines Verkehrsteilnehmers bzw. Fahrzeugs in einer bestimmten Bildposition.

[0089] Nehmen wir einen anfänglichen Satz von Erfassungen an Positionen $\{x\}_1^N$ an, wobei die x die Mitte von Erfassungs-Begrenzungsrechtecken sind und N die Anzahl anfänglicher Erfassungen ist. Es wird auf die Schätzung einer Maßstabfunktion (x) abgezielt, die die Größe des Objekt-Begrenzungsrechtecks in Pixel^2 an jeder Pixelkoordinate präsentiert. Es wird ein Polynom zweiten Ranges angenommen, um die Funktion zu approximieren und anhand der Methode der kleinsten Fehlerquadrate anzupassen:

$$s(x) = p_2 x^2 + p_1 x + p_0 \quad (1)$$

wobei p_2 , p_1 und p_0 Parameter, die zu schätzen sind, sind.

Notizen zur Polynom-Anpassung

[0090] Das Polynom zweiten Ranges reicht, um ein flaches Szenenlayout aus einer homographischen Projektion darzustellen. Diese Annahme ist bei den meisten Straßenszenen plausibel, da Verkehrswege, die von der Überwachungskamera sichtbar sind, meistens flach bzw. eben sind. Die Approximation des Maßstablayouts kompensiert für die Größenschwankungen der Objekte bzw. Fahrzeuge, zum Beispiel Automobile unterschiedlicher Größe und dergleichen, angesichts der ausreichenden Anzahl von Erfassungen. Das wird durch eine Testschätzung gemäß **Fig. 2** veranschaulicht. Das Maßstablayout stellt automatisch eine Horizontschätzung bereit, die auch in den **Fig. 2** bis **Fig. 4** durch das Bezugszeichen **50** angegeben ist.

Vorschlagbeschneiden

[0091] Innerhalb des RPN-Moduls des vorgeschlagenen GP-FRCNN, werden Objektvorschläge gemäß Folgendem beschnitten:

$$\frac{\|s(\hat{x}) - \hat{b}\|}{s(\hat{x})} < \sigma \quad (2)$$

[0092] Hier ist $s(\hat{x})$ die Maßstabschätzung des Objekts an der Position \hat{x} , wie in der Gleichung (1) beschrieben, und \hat{b} ist die tatsächliche Begrenzungs-Rechteckgröße des Objekts. σ stellt die akzeptable Abweichung der Vorschlagsgröße von der Maßstabfunktion dar. Gegenwärtig wird ihr Wert für alle unterschiedlichen Ausführungsformen basierend auf den beobachteten Varianten in den Trainingsdaten auf 0,3 festgelegt.

Über eine Objektklasse hinaus

[0093] Für den DETARC-Challenge werden nur Fahrzeuge als Zielerfassung für die Schätzung des Maßstablayouts berücksichtigt. Zu bemerken ist jedoch, dass die Gleichungen (1), (2) auch für das Schätzen anderer Objektgrößen gelten können.

[0094] Mit einer gegebenen einzelnen korrekten Maßstabsbeobachtung eines anderen Objekts, zum Beispiel eines Busses, als ein gewisses Fahrzeug, kann man die Maßstabschätzungsfunktion $s(x)$ daran durch einfache

$$\hat{b}'$$

ches Skalieren mit dem $\frac{\hat{b}'}{s(\hat{x})}$ anpassen, wobei $s(\hat{x})$ die ursprüngliche Maßstabfunktionsschätzung der Pixelposition ist, und \hat{b}' die Größe der Buserfassung ist.

Feinabstimmung für Faster R-CNN

[0095] Der Einsatz geometrischer Vorschläge ist eine einfache Erweiterung zu dem Vanilla Faster R-CNN, aber die einfache Integration kann die Leistung verschlechtern. Unten wird Anpassung vorgeschlagen, um das Modell bzw. das Verfahren zu verbessern. Die Anpassungen können für unterschiedliche Modelleinstellungen gelten, zum Beispiel die Auswahl des Netzwerks.

Spezifische Anker-Maßstäbe

[0096] Während des Trainings können Faster R-CNN die Objekt-Begrenzungsrechtecke in Anker-Maßstäbe und erwartete Verhältnisse trennen. Standardmäßig kann der Maßstabsatz von Anker-Rechtecken {8, 16, 32} sein. Das kann für die meisten Erfassungs-Benchmarks, wie zum Beispiel PASCAL VOC, geeignet sein. Das Anwenden des Faster R-CNN mit den Standard-Ankermaßstäben auf den UA-DETRAC-Standard kann jedoch schwerwiegend hinter den Erwartungen zurückbleiben, weil die meisten Fahrzeuge viel kleiner sind als der kleinste Standardmaßstab. Die Bereichsvorschläge, die den kleinsten Anker-Rechtecken entsprechen, müssen folglich für jedes Objekt dienen, das kleiner ist als seine eingestellte Größe im Kontrast zu dem eigentlichen Konzept von Anker-Maßstäben. Dieses Problem kann behoben werden, indem der Satz mit kleineren Maßstäben in der Sequenz erweitert wird, das heißt {1, 2, 4}, wofür Resultate in **Fig. 6** gezeigt sein können.

[0097] **Fig. 6** zeigt in einer schematischen Skizze den Maßstab im Vergleich zu Wiedererkennung für ein VGG-M-Netzwerk, wobei Bins als Muster gleichmäßig verteilend bestimmt wurden. **Fig. 6** zeigt eine Skizze **52**, in der eine Ordinate **54** der Wiedererkennung in % zugeordnet ist, und eine Abszisse **56** der mittleren Fahrzeuggröße (Pixel²) zugeordnet ist. Die Tupel von Bins betreffen einen Bin **58**, der dem FRCNN entspricht, einen Bin **60**, der dem FRCNN+GP entspricht, einen Bin **62**, der dem FRCNN+BW entspricht, sowie einen Bin **64**, der dem FRCNN+BW+PG entspricht.

[0098] Bei einer alternativen Ausführungsform kann auch mit quantisierten Maßstäben in Zusammenhang mit den Trainingsdaten experimentiert werden. Es hat sich jedoch herausgestellt, dass beide Techniken in ähnlichen Leistungen resultieren, weshalb unten nur die einfache Erweiterung auf den Maßstabssatz der Anker in RPN berücksichtigt wird.

Funktionskarten mit höherer Auflösung

[0099] Eine zweite Einschränkung des Faster R-CNN auf kleinere Objekte kann durch die Quasar-Auflösung seines CNN-Funktionsblocks gegeben werden. Dieses Problem wurde mehrere Male bei semantischer Segmentierung festgestellt, wodurch eine Quasar-Granularität die pixelweise Auflösung einschränken kann.

[0100] Daher wird vor dem Vorschlag des Faster R-CNN und der Klassifikationsanlage eine Finderfunktionskarte vorgeschlagen. Ausführlicher wird der Funktionsabstand von **16** zu **18** durch Entfernen der letzten Max-Pooling-Schicht von den Basisfunktionsnetzwerken an allen experimentierten Modellen verringert.

[0101] Das kann effektiv zur Erhöhung der Anzahl von Stellen auf dem Bild bzw. der Abbildung führen, an welchen nach dem Objekt zu suchen ist, was in einem signifikanten Gewinn an Wiedererkennung für die kleinen Fahrzeuge resultiert, wie aus **Fig. 6** ersichtlich ist.

[0102] Ferner ist zu bemerken, dass das in einem kleineren jeweiligen Feld auf dem eingegebenen Bild bzw. auf der eingegebenen Abbildung, das/die zu verarbeiten ist, resultieren kann. Obwohl das eventuell nicht alle kleinen Objekte, für die das Modell eine größere Fläche in dem Objekt betrachten muss, betrifft, kann das die Wiedererkennung für die größeren Fahrzeuge in den Experimenten, für die der Kontext zu knapp wird, verringern. Diese Wirkung kann bei kleineren Funktionsmodellen, wie zum Beispiel FGG-M, offensichtlicher sein, während größere Modelle robuster scheinen und potentiell ausreichend große jeweilige Felder aufrecht-erhalten.

Mehrstufraining

[0103] Das Lernen der Parameter für alle Faltungsschichten der Erfassungsaufgabe ist nicht einfach, weshalb bei der Standardstrategie zum Trainieren des Faster R-CNN-Modells die Parameter bevorzugt mit dem vor-trainierten ImageNet-Modell initialisiert werden, und das Lernen der ersten **4** Faltungsschichten übersprungen werden kann. Das bedeutet, dass die Funktionen niedrigen Niveaus in dem grundlegenden Faster R-CNN-Modell immer noch diejenigen sein können, die tatsächlich nur für die Bild-Nettoklassifizierungsaufgabe trainiert werden. Das braucht natürlich keine optimale Einstellung zu sein. Es kann jedoch ein mehrstufiger Trainingsansatz herangezogen werden, und diese anfänglichen Faltungsschichten, die in dem Standard-Trainingsstadium der Faster R-CNN übersprungen werden, können ebenfalls gelernt werden. In dem ersten Stadium kann die Standardstrategie beibehalten werden, und die Parameter der intimalen Faltungsschichten der Netzwerke werden nicht gelernt. Bei dem zweiten Stadium kann die Trainingsvorgehensweise auf dem vollständigen Netzwerk fortgesetzt werden, nachdem die ursprünglichen Faltungsschichten ebenfalls entriegelt sind. Alternativ könnte man auch eine Strategie einem Aufwärmtraining ähnlich, mit sehr kleinen Lern-Raids untersuchen.

Experimente und Resultate

[0104] Die Einzelheiten der Experimente und die Resultate der erfindungsgemäßen Ansätze auf dem UA-DETRAC werden bereitgestellt, was ein sehr umfassender Datensatz für Überwachungsszenarien ist. Der Datensatz besteht aus **100** Videosequenzen (**60** für das Training, **40** für das Testen), die reale Verkehrsszenen bei unterschiedlichen Wetterbedingungen präsentieren.

Netzwerkaufbauten

[0105] Zwei unterschiedliche Varianten der VGG-Netzwerkaufbauten werden verwendet. Die erste ist VGG_CNN_M_1024 mit **5** Faltungs- und **3** vollständig verbundenen Schichten. Unten wird dieses Netzwerk VGG-M genannt. Die zweite ist VGG-**16** mit **13** Faltungs- und **3** vollständig verbundenen Schichten.

Trainingsstrategie

[0106] Die folgende Strategie ist in dem ursprünglichen Bericht des UA-DETRAC-Datensatzes zum Auswählen von Round-Truth-Fahrzeuganmerkungen zum Trainieren der Modelle beschrieben. Das bedeutet, dass nur Fahrzeuge mit weniger oder gleich 50 % Okklusion und 50 % Trunkierung enthalten sind. Um die Modelle robust zu machen und übermäßige Anpassung an DETRAC-Szenarien zu vermeiden, wurden PASCAL VOC 2007- und 2012-Trainval-Bildsätze gemeinsam mit den DETRAC-Bildern verwendet, um die Modelle zu trainieren. Das Training wird für alle **20** Standardklassen des PASCAL VOC-Datensatzes erstellt. Die Bewertungen zeigen, dass das Training für alle **20** Klassen im Allgemeinen leicht besser sein kann als das Training nur für Fahrzeuge. Für alle Experimente wird der Ende-an-Ende-Ansatz des Faster R-CNN zum Modelltraining verwendet, der sowohl RPN als auch das Klassifizierungsnetzwerk gleichzeitig trainiert.

[0107] Die Auswahl des NMS-Schwellenwerts ist für typische Objektdetektoren recht kritisch. Da nur Fahrzeuge mit weniger oder gleich 50 % Okklusion in dem Validierungssatz berücksichtigt werden, kann es theoretisch sinnvoll sein, einen NMS-Schwellenwert von 0,5 zu verwenden. Außerdem kann ein strikterer Wert für den Faster R-CNN-Parameter FG_THRESH, das heißt 0,7 an Stelle von 0,5, für eine Erfassung, die während des Trainings des Klassifizierungsnetzwerks als eine positive Klasse zu betrachten ist, verwendet werden. Dieser Wert ist für alle gegenwärtigen Resultate optimal, da die Online-UA-DETRAC-Bewertung IoU von 0,7 verwendet, um eine Erfassung als richtig zu zählen.

Validierung

[0108] Es wird davon ausgegangen, dass die Verteilungen der Fahrzeugdaten in den Trainings- und Test-sätzen ähnlich sind, weshalb es sinnvoll ist, einen Validierungssatz für umfassende Bewertungen herauszunehmen. Bei der vorliegenden Ausführungsform werden **36** Videos ausgewählt, die unterschiedliche Gesichtspunkte und Wetterbedingungen in einem Validierungssatz aufweisen. Die restlichen **24** Videos werden verwendet, um die Modelle in dieser Validierungsphase zu trainieren.

[0109] Um das Modell zu trainieren, werden alle Rahmen aus den ausgewählten **24** Videos des Trainingssatzes verwendet. Zum Testen wird jeder zehnte Rahmen aus den 36 Sequenzen des Validierungssatzes berücksichtigt. Das erlaubt es, schnell mehrere Ansätze zu bewerten, darunter Verbesserungen und Verfeinerungen für das Faster R-CNN-Rahmenwerk. In Tabelle 1 bemerkt man, dass eine signifikante Verbesserung hinsichtlich der mittleren Präzision (Average Precision - AP) nach dem Aufnehmen von PASCAL VOC-Datensätzen erzielt werden kann. Das vorgeschlagene Verfahren kann ausführlich unter Verwendung eines kleineren VGG-M-Netzwerkes in Zusammenhang mit unterschiedlichen Aspekten bewertet werden, darunter die Fähigkeit, Maßstabänderungen zu bewältigen, und unterschiedliche Niveaus von Okklusion.

Tabelle 1

Verfahren	AP (%)	Geschwindigkeit (fps)
Faster RCNN	58,9	12
Faster RCNN	64,1	11
+ Extraanker	68,8	11
+ hochauflösende Funktionskarten	72,3	8
+ GP GP-FRCNN	78,7	8
+ Mehrstufentraining	80,9	8

Maßstabinvarianz

[0110] Hier wird berichtet, dass das abschließende Modell fähig ist, die Maßstabinvarianzeigenschaft des ursprünglichen Detektors zu verstärken. Es wird gezeigt, dass die Fusion geometrischer Vorschläge und einer modifizierten Version des Faster R-CNN fähig ist, die Detektor-Wiedererkennung ungeachtet des Maßstabs des Objekts signifikant zu verbessern.

[0111] Die Resultate werden gemäß **Fig. 6** gemeldet. Man bemerkt, dass der grundlegende Ansatz die Leistung für das Erfassen von Fahrzeugen mit kleineren Maßstäben signifikant verfehlt. Das Hinzufügen der geometrischen Vorschläge (Geometric Proposals-GP) durch Verwenden des erfinderischen vorgeschlagenen Maßstablayouts (**Fig. 1** bis **Fig. 4**) hilft eventuell überhaupt nicht. Das gibt an, dass grundlegende Modelle nicht genug Fähigkeit zum Handhaben von Objekten mit kleineren Maßstäben, wie oben besprochen, haben.

[0112] In Zusammenhang mit kleinen Objekten kann jedoch eine Verbesserung erzielt werden, indem Bells und Whistles (BW) in das Netzwerk eingeführt werden, zum Beispiel Extraanker, die auf kleinere Objekte abzielen, und der Funktionsabstand verringert wird, um Funktionskarten mit höherer Auflösung zu erlauben. Obwohl höhere Wiedererkennung für die kleineren Fahrzeuge erzielt werden kann, kann diese Änderungen eine negative Auswirkung auf das Erfassen größerer Fahrzeuge haben.

[0113] In **Fig. 6** ist sichtbar, dass Wiedererkennung konsistent nach BW sinkt. Das zeigt, dass das Erhöhen der Auflösung der Funktionskarten eventuell nicht immer helfen kann. Da dadurch das effektive rezeptive Feld auf dem zu verarbeitenden Bild bzw. der zu verarbeitenden Abbildung verringert werden kann, kann das Erfassen größerer Objekte für das Netzwerk schwierig zu handhaben werden. Die vorgeschlagenen geometrischen Vorschläge (GP), die eventuell zuvor ineffektiv waren, können jedoch einen substanziellen Gewinn hinsichtlich der Wiedererkennung für kleinere und mittlere Fahrzeuge bringen, und auch im Gleichlauf mit dem Linienmodell für die größeren Objekte arbeiten. Es ist eine beeindruckende Feststellung, dass geometrisch konsistente Vorschläge die Kapazität des Detektors für kleinere Objekte wesentlich verstärken können, ohne die Leistung für die größeren Objekte zu verschlechtern.

Bessere Okklusionshandhabung

[0114] Hier kann bewertet werden, wie gut die Modelle hinsichtlich der Handhabung unterschiedlicher Okklusionsniveaus sind. Interessanterweise bemerkt man eine ähnliche Tendenz, dass das Einführen der Bells and Whistles (BW) gemeinsam mit den geometrischen Vorschlägen (GP) fähig sein kann, die Einstufungsversagensfälle des RPN zu meistern und die besten Resultate bereitstellen kann.

[0115] Gemäß **Fig. 7** wird die Wiedererkennung aller Modelle in Zusammenhang mit unterschiedlichen Okklusionsniveaus vorgeführt. **Fig. 7** zeigt eine schematische Skizze **66**, die eine Ordinate als **Fig. 6** und eine Abszisse **68** hat, die einem Hauptokklusionsverhältnis zugewiesen ist. In der Skizze **66** ist das Tupel der Bins **58** bis **64** entsprechend den Bins gemäß **Fig. 6** gezeigt. Die Wiedererkennung verbessert sich nur leicht, wenn die geometrischen Vorschläge an das Vanilla Faster R-CNN angewandt werden, was die beschränkte Fähigkeit des Modells aufzeigt. Man sieht jedoch, dass das Faster R-CNN mit Bells and Whistles mit zunehmendem Okklusionsverhältnis signifikant schlechter funktioniert, und sich in der Tat für kein Okklusionsverhältnissniveau verbessert. Dieses Resultat erklärt die Bedeutung eines größeren rezeptiven Felds, das größeren Kontext für das Objekt auf dem zu verarbeitenden Bild bzw. der zu verarbeitenden Abbildung bereitstellt und daher bessere Okklusionshandhabung erlaubt.

[0116] Die oben stehenden Feststellungen können auch hinsichtlich der mittleren Präzision erfolgen. Diese Resultate sind in Tabelle 1 bereitgestellt. In dem Fall des UA-DTERAC, ist das Verhältnis der Objektgröße zur Bildgröße im Vergleich dazu ziemlich klein, was in einem typischen Bild eines PASCAL VOC-Datensatzes beobachtet wird. Die Aufgabe wird leichter, wenn mehr geeignete Ankermaßstäbe erlaubt werden, um die Regressionsparameter zu lernen. Man kann bemerken, dass das Hinzufügen zusätzlicher Maßstäbe für die Anker-Rechtecke zum Erfassen kleinerer Fahrzeuge den AP signifikant auf 68,8 % verbessert.

[0117] Wie oben beschrieben, kann das Erhöhen der Auflösung der Funktionskarten mehr Vorschläge erlauben und kann weiter beim Erfassen kleinerer Objekte helfen. Natürlich kann das die Rechenzeit erhöhen. Des Weiteren kann man bemerken, dass die vorgeschlagene Erweiterung auf das RPN durch Eingliedern von Maßstablayout, das geometrische Vorschläge bietet, den AP um mehr als 6 % anhebt. Dieses Resultat fasst den Gewinn zusammen, den man in den **Fig. 6** und **Fig. 7** für unterschiedliche Objektmaßstäbe und Okklusionsniveaus bemerken kann. Schließlich kann Mehrstufentraining, wie oben beschrieben, den AP weiter auf beachtliche 89,9 % verbessern.

[0118] Letztendlich kann auch der vorgeschlagene Ansatz unter Verwenden des größeren VGG-16-Netzwerks bewertet werden. Diese Resultate sind in Tabelle 2 gezeigt. Insgesamt können leicht bessere Resultate erzielt werden, verlieren aber die Hälfte der Rahmenrate. Kooperativ kann beobachtet werden, dass kleinere Netzwerke (VGG-M) signifikant mehr von den geometrischen Vorschlägen profitieren. Man kann bemerken, dass die Maßstabänderungen das größere Modell (VGG-16) signifikant besser funktionieren lassen als das kleinere (VGG-M). Die vorgeschlagenen geometrischen Vorschläge verringern jedoch die Leistungsspalte wieder.

Tabelle 2

Verfahren	AP (%)	Geschwindigkeit (fps)
Faster RCNN	66,8	5,5
+ Extraanker	71,8	5
+ hochauflösende Funktionskarten	77,9	4
+ GP -> GP-FRCNN	80,3	4
+ Mehrstufentraining	82,1	4

Der UA-DETRAC-Challenge

[0119] Unten werden die Resultate für den UA-DETRAC-Erfassungs-Challenge berichtet, und im Zeitpunkt der Verfassung sind alle derzeit auf der Website hinsichtlich der mittleren Präzision verfügbaren Resultate besser.

[0120] Um die Modelle zu trainieren, wird der vollständige UA-DETRAC-Train-Set (**60** Videosequenzen oder Bilder) verwendet. PASCAL VOC 2007 und 2012 Trainval-Bildsätze sind ebenfalls enthalten, wie es in der Validierungsphase erfolgt. Man kann beobachten, dass ein Teil der Verkehrsszenarien in dem UA-DETRAC-

Test-Satz relativ dichter und überladener sind als vergleichsweise die Videosequenzen in dem Trainingssatz. Die meisten Resultate, die für den Online-Challenge erhalten werden, stimmen jedoch mit diesen Bewertungen während der Validierungsphase überein. Insgesamt kann es um beeindruckende 19,5 % hinsichtlich des AP im Vergleich zu dem Vanilla Faster R-CNN verbessert werden, das heißt von 57,08 % auf 67,57 %. Man kann bemerken, dass die Auswirkung des Hinzufügens geometrischer Vorschläge nicht so stark ist wie während der Validierung beobachtet. Man vermutet, dass dies auf die Tatsache zurückzuführen ist, dass eine große Anzahl von Objekten kleinen Maßstabs während Online-Bewertung ignoriert werden. Das kann an den markierten ignorierten Erkennungen in dem Bild bzw. der Abbildung, das/die zu verarbeiten ist, liegen.

[0121] Diese Resultate sind in Tabelle 3 gezeigt.

Tabelle 3

Verfahren	AP (%)	Geschwindigkeit (fps)
Faster RCNN	57,08	5, 5
+ Extraanker	65,03	5
+ hochauflösende Funktionskarten	72,80	4
+ GP → GP-FRNCNN	73,88	4
+ Mehrstufentraining → GP-FRCNNm	76,57	4

[0122] **Fig. 1** zeigt in einer schematischen Ansicht eine **Abb. 16**, die zu verarbeiten ist, mit allen Bereichsvorschlägen **18**, die durch einen Bereichsvorschlaggenerator erzeugt werden. Die zu verarbeitende **Abb. 16** zeigt ferner einen Verkehrsweg **14**, der eine Vielzahl von Fahrbahnen hat, auf welchen Fahrzeuge **12** als Verkehrsteilnehmer fahren.

[0123] Wie aus **Fig. 1** ersichtlich ist, werden drei Rechtecke **20**, **22**, **24** bereitgestellt. Das Rechteck **20** befindet sich in einem oberen Abschnitt der zu verarbeitenden **Abb. 16**, so dass dieses Rechteck zu groß ist, weil die Verkehrsteilnehmer, die in diesem Abschnitt der zu verarbeitenden **Abb. 16** zu erfassen sind, viel kleiner sind. Das Rechteck **24** in dem unteren Abschnitt der zu verarbeitenden **Abb. 16** ist hingegen zu klein, um einen Verkehrsteilnehmer bzw. ein Fahrzeug zu enthalten. Im Gegensatz dazu ist ein Rechteck **22** in dem unteren Abschnitt der zu verarbeitenden **Abb. 16** angepasst, um einen Verkehrsteilnehmer bzw. ein Fahrzeug zu enthalten. Die Größe des Rechtecks stimmt mit dem Fahrzeug **12** überein. Die zu verarbeitende **Abb. 16** ist eine einzelne Abbildung eines Videostreams einer Kamera, die nicht gezeigt ist.

[0124] **Fig. 2** zeigt eine dreidimensionale Skizze, die Filtervorschläge zeigt, die die geschätzten Maßstabfilter bzw. die Bildmaßstabfunktion verwenden. Eine vertikale Achse **28** entspricht der Größe des Objekts in Einheiten von Pixeln². Die Achsen **30**, **32** verweisen auf Positionen. Eine Ebene **26** definiert eine Fläche, wo Flächenpunkte **28** angeordnet sein können. Die Punkte **28** verweisen auf Positionen zuverlässiger Erfassungen in der Abbildung oder dem Bild **16**, die/das zu verarbeiten ist, die verwendet werden können, um die Maßstabfunktionen zu schätzen.

[0125] Wie man sieht, definiert eine Linie den Horizont **50**. **Fig. 3** zeigt eine Projektion der Ebene **26** in die Richtung der Achse **28** von oben. Wie man sieht, liegt der Horizont **50** parallel zu der Achse **32**, die die Achse **30** an dem Wert **100** kreuzt. Die Punkte **28** befinden sich nur unter dem Horizont **50**, das heißt an Werten größer als **100** der Achse **30**. In dem Bereich des Horizonts **50**, befindet sich ein kleiner Maßstab, wobei sich in dem Bereich der Achse **32** ein großer Maßstab befindet.

[0126] **Fig. 4** zeigt eine Wirkung der Erfindung, nämlich dass Bereichsvorschläge viel präziser sind. Zuerst werden sie nur unter dem Horizont **50** bereitgestellt. Außerdem ist die Größe der Bereichsvorschläge **34** besser an den Maßstab angepasst. Das resultiert in präziserem und schnellerem Erfassen von Fahrzeugen **12** durch die Rechtecke **36**. Der Grund dafür ist, dass die Rechtecke **36** derart bemessen sind, dass sie in Abhängigkeit von einem Maßstab und einer Position in der zu verarbeitenden **Abb. 16** besser an Größen von Fahrzeugen **12** angepasst sind. Objektvorschläge werden, während sie sich dem Horizont **50** nähern, kleiner.

[0127] **Fig. 5** zeigt in einer beispielhaften Ausführungsform ein schematisches Flussdiagramm eines Verfahrens **10** zum Erfassen von Verkehrsteilnehmern **12**, hier von Fahrzeugen, auf einem Verkehrsweg **14** in einer

Sequenz aufeinanderfolgender Abbildungen, die durch eine Kamera bereitgestellt werden, die die Sequenz von Abbildungen des Verkehrswegs **14** aufzeichnet. Die Sequenz von Abbildungen wird gegenwärtig durch einen Videostream bereitgestellt.

[0128] Bei Schritt **40** wird eine erste Abbildung der Sequenz aufeinanderfolgender Abbildungen ausgewählt, und die erste Abbildung wird als eine **Abb. 16**, die zu verarbeiten ist, definiert. Dann wird bei Schritt **42** ein Bereichsvorschlaggenerator auf die **Abb. 16**, die zu verarbeiten ist, angewandt, der durch selektive RPN(Faster R-CNN)-Suche, Gleitfenster und dergleichen bereitgestellt werden kann. Der Bereichsvorschlaggenerator erzeugt eine Vielzahl von Bereichsvorschlägen **18** für mögliche Objekte, die in der zu verarbeitenden **Abb. 16** aufgezeichnet sind.

[0129] Dann, bei Schritt **44**, wird geprüft, ob Filter verfügbar sind. Falls nicht, setzt das Verfahren mit Schritt **46** fort, indem alle Bereichsvorschläge **18** des Bereichsvorschlaggenerators, der in **Fig. 1** ebenfalls gezeigt ist, präsentiert werden. Dann setzt das Verfahren mit Schritt **72** fort, wobei Objekterfassung für alle Bereichsvorschläge **18** bereitgestellt wird, um die Verkehrsteilnehmer **12** zu erfassen, wobei ein vorab definiertes Vertrauensniveau berücksichtigt wird. In dem Verfahrensschritt **74** werden zuverlässige Erfassungen (mit hohem Resultat) gesammelt.

[0130] Bei Schritt **76** wird geprüft, ob ausreichend Erfassungen erzielt wurden, um eine Filterschätzung zu erlauben. Falls nicht (n), setzt das Verfahren mit Schritt **40** fort und wählt eine andere Abbildung der Sequenz aufeinanderfolgender Abbildungen aus, um die Anzahl erzielter Erfassungen zu verbessern.

[0131] Falls in Schritt **76** ausreichend Erfassungen erzielt wurden (y), setzt das Verfahren mit Schritt **78** fort. Bei Schritt **78** werden Filter geschätzt, wie zum Beispiel Maßstabfilter, Seitenverhältnisfilter, ein Horizont und dergleichen. Die geschätzten Filter werden geliefert, um eine Vorschlagfilterung, wie unten besprochen, zu steuern. Wenn die Filter in Schritt **78** geschätzt werden, geht das Verfahren weiter zu Schritt **40**, wie oben beschrieben.

[0132] Wenn bei Schritt **44** ermittelt wird, dass Filter verfügbar sind (y), geht das Verfahren weiter zu Schritt **48**, indem es Vorschlagfiltern wie oben besprochen anwendet. Das Verfahren setzt dann mit Schritt **70** fort, indem gefilterte Vorschläge auf die zu verarbeitende **Abb. 16** angewandt werden, was ebenfalls in **Fig. 4** gezeigt ist. Dann geht das Verfahren, wie oben besprochen, weiter zu Schritt **73**.

[0133] Optional können bei Schritt **80** Kamerakalibrierungsinformationen bzw. -daten empfangen und zu Schritt **78** geliefert werden, um die Filterschätzung zu verstärken.

[0134] Wie durch diese Offenbarung gezeigt, hat der vorgeschlagene GP-FRCNN-Ansatz das Potenzial, die Einstufungsversagen des grundlegenden RPN zu überkommen, und kann als ein Resultat mehr oder weniger ähnliche Leistung unabhängig von dem Maßstab des Objekts erzielen. Die Feststellungen der Erfinder deuten auch an, dass man nicht einfach das geometrische Layout aufnehmen kann, um Vorschläge neu einzustufen, und dann gewünschte Verbesserungen erwarten kann, stattdessen wird eine Anzahl von Maßstabänderungen bevorzugt bereitgestellt.

[0135] Falls gewünscht, können die unterschiedlichen Funktionen und Ausführungsformen, die hier besprochen sind, in einer unterschiedlichen abweichenden Reihenfolge und/oder gleichzeitig miteinander auf diverse Arten ausgeführt werden. Ferner kann nach Wunsch eine oder mehrere der oben beschriebenen Funktionen und/oder Ausführungsformen optional sein oder bevorzugt auf eine willkürliche Art kombiniert werden.

[0136] Obwohl diverse Aspekte der Erfindung in den unabhängigen Ansprüchen dargelegt sind, umfassen andere Aspekte der Erfindung andere Kombinationen von Merkmalen aus den beschriebenen Ausführungsformen und/oder den abhängigen Ansprüchen, mit den Merkmalen der unabhängigen Ansprüche, und nur die Kombination, die in den Ansprüchen dargelegt ist.

[0137] Es wird auch vermerkt, dass, obwohl oben Stehendes beispielhafte Ausführungsformen der Erfindung beschreibt, diese Beschreibung nicht als den Schutzbereich einschränkend betrachtet werden sollte. Vielmehr gibt es mehrere Variationen und Änderungen, die ohne Abweichen von dem Schutzbereich der vorliegenden Erfindung, wie er in den abhängigen Ansprüchen geschützt wird, vorgenommen werden können.

Bezugszeichenliste

10	Verfahren
12	Fahrzeug
14	Straße
16	Abbildung
18	Bereichsvorschlag
20	Rechteck
22	Rechteck
24	Rechteck
26	Ebene
28	Achse
30	Achse
32	Achse
34	Bereichsvorschlag
36	Rechteck
38	Punkte
40	Schritt
42	Schritt
44	Schritt
46	Schritt
48	Schritt
50	Horizont
52	Skizze
54	Ordinate
56	Abszisse
58	Bin
60	Bin
62	Bin
64	Bin
66	Skizze
68	Abszisse
70	Schritt
72	Schritt
73	Schritt
74	Schritt
76	Schritt
78	Schritt

80 Schritt

y ja

n nein

Patentansprüche

1. Verfahren (10) zum Erfassen von Verkehrsteilnehmern (12) auf einem Verkehrsweg (14) in einer Abbildung, die durch eine Kamera bereitgestellt wird, die die Abbildung aufnimmt, wobei das Verfahren Folgendes umfasst:

- Erzeugen (42) einer Vielzahl von Bereichsvorschlägen (18) für mögliche Objekte, die in der Abbildung (16) aufgezeichnet sind, durch Anwenden eines Bereichsvorschlaggenerators auf die Abbildung (16);
- Bereitstellen von Objekterfassung (72) für alle Bereichsvorschläge (18) durch Anwenden eines Objekterfassungsgeräts auf alle Bereichsvorschläge (18), um den Verkehrsweg (14) und/oder die Verkehrsteilnehmer (12) durch Einstufen zu erfassen, wobei ein vorbestimmtes Vertrauensniveau berücksichtigt wird; und
- Ausgeben von Erfassungsdaten, die von der Objekterfassung empfangen werden, für den erfassten Verkehrsweg (14) und/oder die erfassten Verkehrsteilnehmer (12), **gekennzeichnet durch**:
- Bereitstellen eines Filters (48) für die Bereichsvorschläge (18) vor dem Schritt des Bereitstellens von Objekterfassung, wobei das Filtern basierend auf jeweiligen Filterdaten ausgeführt wird, die basierend auf einer Relevanz der Bereichsvorschläge (18) in Zusammenhang mit den Verkehrsteilnehmern (12) und/oder dem Verkehrsweg (14) geschätzt werden.

2. Verfahren nach Anspruch 1, wobei Maßstabinformationen vor dem Schritt des Ausführens der Objekterfassung bereitgestellt werden, wobei Maßstabinformationen aus Langzeitbeobachtung der auf der Bodenfläche erfassten Fahrzeuge erfasst werden.

3. Verfahren nach Anspruch 1 oder 2, wobei die Filterdaten basierend auf einer Szene geschätzt werden, die durch die Abbildung (16), die zu verarbeiten ist, aufgezeichnet wird, insbesondere der Verkehrsweg (14), der in der Szene enthalten ist.

4. Verfahren nach einem der Ansprüche 1 bis 3, wobei die Filterdaten einen Beobachtungsbereich für mögliche Objekte, die durch die Objekterfassung als Verkehrsteilnehmer (12) eingestuft werden, bereitstellen.

5. Verfahren nach einem der Ansprüche 1 bis 4, wobei die Filterdaten in Abhängigkeit von Daten der Objekterfassung aktualisiert werden.

6. Verfahren nach einem der Ansprüche 1 bis 5, wobei Deep-Learning-Verfahren und/oder künstliche Intelligenz verwendet wird.

7. Verfahren nach einem der Ansprüche 1 bis 6, wobei die Ausgangsdaten von mindestens zwei Abbildungen der Sequenz von Abbildungen verarbeitet werden, um mindestens eine Bewegungsbahn für einen bestimmten der erfassten Verkehrsteilnehmer (12) zu bestimmen.

8. Verfahren nach einem der Ansprüche 1 bis 7, wobei die Filterdaten geschätzt werden, indem Kamerakalibrierungsdaten berücksichtigt werden.

9. Vorrichtung zum Erfassen von Verkehrsteilnehmern (12) auf einem Verkehrsweg (14) in einer Sequenz aufeinanderfolgender Abbildungen, wobei die Vorrichtung Folgendes umfasst:

- mindestens eine Kamera, die die Sequenz von Abbildungen des Verkehrswegs (14) aufnimmt, und
- ein Gerät, das konfiguriert ist, um
- eine erste Abbildung der Sequenz aufeinanderfolgender Abbildungen auszuwählen (40) und die erste Abbildung als eine Abbildung (16), die zu verarbeiten ist, zu definieren;
- eine Vielzahl von Bereichsvorschlägen (18) für mögliche Objekte, die in der zu verarbeitenden Abbildung (16) aufgezeichnet sind, durch Anwenden eines Bereichsvorschlaggenerators auf die zu verarbeitende Abbildung (16) zu erzeugen (42);
- Objekterfassung (72) für alle Bereichsvorschläge (18) bereitzustellen, um den Verkehrsweg (14) und/oder die Verkehrsteilnehmer (12) durch Klassifizieren unter Berücksichtigung eines vordefinierten Vertrauensniveaus zu erfassen;
- Erfassungsdaten, die von der Objekterfassung für den erfassten Verkehrsweg (14) und/oder die erfassten Verkehrsteilnehmer (12) empfangen wurden, auszugeben; und

- durch Auswählen einer anderen Abbildung der Sequenz aufeinanderfolgender Abbildungen als die Abbildung (16), die zu verarbeiten ist, fortzusetzen;

dadurch gekennzeichnet, dass

das Gerät ferner konfiguriert ist, um

- ein Filtern (48) für die Bereichsvorschläge (18) vor dem Schritt des Bereitstellens von Objekterfassung (72) bereitzustellen, wobei das Filtern (48) basierend auf jeweiligen Filterdaten ausgeführt wird, die basierend auf einer Relevanz der Bereichsvorschläge (18) in Zusammenhang mit den Verkehrsteilnehmern (12) und/oder dem Verkehrsweg (14) geschätzt werden.

10. Verfahren zum Bereitstellen von Verkehrsführung, umfassend:

- Erfassen von Verkehrsteilnehmern (12) auf einem Verkehrsweg (14) in einer Sequenz aufeinanderfolgender Abbildungen, die durch eine Kamera bereitgestellt wird, die die Sequenz von Abbildungen des Verkehrswegs (14) aufnimmt;

- Bestimmen einer verwendeten Kapazität des Verkehrswegs (14), die durch Erfassen der Verkehrsteilnehmer (12) bereitgestellt wird;

- Identifizieren individueller tatsächlicher Geschwindigkeit und/oder Position jedes der Verkehrsteilnehmer (12), um jeweilige individuelle tatsächliche Bahnen der Verkehrsteilnehmer (12) zu bestimmen;

- Bestimmen mindestens einer jeweiligen Nennbahn für jeden der Verkehrsteilnehmer (12); und

- Kommunizieren der bestimmten Nennbahnen zu den jeweiligen Verkehrsteilnehmern (12); **dadurch gekennzeichnet**, dass

- das Erfassen von Verkehrsteilnehmern (12) auf dem Verkehrsweg (14) durch ein Verfahren nach einem der Ansprüche 1 bis 8 bereitgestellt wird.

11. Verfahren nach Anspruch 10, wobei mindestens einer der Verkehrsteilnehmer (12) mindestens teilweise automatisch angetrieben ist.

12. Verfahren nach Anspruch 10 oder 11, wobei mindestens einer der Verkehrsteilnehmer (12) eine gewünschte Bahn anfordert, was Folgendes veranlasst:

- Bestimmen einer bestimmten Nennbahn für den Verkehrsteilnehmer (12); und

- Kommunizieren der bestimmten Nennbahn zu dem Verkehrsteilnehmer (12).

13. Verfahren nach einem der Ansprüche 10 bis 12, das das Prüfen einer bestimmten Nennbahn durch Simulation für den Verkehrsteilnehmer (12) umfasst.

14. Verfahren nach einem der Ansprüche 10 bis 13, das das Abfragen einer tatsächlichen Geschwindigkeit und/oder Position mindestens eines der Verkehrsteilnehmer (12) umfasst, um seine tatsächliche Bahn zu bestimmen.

15. Verfahren nach einem der Ansprüche 10 bis 14, wobei in Abhängigkeit von der verwendeten Kapazität des Verkehrswegs (14) die Geschwindigkeit der Verkehrsteilnehmer (12) automatisch angepasst wird.

16. Verfahren nach einem der Ansprüche 10 bis 15, das das automatische Anpassen der Geschwindigkeit mindestens eines der Verkehrsteilnehmer (12) umfasst, während sich der Verkehrsteilnehmer (12) einem vorbestimmten Durchfahrtmuster für den Verkehrsweg (14) nähert.

17. Verkehrsleitsystem zum Bereitstellen von Verkehrsführung, umfassend:

- mindestens eine Vorrichtung zum Erfassen von Verkehrsteilnehmern (12) auf einem Verkehrsweg (14) in einer Sequenz aufeinanderfolgender Abbildungen, und

- ein Gerät, das konfiguriert ist, um

- Verkehrsteilnehmer (12) auf dem Verkehrsweg (14) in der Sequenz aufeinanderfolgender Abbildungen zu erfassen;

- eine verwendete Kapazität des Verkehrswegs (14), die durch die Verkehrsteilnehmer (12) bereitgestellt wird, zu bestimmen;

- individuelle tatsächliche Geschwindigkeit und/oder Position jedes der Verkehrsteilnehmer (12) zu identifizieren, um jeweilige individuelle tatsächliche Bahnen zu bestimmen;

- mindestens eine jeweilige Nennbahn für jeden der Verkehrsteilnehmer (12) zu bestimmen; und

- die bestimmten Nennbahnen zu den jeweiligen Verkehrsteilnehmern (12) zu kommunizieren; **dadurch gekennzeichnet**, dass

die Vorrichtung zum Erfassen von Verkehrsteilnehmern (12) auf einem Verkehrsweg (14) gemäß Anspruch 9 konfiguriert ist.

18. Computerprogrammprodukt, das ein Programm für eine Verarbeitungsvorrichtung aufweist, das Softwarecodeabschnitte eines Programms zum Ausführen der Schritte eines Verfahrens nach einem der Ansprüche 1 bis 8 und/oder zum Ausführen der Schritte eines Verfahrens nach einem der Ansprüche 10 bis 16 umfasst, wenn das Programm auf der Verarbeitungsvorrichtung ausgeführt wird.

Es folgen 5 Seiten Zeichnungen

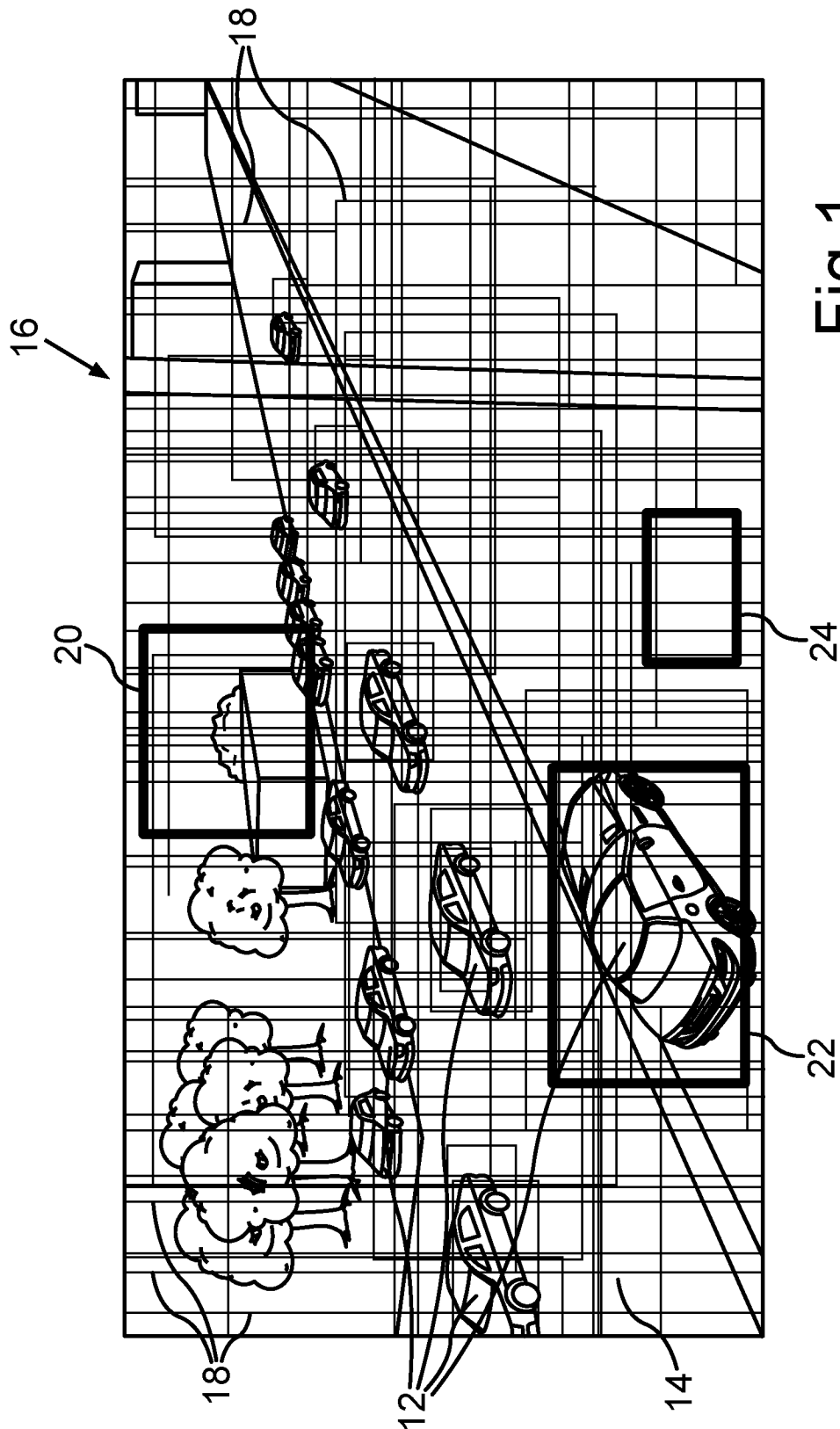
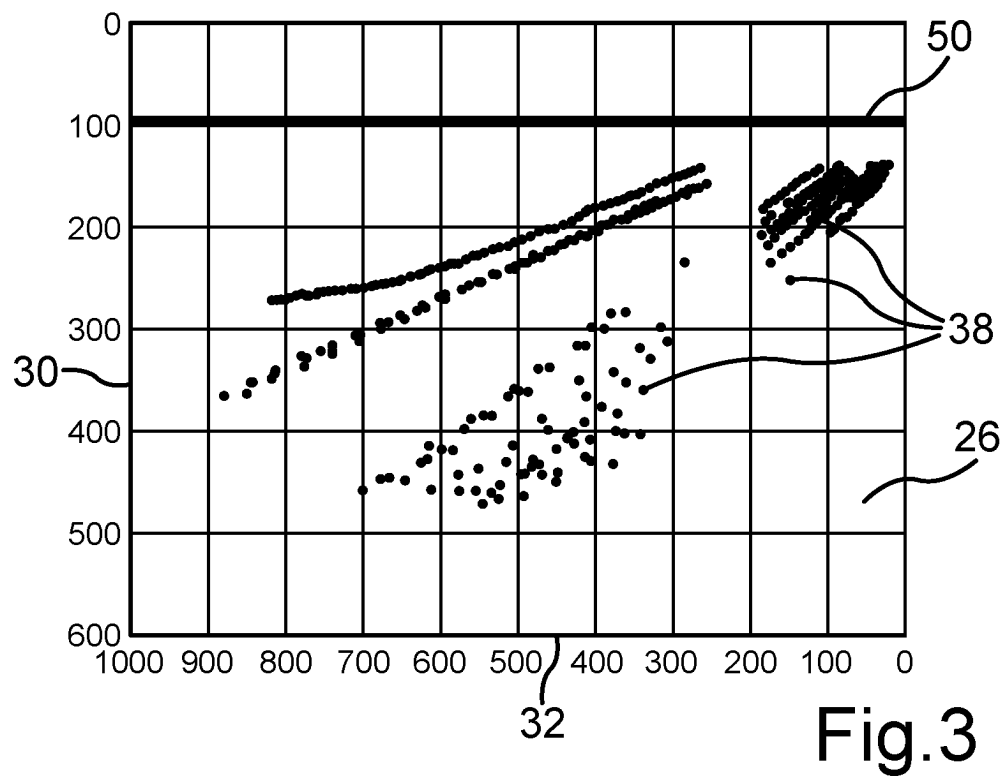
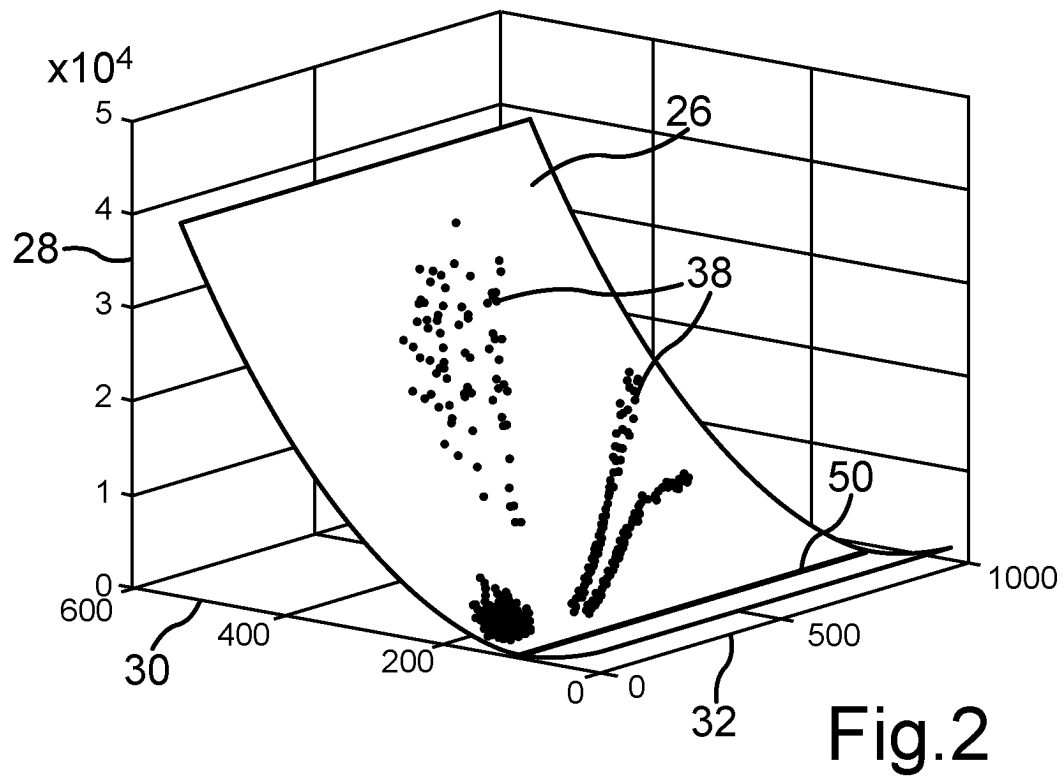


Fig. 1



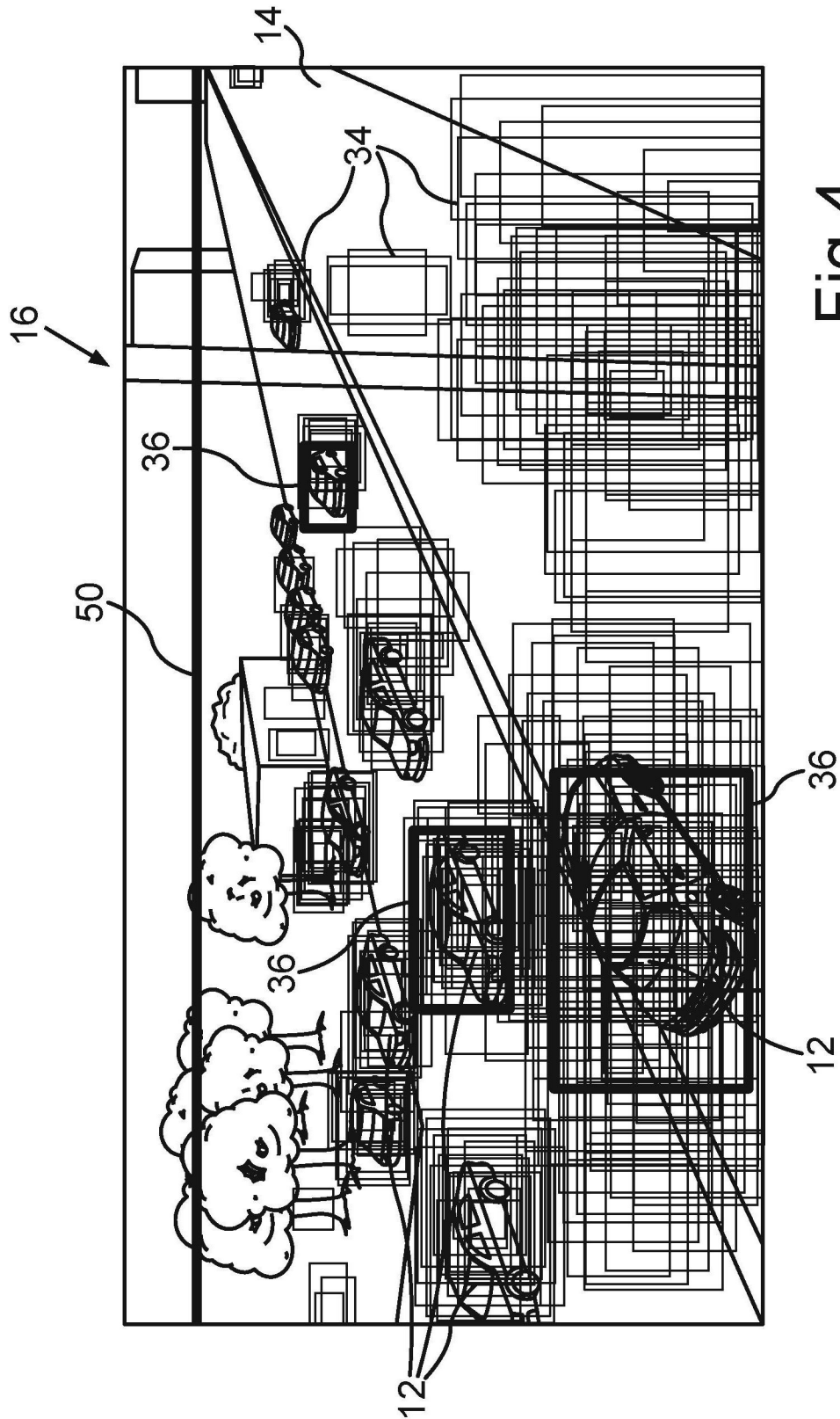


Fig. 4

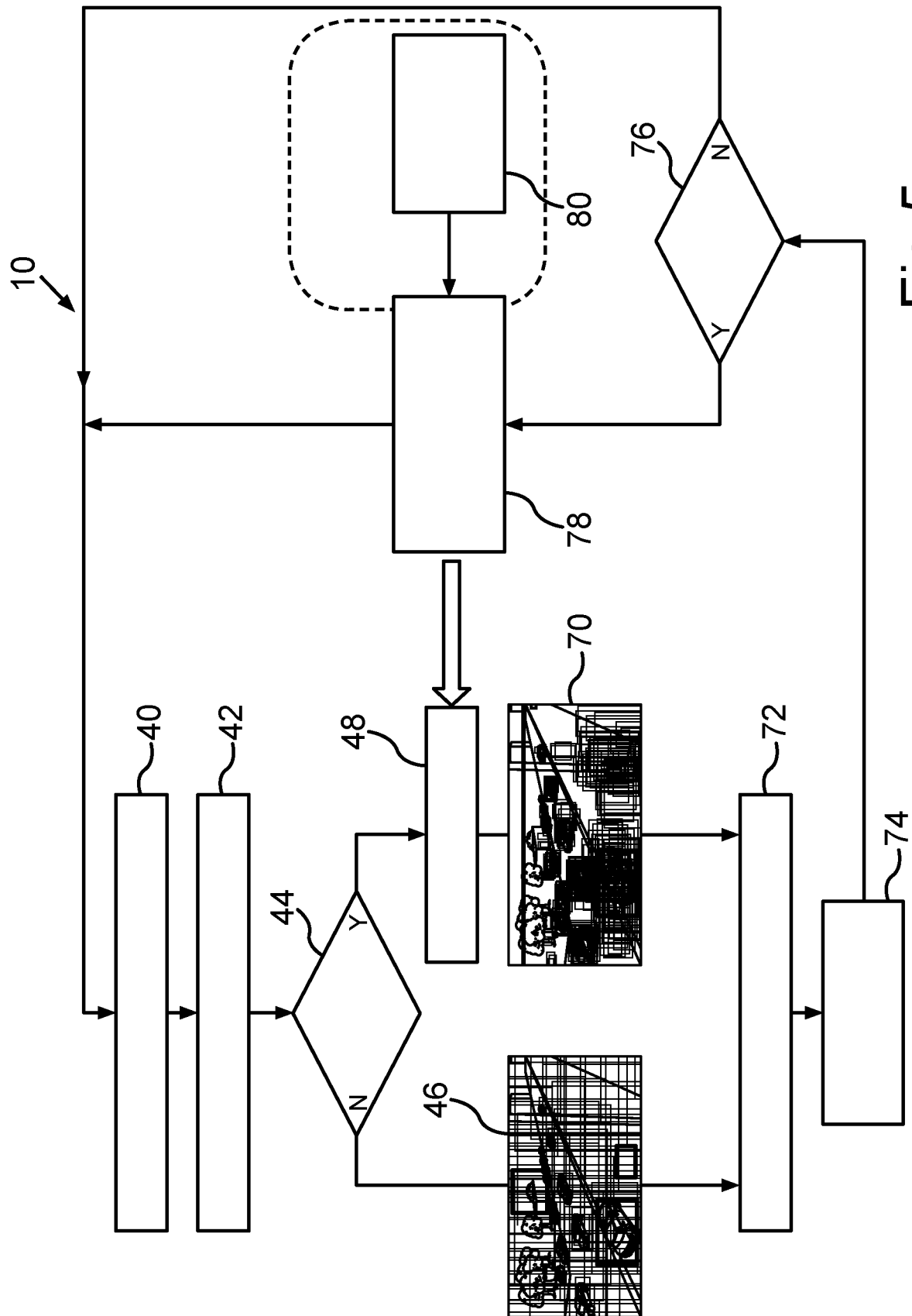


Fig.5

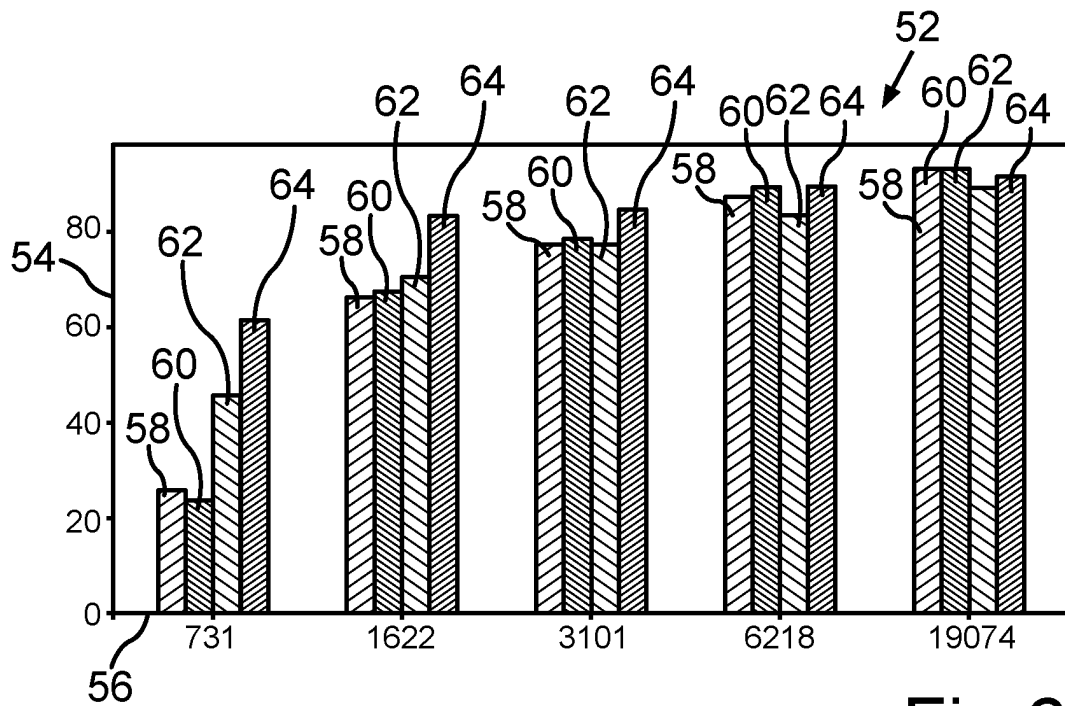


Fig. 6

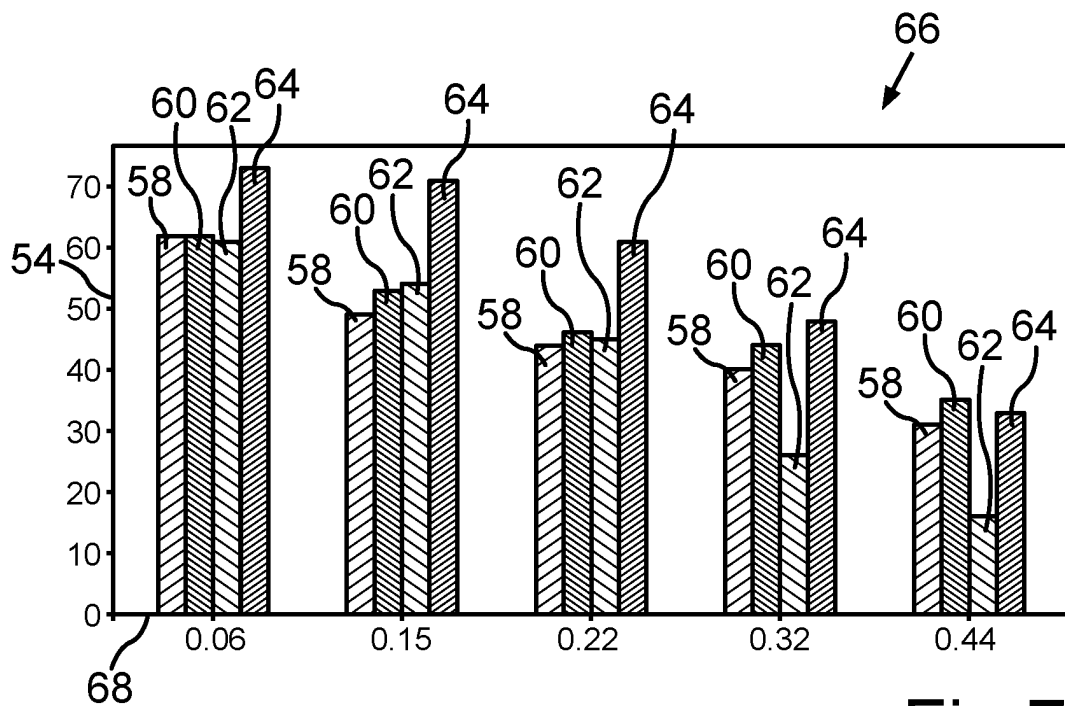


Fig. 7