



(10) **DE 11 2017 001 311 T5** 2018.11.29

(12)

Veröffentlichung

der internationalen Anmeldung mit der
(87) Veröffentlichungs-Nr.: **WO 2017/156628**
in der deutschen Übersetzung (Art. III § 8 Abs. 2
IntPatÜG)
(21) Deutsches Aktenzeichen: **11 2017 001 311.9**
(86) PCT-Aktenzeichen: **PCT/CA2017/050334**
(86) PCT-Anmeldetag: **14.03.2017**
(87) PCT-Veröffentlichungstag: **21.09.2017**
(43) Veröffentlichungstag der PCT Anmeldung
in deutscher Übersetzung: **29.11.2018**

(51) Int Cl.: **G06F 15/18** (2006.01)
G06K 9/46 (2006.01)
G06K 9/62 (2006.01)
G06K 9/80 (2006.01)
G06T 7/194 (2017.01)

(30) Unionspriorität:
62/309,777 **17.03.2016** **US**

(71) Anmelder:
**Avigilon Corporation, Vancouver, British
Columbia, CA**

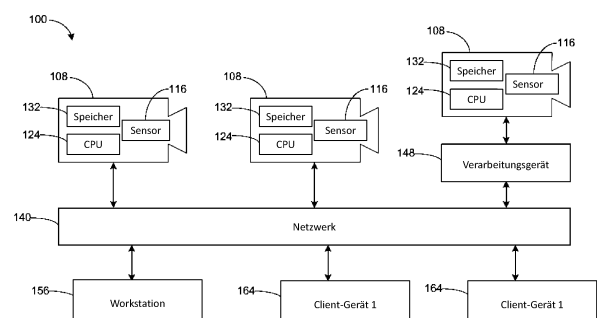
(74) Vertreter:
**Zimmermann & Partner Patentanwälte mbB,
80331 München, DE**

(72) Erfinder:
Shrivastava, Ashish, Vancouver, BC, CA

Die folgenden Angaben sind den vom Anmelder eingereichten Unterlagen entnommen.

(54) Bezeichnung: **System und Verfahren zum Trainieren eines Objektklassifikators durch maschinelles Lernen**

(57) Zusammenfassung: System und Verfahren zum Trainieren eines computerimplementierten Objektklassifikators umfasst Detektieren eines visuellen Vordergrund-Objekts innerhalb einer Teilregion einer Szene, Bestimmen eines Hintergrundmodells der Teilregion der Szene, wobei das Hintergrundmodell die Teilregion repräsentiert, wenn ein visuelles Vordergrund-Objekt in dieser Teilregion nicht vorhanden ist, und Trainieren des Objektklassifikators durch computerimplementiertes maschinelles Lernen unter Verwendung des Hintergrundmodells der Teilregion als ein negatives Trainingsbeispiel.



Beschreibung

FELD

[0001] Das vorliegende Thema bezieht sich auf die Klassifizierung von visuellen Objekten, insbesondere auf das Training eines computerimplementierten Objektklassifikators anhand von Hintergrundmodellen von erkannten visuellen Objekten im Vordergrund als negative Trainingsbeispiele.

WERDEGANG

[0002] Computerimplementierte visuelle Objektklassifizierung, auch Objekterkennung genannt, bezieht sich auf die Klassifizierung von visuellen Darstellungen von realen Objekten, die in Standbildern oder Bewegungsvideos, die von einer Kamera aufgenommen wurden, enthalten sind. Durch die visuelle Objektklassifizierung wird jedes in den Standbildern oder Bewegungsvideos gefundene visuelle Objekt nach seinem Typ (z.B. Mensch, Fahrzeug, Tier) klassifiziert.

[0003] Automatisierte Sicherheits- und Überwachungssysteme verwenden in der Regel Videokameras oder andere Bilderfassungsgeräte oder Sensoren zur Erfassung von Bilddaten. In den einfachsten Systemen werden die durch die Bilddaten repräsentierten Bilder zur gleichzeitigen Überprüfung durch das Sicherheitspersonal angezeigt und/oder nach einem Sicherheitsverstoß zur späteren Verwendung aufgezeichnet. In diesen Systemen wird die Aufgabe, visuelle Objekte von Interesse zu erkennen und zu klassifizieren, von einem menschlichen Beobachter ausgeführt. Ein wesentlicher Fortschritt entsteht, wenn das System selbst in der Lage ist, die Objekterkennung und -klassifizierung teilweise oder vollständig durchzuführen.

[0004] In einem typischen Überwachungssystem kann es beispielsweise darum gehen, Objekte wie Menschen, Fahrzeuge, Tiere usw. zu erkennen, die sich durch die Umwelt bewegen. Verschiedene Objekte können unterschiedliche Bedrohungen oder Alarmstufen darstellen. Zum Beispiel kann ein Tier in der Szene normal sein, aber ein Mensch oder ein Fahrzeug in der Szene kann einen Alarm auslösen und die sofortige Aufmerksamkeit eines Sicherheitspersonals erfordern. Die automatisierte computerimplementierte Erkennung und Klassifizierung von Objekten in den Bildern, die durch die von den Kameras erfassten Bilddaten dargestellt werden, kann die Aufgabe der Überprüfung des Sicherheitspersonals sowie die Verbesserung der Aufzeichnung von Bilddaten erheblich erleichtern.

ZUSAMMENFASSUNG

[0005] Die hier beschriebenen Ausführungsformen stellen in einem Aspekt eine Methode zum Training eines computerimplementierten Objektklassifikators bereit. Das Verfahren umfasst Erkennen eines visuellen Vordergrund-Objekts innerhalb eines Teilbereichs einer Szene, Bestimmen eines Hintergrundmodells des Teilbereichs der Szene, wobei das Hintergrundmodell den Teilbereich darstellt, wenn ein visuelles Vordergrund-Objekt davon abwesend ist, und Trainieren des Objektklassifikators durch computerimplementiertes maschinelles Lernen unter Verwendung des Hintergrundmodells des Teilbereichs als negatives Trainingsbeispiel.

[0006] Die hier beschriebenen Ausführungsformen stellen in einem weiteren Aspekt einen computerimplementierten Objektklassifikator bereit. Das System umfasst einen Prozessor, eine computerlesbare Speichervorrichtung, die Programmbefehle speichert, die, wenn sie vom Prozessor ausgeführt werden, das System veranlassen, Operationen durchzuführen, die Erkennen eines visuellen Vordergrund-Objekts innerhalb eines Teilbereichs einer Szene, Bestimmen eines Hintergrundmodells des Teilbereichs der Szene, wobei das Hintergrundmodell den Teilbereich darstellt, wenn ein visuelles Vordergrund-Objekt davon abwesend ist, und Trainieren des Objektklassifikators durch computerimplementiertes maschinelles Lernen unter Verwendung des Hintergrundmodells des Teilbereichs als ein negatives Trainingsbeispiel umfassen.

[0007] Gemäß einiger exemplarischer Ausführungsformen weisen die Methoden und/oder Systeme ferner das Trainieren des Objektklassifikators durch maschinelles Lernen am detektierten visuellen Vordergrundobjekts als positives Trainingsbeispiel auf.

[0008] Gemäß einiger exemplarischer Ausführungsformen umfasst Bestimmen des Hintergrundmodells des Teilbereichs der Szene Auswählen eines historischen Bildrahmens, der aufgenommen ist, als irgendein Vordergrund-Objekt in einem Teilbereich des historischen Bildrahmens fehlt, der dem Teilbereich der Szene entspricht, und das Ausschneiden des Teilbereichs, der dem Teilbereich der Szene entspricht, aus dem histori-

schen Bildrahmen, wobei der ausgeschnittene Bildrahmen das Hintergrundmodell des Teilbereichs der Szene ist.

[0009] Gemäß einiger exemplarischer Ausführungsformen umfasst Bestimmen des Hintergrundmodells des Teilbereichs der Szene Bestimmen, innerhalb jedes einer Vielzahl von historischen Bildrahmen, eines oder mehrere Teilbereiche, die frei von irgendwelchen Vordergrund-Objekten sind, aggregieren der einen oder mehreren Teilregionen aus der Vielzahl der historischen Bilder um ein vollständiges Hintergrundbild, das die gesamte Szene darstellt, und aus dem vollständigen Hintergrundbild einen Teilbereich entsprechend des Teilbereichs der Szene ausschneiden, wobei das ausgeschnittene vollständige Hintergrundbild das Hintergrundmodell des Teilbereichs der Szene ist.

[0010] Gemäß einiger exemplarischer Ausführungsformen umfasst Aggregieren des einen oder der mehreren Teilbereichen aus der Vielzahl der historischen Bilder Zusammenfügen des einen oder der mehreren Teilbereiche um ein Bild zu bilden, das die gesamte Szene darstellt.

[0011] Gemäß einiger exemplarischer Ausführungsformen wird der Objektklassifikator speziell für eine aktuelle Szene trainiert.

[0012] Gemäß einiger exemplarischer Ausführungsformen wird bei einem Wechsel der aktuellen Szene zu einer neuen Szene auf den Objektklassifikator ohne das für die aktuelle Szene spezifische Trainieren zurückgegriffen und der Objektklassifikator wird durch maschinelles Lernen mit Hilfe von Hintergrundmodellen aus der neuen Szene trainiert.

[0013] Gemäß einiger exemplarischer Ausführungsformen wird der Objektklassifikator teilweise unter Verwendung von überwachtem Lernen vorbereitet.

[0014] Gemäß einiger exemplarischer Ausführungsformen wird das computerimplementierte maschinelle Lernen aus Faltungsneuronalen Netzen, Support-Vektor-Maschinen, Entscheidungsbäumen, Random Forests und Kaskaden-Klassifikatoren ausgewählt.

[0015] Gemäß einiger exemplarischer Ausführungsformen weisen die Verfahren und/oder die Systeme ferner das Trainieren des Objektklassifikators durch computerimplementiertes maschinelles Lernen unter Verwendung eines falsch klassifizierten Teilbereichs einer Szene als negatives Trainingsbeispiel auf.

Figurenliste

[0016] Die detaillierte Beschreibung bezieht sich auf die folgenden Figuren, in denen:

Fig. 1A illustriert ein Blockschaltbild der angeschlossenen Geräte eines Videoerfassungs- und Wiedergabesystems gemäß einer exemplarischen Ausführungsform;

Fig. 1B illustriert ein Blockschaltbild einer Reihe von Betriebsmodulen des Videoerfassungs- und Wiedergabesystems gemäß einer exemplarischen Ausführungsform;

Fig. 1C illustriert ein Blockschaltbild eines Satzes von Betriebsmodulen, die in einem Gerät gemäß einer exemplarischen Ausführungsform implementiert sind;

Fig. 2 illustriert ein Flussdiagramm einer exemplarischen Ausführungsform einer Methode zur Durchführung von Videoanalysen an Bilddaten;

Fig. 3A illustriert ein Blockschaltbild eines Satzes von operativen Teilmodulen eines Videoanalysemoduls nach einer exemplarischen Ausführungsform;

Fig. 3B illustriert eine Vielzahl von Objektklassifikatoren eines Objektklassifikationsmoduls gemäß einer exemplarischen Ausführungsform;

Fig. 4 illustriert ein Flussdiagramm eines aus der Technik bekannten Verfahrens zum weiteren Trainieren eines Basisklassifikators;

Fig. 5 illustriert ein Flussdiagramm eines verbesserten computerimplementierten Verfahrens zur Weiterbildung eines Basisklassifikators gemäß einer exemplarischen Ausführungsform;

Fig. 6A bis Fig. 6F sind Teilbereiche von Szenen mit detektierten visuellen Objekten im Vordergrund und deren dazugehörigen Hintergrundmodellen;

Fig. 7A ist ein erster historischer Vollbildrahmen, der ein Beispiel einer Szene darstellt;

Fig. 7B ist ein zweiter historischer Vollbildrahmen, der ein Beispiel der Szene darstellt;

Fig. 8 illustriert ein Flussdiagramm eines verbesserten computerimplementierten Verfahrens zum weiteren Training eines Basisklassifikators nach einer alternativen exemplarischen Ausführungsform; und

Fig. 9 illustriert ein Flussdiagramm eines verbesserten computerimplementierten Verfahrens zum szenenspezifischen Training eines Basisklassifikators gemäß einer exemplarischen Ausführungsform.

[0017] Es wird anerkannt dass aus Gründen der Einfachheit und Klarheit die in den Figuren dargestellten Elemente nicht unbedingt maßstabsgerecht gezeichnet wurden. Beispielsweise können die Abmessungen einiger Elemente im Vergleich zu anderen Elementen zur besseren Übersichtlichkeit übertrieben sein. Darüber hinaus können, wenn dies zweckmäßig erscheint, Referenzziffern zwischen den Figuren wiederholt werden, um entsprechende oder analoge Elemente anzugeben.

DETAILLIERTE BESCHREIBUNG DER VERSCHIEDENEN AUSFÜHRUNGSFORMEN

[0018] Um die hier beschriebenen beispielhaften Ausführungsformen zu verstehen, sind zahlreiche spezifische Details aufgeführt. Es wird jedoch von denjenigen, die mit gewöhnlichen Technischen Kenntnissen vertraut sind, verstanden, dass die hier beschriebenen Ausführungsformen ohne diese spezifischen Details praktiziert werden können. In anderen Fällen wurden bekannte Methoden, Verfahren und Komponenten nicht detailliert beschrieben, um die hier beschriebenen Verkörperungen nicht zu verdecken. Darüber hinaus ist diese Beschreibung nicht in irgendeiner Weise als Einschränkung des Umfangs der hier beschriebenen Ausführungsformen zu verstehen, sondern als bloße Beschreibung der Umsetzung der verschiedenen hier beschriebenen Ausführungsformen.

[0019] Richtungsbegriffe wie „oben“, „unten“, „nach oben“, „nach unten“, „vertikal“ und „seitlich“ werden in der folgenden Beschreibung nur zum Zwecke der relativen Bezugnahme verwendet und sollen nicht irgendwelche Einschränkungen bei der Positionierung eines Artikels während des Gebrauchs oder der Montage in einer Baugruppe oder relativ zu einer Umgebung suggerieren.

[0020] Die Begriffe „ein Aspekt“, „eine Ausführungsform“, „Ausführungsform“, „Ausführungsformen“, „die Ausführungsform“, „eine oder mehrere Ausführungsformen“, „einige Ausführungsformen“, „bestimmte Ausführungsformen“, „eine Ausführungsform“, „eine andere Ausführungsform“ und dergleichen bedeuten „eine oder mehrere (aber nicht alle) Ausführungsformen der offenbarten Erfindung(en)“, sofern nicht ausdrücklich anders angegeben. Ein Verweis auf „eine andere Ausführungsform“ oder „einen anderen Aspekt“ bei der Beschreibung einer Ausführungsform bedeutet nicht, dass sich die referenzierte Ausführungsform mit einer anderen Verkörperung (z.B. einer vor der referenzierten Ausführungsform beschriebenen Ausführungsform) gegenseitig ausschließt, sofern nicht ausdrücklich etwas anderes angegeben ist.

[0021] Die Begriffe „einschließlich“, „umfassend“ und Varianten davon bedeuten „einschließlich, aber nicht beschränkt auf“, sofern nicht ausdrücklich anders angegeben.

[0022] Der Begriff „Vielzahl“ bedeutet „zwei oder mehr“, sofern nicht ausdrücklich anders angegeben. Der Begriff „hierin“ bedeutet „in der vorliegenden Anmeldung, einschließlich allem, was durch Verweis aufgenommen werden kann“, sofern nicht ausdrücklich anders angegeben.

[0023] Der Begriff „z.B.“ und ähnliche Begriffe bedeuten „zum Beispiel“ und schränken somit den Begriff oder die Phrase, die sie erläutern nicht ein.

[0024] Der Begriff „jeweils“ und ähnliche Begriffe bedeuten „einzeln betrachtet“. Wenn also zwei oder mehr Dinge „entsprechende“ Eigenschaften haben, dann hat jedes dieser Dinge seine eigene Eigenschaft, und diese Eigenschaften können sich voneinander unterscheiden, müssen es aber nicht sein. Zum Beispiel bedeutet der Ausdruck „jede von zwei Maschinen hat eine entsprechende Funktion“, dass die erste Maschine eine Funktion hat und die zweite Maschine auch eine Funktion. Die Funktion der ersten Maschine kann mit der Funktion der zweiten Maschine übereinstimmen oder nicht.

[0025] Das Wort „ein“ in Verbindung mit dem Begriff „umfassend“ oder „einschließend“ in den Ansprüchen und/oder der Beschreibung kann „eins“ bedeuten, steht aber auch im Einklang mit der Bedeutung von „eins oder mehr“, „mindestens eins“ und „eins oder mehr als eins“, sofern der Inhalt nichts anderes bestimmt. Ebenso

kann das Wort „ein anderer“ mindestens ein zweiter oder mehr bedeuten, sofern der Inhalt nichts anderes bestimmt.

[0026] Die hier verwendeten Begriffe „gekoppelt“, „koppelnd“ oder „verbunden“ können je nach Kontext, in dem diese Begriffe verwendet werden, verschiedene Bedeutungen haben. Beispielsweise können die Begriffe gekoppelt, gekoppelt oder verbunden eine mechanische oder elektrische Bedeutung haben. Beispielsweise können die hier verwendeten Begriffe gekoppelt, gekoppelt oder verbunden bedeuten, dass zwei Elemente oder Geräte direkt miteinander verbunden sind oder über ein oder mehrere Zwischenelemente oder Geräte über ein elektrisches Element, ein elektrisches Signal oder ein mechanisches Element je nach Kontext miteinander verbunden sind.

[0027] „Bildraten“ beziehen sich auf Daten, die von einem Videoerfassungsgerät erzeugt werden und die vom Videoerfassungsgerät aufgenommene Bilder darstellen. Die Bildraten können eine Vielzahl aufeinanderfolgender Bildrahmen enthalten, die zusammen ein vom Videoerfassungsgerät aufgenommenes Video bilden. Jeder Bildrahmen kann durch eine Matrix von Pixeln dargestellt werden, wobei jedes Pixel einen Pixelbildwert hat. Beispielsweise kann der Pixelbildwert ein numerischer Wert auf Graustufen (z.B. 0 bis 255) oder eine Vielzahl numerischer Werte für farbige Bilder sein. Beispiele für Farbräume zur Darstellung von Pixelbildwerten in Bildraten sind RGB, YUV, CYKM, YCBCR 4:2:2 und YCBCR 4:2:0 Bilder. Es wird davon ausgegangen, dass sich „Bildraten“, wie sie hier verwendet werden, auf „rohe“ Bildraten beziehen können, die von dem Videoerfassungsgerät erzeugt wurden und/oder auf Bildraten, die in irgendeiner Form verarbeitet wurden.

[0028] Ein „visuelles Vordergrund-Objekt“ bezieht sich auf eine visuelle Darstellung eines realen Objekts (z.B. Person, Tier, Fahrzeug), das sich in den vom Videoerfassungsgerät aufgenommenen Bildrahmen befindet. Das visuelle Vordergrund-Objekt ist ein Objekt, das für verschiedene Zwecke, wie z.B. für die Videoüberwachung, von Interesse ist. Beispielsweise kann das visuelle Vordergrund-Objekt, das sich in einer Szene befindet, ein Ereignis darstellen, bei dem z. B. ein Mensch oder ein Fahrzeug anwesend ist. Ein visuelles Vordergrund-Objekt kann ein sich bewegendes Objekt oder ein zuvor bewegtes Objekt sein. Das visuelle Vordergrund-Objekt unterscheidet sich von einem Hintergrundobjekt, das ein Objekt darstellt, welches sich im Hintergrund einer Szene befindet und nicht von Interesse ist.

[0029] Ein „aktueller Bildrahmen“ bezieht sich auf einen Bildrahmen innerhalb der Vielzahl von sequentiellen Bildrahmen eines Videos, das derzeit in verschiedenen hier beschriebenen Systemen und Verfahren analysiert wird. Die Bildraten eines aktuellen Bildrahmens werden analysiert, um Informationen über Objekte zu erzeugen, die innerhalb des aktuellen Bildrahmens und/oder innerhalb einer Vielzahl von Bildrahmen vor dem aktuellen Bild aufgenommen wurden.

[0030] Ein „vorheriger Bildrahmen“ oder ein „historischer Bildrahmen“ eines aktuellen Bildrahmens bezieht sich auf einen Bildrahmen, der vor einem aktuellen Bildrahmen innerhalb der Vielzahl aufeinanderfolgender Bildrahmen eines Videos aufgetreten ist. Beispielsweise kann der vorherige Bildrahmen der Bildrahmen sein, der unmittelbar vor dem aktuellen Bildrahmen vorausgegangen ist. Alternativ kann der vorherige Bildrahmen auch ein früherer Bildrahmen aus der Vielzahl der sequentiellen Bildrahmen sein, ist aber ausreichend nahe am aktuellen Bildrahmen, um für den aktuellen Bildrahmen relevant zu sein.

[0031] „Verarbeitung von Bildraten“ oder Varianten davon bezieht sich auf eine oder mehrere computerimplementierte Funktionen, die auf Bildraten ausgeführt werden. Die Verarbeitung von Bildraten kann beispielsweise, ist aber nicht darauf beschränkt, Bildverarbeitungsvorgänge, Analysieren, Verwalten, Komprimieren, Kodieren, Speichern, Übertragen und/oder Abspielen der Videodaten umfassen. Die Analyse der Bildraten kann die Segmentierung von Bildbereichen und die Detektion visueller Objekte, die Verfolgung und/oder Klassifizierung visueller Objekte innerhalb der durch die Bildraten dargestellten Szene umfassen. Die Verarbeitung der Bildraten kann dazu führen, dass veränderte Bildraten erzeugt werden, wie z.B. komprimierte (z.B. verminderte Qualität) und/oder neu kodierte Bildraten. Die Verarbeitung der Bildraten kann auch dazu führen, dass zusätzliche Informationen über die in den Bildern erfassten Bildraten oder visuellen Objekte ausgegeben werden. Solche Zusatzinformationen werden beispielsweise allgemein als Metadaten verstanden. Die Metadaten können auch für die weitere Verarbeitung der Bildraten verwendet werden, wie z.B. das Zeichnen von Begrenzungsrahmen um detektierte visuelle Objekte in den Bildrahmen.

[0032] Wenn zwei oder mehr Begriffe oder Phrasen gleichbedeutend sind (z.B. wegen einer ausdrücklichen Aussage, dass die Begriffe oder Phrasen gleichbedeutend sind), bedeutet es nicht, dass Beispiele eines solchen Begriffs oder einer solchen Phrase eine andere Bedeutung haben müssen als Beispiele eines anderen Begriffs oder einer anderen Phrase. Zum Beispiel, wenn eine Aussage die Bedeutung von „einschließlich“

mit „einschließlich, aber nicht beschränkt auf“ gleichbedeutend macht, bedeutet die bloße Verwendung des Ausdrucks „einschließlich, aber nicht beschränkt auf“ nicht, dass der Begriff „einschließlich“ etwas anderes als „einschließlich, aber nicht beschränkt auf“ bedeutet.

[0033] Weder der Titel (am Anfang der ersten Seite der vorliegenden Anmeldung) noch der Abstract (am Ende der vorliegenden Anmeldung) sind in irgendeiner Weise als Einschränkung des Umfangs der offenbaren Erfindung(en) zu verstehen. Ein Abstract wurde nur deshalb in diesen Antrag aufgenommen, weil ein Abstract von nicht mehr als **150** Wörtern nach **37** C.F.R. Abschnitt 1.72(b) oder einem ähnlichen Gesetz in anderen Rechtssystemen erforderlich ist. Der Titel der vorliegenden Anmeldung und die Überschriften der Abschnitte in der vorliegenden Anmeldung dienen lediglich der Übersichtlichkeit und sind nicht als Einschränkung der Offenlegung zu verstehen.

[0034] Zahlreiche Ausführungsformen sind in der vorliegenden Anwendung beschrieben und dienen nur zur Veranschaulichung. Die beschriebenen Ausführungsformen sind in keiner Weise einschränkend und sollen es auch nicht sein. Der/die derzeit offenbarte(n) Aspekt(e) sind auf zahlreiche Ausführungsformen anwendbar, wie aus der Offenbarung hervorgeht. Ein Fachmann mit üblichen Fähigkeiten in diesem Gebiet wird erkennen, dass der/die offenbarte(n) Aspekt(e) mit verschiedenen Modifikationen und Veränderungen, wie z.B. strukturellen und logischen Modifikationen, ausgeübt werden kann. Obwohl besondere Merkmale des oder der offenbaren Aspekte mit Bezug auf eine oder mehrere besondere Ausführungsformen und/oder Zeichnungen beschrieben werden können, sollte man sich darüber im Klaren sein, dass diese Merkmale nicht auf die Verwendung in der einen oder den mehreren besonderen Ausführungsformen oder Zeichnungen, auf die sie bezogen sind, beschränkt sind, sofern nicht ausdrücklich etwas anderes angegeben ist.

[0035] Keine Ausführungsform der in dieser Anwendung beschriebenen Verfahrensschritte oder Produktelemente ist wesentlich oder koextensiv, es sei denn, dies wird in dieser Beschreibung ausdrücklich angegeben oder in einem Anspruch ausdrücklich erwähnt.

[0036] Wie von einem Fachmann geschätzt wird, können die verschiedenen hier beschriebenen Beispielausführungsformen als Verfahren, System oder Computerprogrammprodukt ausgestaltet werden. Dementsprechend können die verschiedenen exemplarischen Ausführungsformen in Form einer vollständigen Hardwareausführungsform, einer vollständigen Softwareausführungsform (einschließlich Firmware, residenter Software, Mikrocode usw.) oder einer Ausführungsform, die Software- und Hardwareaspekte kombiniert, die im Allgemeinen hierin als „Schaltkreis“, „Modul“ oder „System“ bezeichnet werden können, vorliegen. Darüber hinaus können die verschiedenen Beispielausführungsformen in Form eines Computerprogrammprodukts auf einem computerverwendbaren Speichermedium mit einem computerverwendbaren Programmcode auf dem Medium vorliegen.

[0037] Es kann jedes geeignete computerverwendbare oder computerlesbare Medium verwendet werden. Das computerverwendbare oder computerlesbare Medium kann beispielsweise ein elektronisches, magnetisches, optisches, elektromagnetisches, Infrarot- oder Halbleitersystem, Vorrichtung, Gerät oder Ausbreitungsmedium sein. Im Zusammenhang mit diesem Dokument kann ein computerverwendbares oder computerlesbares Medium jedes Medium sein, das das Programm zur Verwendung durch oder in Verbindung mit dem Befehlsausführungssystem, - Vorrichtung oder -gerät enthalten, speichern, kommunizieren, verbreiten oder transportieren kann.

[0038] Computerprogrammcode zur Ausführung von Operationen verschiedener Beispielausführungsformen kann in einer objektorientierten Programmiersprache wie Java, Smalltalk, C++, Python oder ähnlichem geschrieben werden. Der Computerprogrammcode zur Durchführung von Operationen verschiedener Beispielausführungsformen kann aber auch in herkömmlichen prozeduralen Programmiersprachen wie der Programmiersprache „C“ oder ähnlichen Programmiersprachen geschrieben werden. Der Programmcode kann vollständig auf einem Rechner, teilweise auf dem Rechner, als eigenständiges Softwarepaket, teilweise auf dem Rechner und teilweise auf einem entfernten Rechner oder ganz auf dem entfernten Rechner oder Server ausgeführt werden. Im letzteren Fall kann der entfernte Computer über ein lokales Netzwerk (LAN) oder ein Wide Area Network (WAN) mit dem Computer verbunden sein oder die Verbindung zu einem externen Computer (z.B. über das Internet über einen Internet Service Provider) hergestellt werden.

[0039] Im Folgenden werden verschiedene Beispielausführungsformen anhand von Flussdiagrammen und/oder Blockdiagrammen von Verfahren, Vorrichtungen (Systemen) und Computerprogrammprodukte entsprechend den Ausführungsformen der Erfindung beschrieben. Es wird davon ausgegangen, dass jeder Block der Flussdiagramme und/oder Blockdiagramme und Kombinationen von Blöcken in den Flussdiagrammen und/

oder Blockdiagrammen durch Computerprogrammanweisungen implementiert werden können. Diese Computerprogrammbefehle können einem Prozessor eines Universalrechners, Spezialrechners oder einer anderen programmierbaren Datenverarbeitungsvorrichtung zur Verfügung gestellt werden, um eine Maschine herzustellen, so dass die Befehle, die über den Prozessor des Computers oder einer anderen programmierbaren Datenverarbeitungsvorrichtung ausgeführt werden, Mittel zur Ausführung der im Flussdiagramm und/oder Blockdiagrammblock oder Blöcken angegebenen Funktionen/Aktionen erzeugen.

[0040] Diese Computerprogrammbefehle können auch in einem computerlesbaren Speicher gespeichert werden, der einen Computer oder ein anderes programmierbares Datenverarbeitungsgerät so anweisen kann, dass die im computerlesbaren Speicher gespeicherten Befehle einen Herstellungsgegenstand mit Befehlen erzeugen, die die im Flussdiagramm und/oder Blockdiagramm angegebene Funktion/Aktion ausführen.

[0041] Die Computerprogrammbefehle können auch auf einen Computer oder ein anderes programmierbares Datenverarbeitungsgerät geladen werden, um eine Reihe von Arbeitsschritten auf dem Computer oder einem anderen programmierbaren Gerät auszuführen, um einen computerimplementierten Prozess zu erzeugen, so dass die Befehle, die auf dem Computer oder einem anderen programmierbaren Gerät ausgeführt werden, Schritte zur Implementierung der im Flussdiagramm und/oder in Blockdiagrammblocken angegebenen Funktionen/Akten enthalten.

[0042] In **Fig. 1A** ist nun ein Blockschaltbild der angeschlossenen Geräte des Videoerfassungs- und Wiedergabesystems **100** nach einer Beispielausführungsform dargestellt. Beispielsweise kann das Videoerfassungs- und Wiedergabesystem **100** als Videoüberwachungssystem eingesetzt werden. Das Videoerfassungs- und Wiedergabesystem **100** beinhaltet Hard- und Software, die die hier beschriebenen Prozesse und Funktionen ausführen.

[0043] Das Videoerfassungs- und Wiedergabesystem **100** umfasst mindestens ein Videoerfassungsgerät **108**, das so betrieben werden kann, dass es eine Vielzahl von Bildern erfasst und Bilddaten erzeugt, die die Vielzahl der erfassten Bilder darstellen.

[0044] Jedes Videoerfassungsgerät **108** enthält mindestens einen Bildsensor **116** zur Erfassung mehrerer Bilder. Das Videoerfassungsgerät **108** kann eine digitale Videokamera sein und der Bildsensor **116** kann das aufgenommene Licht als digitale Daten ausgeben. Beispielsweise kann der Bildsensor **116** ein CMOS, NMOS oder CCD sein.

[0045] Der mindestens eine Bildsensor **116** kann zur Erfassung von Licht in einem oder mehreren Frequenzbereichen eingesetzt werden. Beispielsweise kann der mindestens eine Bildsensor **116** so betrieben werden, dass er Licht in einem Bereich erfasst, der im Wesentlichen dem Frequenzbereich des sichtbaren Lichts entspricht. In anderen Beispielen kann der mindestens eine Bildsensor **116** so betrieben werden, dass er Licht außerhalb des sichtbaren Bereichs, z.B. im Infrarot- und/oder Ultraviolettbereich, erfasst. In anderen Beispielen kann das Videoerfassungsgerät **108** eine Multisensorkamera sein, die zwei oder mehr Sensoren enthält, die Licht in verschiedenen Frequenzbereichen erfassen können.

[0046] Das mindestens eine Videoerfassungsgerät **108** kann eine bestimmte Kamera enthalten. Es wird davon ausgegangen, dass sich eine bestimmte Kamera auf eine Kamera bezieht, deren Hauptmerkmale darin bestehen, Bilder oder Videos aufzunehmen. In einigen Beispielausführungsformen kann die bestimmte Kamera Funktionen ausführen, die den aufgenommenen Bildern oder Videos zugeordnet sind, wie z. B., aber nicht darauf beschränkt, die Verarbeitung der von ihr oder einem anderen Videoerfassungsgerät **108** erzeugten Bilddaten. Die bestimmte Kamera kann z.B. eine Überwachungskamera sein, z.B. eine Schwenk-Neige-Zoom-Kamera, eine Dome-Kamera, eine Deckenkamera, eine Box-Kamera und eine Kugelkamera.

[0047] Zusätzlich oder alternativ kann das mindestens eine Videoerfassungsgerät **108** eine eingebettete Kamera enthalten. Es wird davon ausgegangen, dass sich eine eingebettete Kamera auf eine Kamera bezieht, die in ein Gerät eingebettet ist, das betriebsbereit ist, um Funktionen auszuführen, die nichts mit dem aufgenommenen Bild oder Video zu tun haben. Zum Beispiel kann die eingebettete Kamera eine Kamera sein, die sich auf einem beliebigen Laptop, Tablett, Drohnengerät, Smartphone, Videospielkonsole oder Controller befindet.

[0048] Jedes Videoerfassungsgerät **108** enthält einen oder mehrere Prozessoren **124**, ein oder mehrere mit den Prozessoren gekoppelte Speichergeräte **132** und eine oder mehrere Netzwerkschnittstellen. Das Speichergerät kann einen lokalen Speicher (z.B. einen Direktzugriffsspeicher und einen Cache-Speicher) enthalten, der bei der Ausführung von Programmbefehlen verwendet wird. Der Prozessor führt Computerprogramm-

befehle (z.B. ein Betriebssystem und/oder Anwendungsprogramme) aus, die im Speichergerät gespeichert werden können.

[0049] In verschiedenen Ausführungen kann der Prozessor **124** von jeder Verarbeitungsschaltung mit einer oder mehreren Schaltungseinheiten implementiert werden, einschließlich eines digitalen Signalprozessors (DSP), eines eingebetteten Grafikprozessors (GPU) usw., und jeder Kombination davon, die unabhängig oder parallel arbeitet, einschließlich möglicherweise redundantem Betrieb. Eine solche Verarbeitungsschaltung kann durch eine oder mehrere integrierte Schaltungen (IC) implementiert werden, einschließlich einer monolithischen integrierten Schaltung (MIC), einer anwendungsspezifischen integrierten Schaltung (ASIC), einem Field Programmable Gate Array (FPGA) usw. oder einer beliebigen Kombination davon. Zusätzlich oder alternativ kann eine solche Verarbeitungsschaltung z.B. als speicherprogrammierbare Steuerung (SPS) realisiert werden. Der Prozessor kann eine Speicherschaltung, wie z.B. digitale Daten, enthalten und kann z.B. die Speicherschaltung enthalten oder mit der Speicherschaltung verkabelt sein.

[0050] In verschiedenen Beispielausführungsformen ist das mit dem Prozessorschaltkreis gekoppelte Speichergerät **132** zur Speicherung von Daten und Computerprogrammbefehlen betreibbar. Typischerweise ist die Speichervorrichtung ganz oder teilweise eine digitale elektronische integrierte Schaltung oder besteht aus einer Vielzahl von digitalen elektronischen integrierten Schaltungen. Das Speichermedium kann als Nur-Lese-Speicher (ROM), Programmierbarer Nur-Lese-Speicher (PROM), Löschrbarer Programmierbarer Nur-Lese-Speicher (EPROM), Elektrisch Löschrbarer Programmierbarer Nur-Lese-Speicher (EEPROM), Flash-Speicher, ein oder mehrere Flash-Laufwerke, Universal Serial Bus (USB) angeschlossene Speichereinheiten, magnetischer Speicher, optischer Speicher, magneto-optischer Speicher usw. oder eine beliebige Kombination davon ausgeführt werden. Der Speicher kann so betrieben werden, dass er als flüchtiger Speicher, nichtflüchtiger Speicher, dynamischer Speicher usw. oder eine beliebige Kombination davon gespeichert werden kann.

[0051] In verschiedenen Beispielausführungsformen können mehrere Komponenten des Bilderfassungsgerätes **108** gemeinsam in einem System auf einem Chip (SOC) realisiert werden. Beispielsweise können der Prozessor **124**, das Speichergerät **116** und die Netzwerkschnittstelle innerhalb eines SOC implementiert werden. Außerdem können auf diese Weise ein Universalprozessor und ein oder mehrere GPUs und DSPs gemeinsam im SOC implementiert werden.

[0052] Fortfahrend mit **Fig. 1A**, jedes der mindestens einen Videoerfassungsgeräte **108** ist an ein Netzwerk **140** angeschlossen. Jedes Videoerfassungsgerät **108** kann Bilddaten ausgeben, die Bilder darstellen, die es aufnimmt, und die Bilddaten über das Netzwerk übertragen.

[0053] Es wird davon ausgegangen, dass das Netzwerk **140** jedes Kommunikationsnetz sein kann, das den Empfang und die Übertragung von Daten ermöglicht. Das Netzwerk **140** kann z.B. ein lokales Netzwerk, ein externes Netzwerk (z.B. WAN, Internet) oder eine Kombination davon sein. In anderen Beispielen kann das Netzwerk **140** ein Cloud-Netzwerk enthalten.

[0054] In einigen Beispielen enthält das Videoerfassungs- und Wiedergabesystem **100** ein Verarbeitungsgerät **148**. Das Verarbeitungsgerät **148** kann die von einem Videoerfassungsgerät **108** ausgegebenen Bilddaten verarbeiten. Das Verarbeitungsgerät **148** umfasst auch einen oder mehrere Prozessoren und einen oder mehrere an einen Prozessor gekoppelte Speichergeräte. Die Verarbeitungsgerät **148** kann auch eine oder mehrere Netzwerkschnittstellen enthalten.

[0055] Zum Beispiel, und wie abgebildet, wird das Verarbeitungsgerät **148** an ein Videoerfassungsgerät **108** angeschlossen. Das Verarbeitungsgerät **148** kann ferner an das Netzwerk **140** angeschlossen werden.

[0056] Das Videoerfassungs- und Wiedergabesystem **100** umfasst gemäß einer beispielhaften Ausführungsform, wie in **Fig. 1A** dargestellt, mindestens eine Workstation **156** (z.B. Server) mit jeweils einem oder mehreren Prozessoren. Die mindestens eine Workstation **156** kann auch Speicherplatz enthalten. Die Workstation **156** empfängt Bilddaten von mindestens einem Videoerfassungsgerät **108** und führt die Verarbeitung der Bilddaten durch. Die Workstation **156** kann weiterhin Befehle zur Verwaltung und/oder Steuerung eines oder mehrerer Bilderfassungsgeräte **108** senden. Die Workstation **156** kann Rohbilddaten vom Videoerfassungsgerät **108** empfangen. Alternativ oder zusätzlich kann die Workstation **156** Bilddaten empfangen, die bereits eine Zwischenverarbeitung durchlaufen haben, wie z.B. die Verarbeitung am Videoerfassungsgerät **108** und/oder an einem Verarbeitungsgerät **148**. Die Workstation **156** kann auch Metadaten aus den Bilddaten empfangen und die Bilddaten weiterverarbeiten.

[0057] Es wird davon ausgegangen, dass eine einzelne Workstation **156** in **Fig. 1A** dargestellt ist, die Workstation jedoch als Aggregation einer Vielzahl von Workstations implementiert werden kann.

[0058] Das Videoerfassungs- und Wiedergabesystem **100** enthält außerdem mindestens ein Client-Gerät **164**, das an das Netzwerk **140** angeschlossen ist. Das Client-Gerät **164** wird von einem oder von mehreren Benutzern zur Interaktion mit dem Videoerfassungs- und Wiedergabesystem **100** verwendet. Dementsprechend umfasst das Client-Gerät **164** mindestens ein Anzeigegerät und mindestens ein Benutzereingabegerät (z.B. Maus, Tastatur, Touchscreen). Das Client-Gerät **164** kann auf seinem Anzeigegerät eine Benutzeroberfläche zur Anzeige von Informationen, zum Empfang von Benutzereingaben und zur Wiedergabe von Videos anzeigen. Zum Beispiel kann das Client-Gerät ein beliebiger PC, Laptop, Tablett, PDA, Handy, Smartphone, Spielgerät und anderes mobiles Gerät sein.

[0059] Das Client-Gerät **164** ist für den Empfang von Bilddaten über das Netzwerk **140** und für die Wiedergabe der empfangenen Bilddaten geeignet. Ein Client-Gerät **164** kann auch über Funktionen zur Verarbeitung von Bilddaten verfügen. Beispielsweise können die Verarbeitungsfunktionen eines Client-Geräts **164** auf die Verarbeitung der empfangenen Bilddaten beschränkt werden. In anderen Beispielen können Bildverarbeitungsfunktionen zwischen der Workstation **156** und einem oder mehreren Client-Geräten **164** geteilt werden.

[0060] In einigen Beispielen kann das Bilderfassungs- und Wiedergabesystem **100** auch ohne die Workstation **156** realisiert werden. Dementsprechend können Bildverarbeitungsfunktionen vollständig auf einem oder mehreren Videoerfassungsgeräten ausgeführt werden **108**. Alternativ können die Bildverarbeitungsfunktionen von zwei oder mehr der Videoerfassungsgeräte **108**, der Verarbeitungsgeräte **148** und der Client-Geräte **164** gemeinsam genutzt werden.

[0061] Bezogen auf **Fig. 1B** ist darin ein Blockschaltbild eines Satzes **200** von Betriebsmodulen des Videoerfassungs- und Wiedergabesystems **100** nach einem Beispiel dargestellt. Die Betriebsmodule können in Hardware, Software oder beides auf einem oder mehreren Geräten des Videoerfassungs- und Wiedergabesystems **100** implementiert werden, wie in **Fig. 1A** dargestellt.

[0062] Der Satz **200** von Betriebsmodule enthält mindestens ein Videoerfassungsmodul **208**. Beispielsweise kann jedes Videoerfassungsgerät **108** ein Videoerfassungsmodul **208** implementieren. Das Videoerfassungsmodul **208** kann eine oder mehrere Komponenten (z.B. Sensor **116**, etc.) eines Videoerfassungsgeräts **108** steuern, um Bilder aufzunehmen.

[0063] Der Satz **200** von Betriebsmodulen enthält eine Untergruppe **216** von Bildverarbeitungsmodulen. Beispielsweise, und wie gezeigt, enthält die Untergruppe **216** der Bilddatenverarbeitungsmodule ein Videoanalysemodul **224** und ein Videoverwaltungsmodul **232**.

[0064] Das Videoanalysemodul **224** empfängt Bilddaten und analysiert diese, um Eigenschaften oder Charakteristika des aufgenommenen Bildes oder Videos und/oder von Objekten in der durch das Bild oder Video dargestellten Szene zu bestimmen. Das Videoanalysemodul **224** kann aufgrund der durchgeführten Bestimmungen weitere Metadaten ausgeben, die Informationen über die Bestimmungen liefern. Beispiele für Bestimmungen des Videoanalysemoduls **224** können eine oder mehrere von Vordergrund-/Hintergrundsegmentierung, Objekterkennung, Objektverfolgung, Objektklassifizierung, virtueller Stolperdraht, Anomalie Erkennung, Gesichtsdetektion, Gesichtserkennung, Kennzeichenerkennung, Identifizierung von „zurückgelassenen“ Objekten, Überwachung von Objekten (z.B. zum Schutz vor Diebstahl) und Business Intelligence sein. Es wird jedoch davon ausgegangen, dass auch andere in Technik bekannte Videoanalysefunktionen durch das Videoanalysemodul **224** implementiert werden können.

[0065] Das Videoverwaltungsmodul **232** empfängt Bilddaten und führt Verarbeitungsfunktionen für die Bilddaten der Videoübertragung, -wiedergabe und/oder -speicherung durch. Beispielsweise kann das Videoverwaltungsmodul **232** die Bilddaten verarbeiten, um die Übertragung der Bilddaten je nach Bandbreitenbedarf und/oder Kapazität zu ermöglichen. Das Videoverwaltungsmodul **232** kann auch die Bilddaten entsprechend den Wiedergabemöglichkeiten eines Client-Geräts **164** verarbeiten, das das Video wiedergeben soll, wie z.B. Rechenleistung und/oder Auflösung des Displays des Client-Geräts **164**. Das Videoverwaltungsmodul **232** kann die Bilddaten auch entsprechend der Speicherkapazität innerhalb des Videoerfassungs- und Wiedergabesystems **100** zur Speicherung der Bilddaten verarbeiten.

[0066] Es wird verständlich sein, dass die Untergruppe **216** der Videoverarbeitungsmodule nach einigen Beispielen nur eines von dem Videoanalysemodul **224** und dem Videoverwaltungsmodul **232** enthalten kann.

[0067] Der Satz **200** von Betriebsmodule enthält außerdem eine Untergruppe **240** von Speichermodule. Die Untergruppe **240** von Speichermodule umfasst beispielsweise ein Videospeichermodule **248** und ein Metadaten Speichermodule **256**. Das Videospeichermodule **248** speichert Bilddaten, die vom Videoverwaltungsmodul verarbeitet werden können. Das Metadaten Speichermodule **256** speichert die vom Videoanalysemodul **224** ausgegebenen Informationsdaten.

[0068] Es wird verständlich sein, dass das Videospeichermodule **248** und das Metadaten Speichermodule **256** zwar als separate Module dargestellt werden, jedoch innerhalb eines Hardware-Speichermediums implementiert werden können, wobei logische Regeln implementiert werden, um gespeicherte Videos von gespeicherten Metadaten zu trennen. In anderen Beispielausführungsformen kann das Videospeichermodule **248** und/oder das Metadaten Speichermodule **256** innerhalb einer Vielzahl von Hardware-Speichergeräten implementiert werden, in denen ein verteiltes Speicherschema implementiert werden kann.

[0069] Der Satz von Betriebsmodulen enthält außerdem mindestens ein Videowiedergabemodule **264**, das zum Empfang von Bilddaten und zur Wiedergabe der Bilddaten als Video betrieben werden kann. Beispielsweise kann das Videowiedergabemodule **264** auf einem Client-Gerät **164** implementiert werden.

[0070] Die Betriebsmodule des Satzes **200** können auf einem oder mehreren von dem Bilderfassungsgerätes **108**, Verarbeitungsgerätes **148**, Workstation **156** und Client-Gerätes **164** implementiert werden. In einigen Beispielausführungsformen kann ein Betriebsmodul vollständig auf einem einzigen Gerät implementiert werden. Beispielsweise kann das Videoanalysemodul **224** vollständig auf der Workstation **156** implementiert werden. Ebenso kann das Videoverwaltungsmodul **232** vollständig auf der Workstation **156** implementiert werden.

[0071] In anderen Beispielausführungsformen können einige Funktionalitäten eines Betriebsmoduls des Satzes **200** teilweise auf einem ersten Gerät implementiert werden, während andere Funktionalitäten eines Betriebsmoduls auf einem zweiten Gerät implementiert werden können. Beispielsweise können Videoanalysefunktionen zwischen einem oder mehreren Bilderfassungsgeräten **108**, einer Verarbeitungseinheit **148** und einer Workstation **156** aufgeteilt werden. Ebenso können die Videoverwaltungsfunktionen zwischen einem oder mehreren Bilderfassungsgeräten **108**, einem Verarbeitungsgerät **148** und einer Workstation **156** aufgeteilt werden.

[0072] Bezogen auf **Fig. 1C** ist darin ein Blockschaltbild eines Satzes **200** von Betriebsmodulen des Videoerfassungs- und Wiedergabesystems **100** nach einer bestimmten Beispielausführungsform dargestellt, wobei das Videoanalysemodul **224**, das Videoverwaltungsmodul **232** und das Speichergerät **240** vollständig auf dem einen oder den mehreren Bilderfassungsgeräten **108** implementiert sind. Dementsprechend benötigt das Videoerfassungs- und Wiedergabesystem **100** möglicherweise keine Workstation **156** und/oder ein Verarbeitungsgerät **148**.

[0073] Es wird begrüßt werden, dass die Möglichkeit, die Untergruppe der Bilddatenverarbeitungsmodulen auf einem einzigen Gerät oder auf verschiedenen Geräten des Videoerfassungs- und Wiedergabesystems **100** zu implementieren, Flexibilität beim Aufbau des Systems **100** ermöglicht.

[0074] Beispielsweise kann man ein bestimmtes Gerät mit bestimmten Funktionalitäten mit einem anderen Gerät ohne diese Funktionalitäten verwenden. Dies kann bei der Integration von Geräten verschiedener Parteien (z.B. Hersteller) oder bei der Nachrüstung eines bestehenden Videoerfassungs- und Wiedergabesystems sinnvoll sein.

[0075] Unter Bezugnahme auf **Fig. 2** ist darin ein Flussdiagramm einer exemplarischen Ausführungsform eines Verfahrens **272** zur Durchführung von Videoanalysen an einem oder mehreren Bildrahmen eines von einem Videoerfassungsgerät **108** erfassten Videos dargestellt. Die Videoanalyse kann vom Videoanalysemodul **224** durchgeführt werden, um Eigenschaften oder Charakteristika des erfassten Bildes oder Videos und/oder von visuellen Objekten zu bestimmen, die sich in der im Video aufgenommenen Szene befinden.

[0076] Bei **300** wird mindestens ein Bildrahmen des Videos in Vordergrund- und Hintergrundbereiche segmentiert. Die Segmentierung trennt Bereiche des Bildrahmens, die bewegten Objekten (oder zuvor bewegten Objekten) in der aufgenommenen Szene entsprechen, von stationären Bereichen der Szene.

[0077] Bei **302** werden ein oder mehrere visuelle Vordergrund-Objekte in der durch den Bildrahmen dargestellten Szene anhand der Segmentierung von **300** detektiert. Beispielsweise kann jeder einzelne zusammenhängende Vordergrundbereich oder „Blob“ als visuelles Vordergrund-Objekt in der Szene identifiziert werden.

Beispielsweise werden nur zusammenhängende Vordergrundbereiche größer als eine bestimmte Größe (z.B. Anzahl der Pixel) als visuelles Vordergrund-Objekt in der Szene identifiziert.

[0078] Metadaten können weiter generiert werden, die sich auf einen oder mehrere detektierten Vordergrundbereiche beziehen. Die Metadaten können die Position des visuellen Vordergrund-Objekts innerhalb des Bildrahmens definieren. Beispielsweise können die Standortmetadaten weiter verwendet werden, um eine Begrenzungsbox (z.B. beim Codieren von Video oder bei der Wiedergabe von Video) zu erzeugen, die das detektierte visuelle Vordergrund-Objekt umreißt.

[0079] Ein visueller Indikator kann dem Bildrahmen hinzugefügt werden, um jedes der einen oder mehreren detektierten visuellen Vordergrund-Objekte im visuell zu identifizieren. Der visuelle Indikator kann ein Begrenzungsrahmen sein, der jedes der visuellen Vordergrund-Objekte innerhalb des Bildrahmens umgibt.

[0080] Nach verschiedenen exemplarischen Ausführungsbeispielen kann die Videoanalyse mit der Detektion von Objekten in der erfassten Szene enden.

[0081] In anderen Beispielausführungsformen kann die Videoanalyse auch, bei **304**, die Klassifizierung der bei **302** detektierten visuellen Vordergrund-Objekte umfassen. Beispielsweise kann eine Mustererkennung durchgeführt werden, um die visuellen Vordergrund-Objekte zu klassifizieren. Ein visuelles Vordergrund-Objekt kann nach Klassen klassifiziert werden, z.B. eine Person, ein Auto oder ein Tier. Zusätzlich oder alternativ kann ein visuelles Objekt durch Aktionen klassifiziert werden, wie z.B. Bewegung und Bewegungsrichtung des visuellen Objekts. Es können auch andere Klassifikatoren wie Farbe, Größe, Ausrichtung usw. bestimmt werden. In konkreteren Beispielen kann die Klassifizierung des visuellen Objekts die Identifizierung einer Person anhand der Gesichtserkennung und die Erkennung von Text, wie z.B. eines Nummernschildes, beinhalten. Die visuelle Klassifikation kann nach Systemen und Methoden durchgeführt werden, die im gemeinsamen US-Patent Nr. 8,934,709 beschrieben sind, das hier durch Verweis in seiner Gesamtheit enthalten ist.

[0082] Die Videoanalyse kann ferner, bei **306**, das Detektieren, ob ein Ereignis eingetreten ist, und die Art des Ereignisses umfassen. Die Detektion des Ereignisses kann auf einem Vergleich der Klassifizierung eines oder mehrerer visueller Vordergrund-Objekte mit einer oder mehreren vordefinierten Regeln basieren. Das Ereignis kann ein Ereignis in der Anomalie Erkennung oder Business Intelligence sein, z.B. ob ein Video-Tripwire ausgelöst wurde, die Anzahl der in einem Bereich anwesenden Personen, ob ein Objekt in der Szene „zurückgelassen“ wurde oder ob ein Objekt in der Szene entfernt wurde.

[0083] In **Fig. 3A** ist nun ein Blockschaltbild eines Satzes **400** von operativen Untermodulen des Videoanalysemoduls nach einer Beispielausführungsform dargestellt. Das Videoanalysemodul **400** enthält eine Reihe von Modulen zur Durchführung verschiedener Aufgaben. Zum Beispiel enthält das Videoanalysemodul **400** ein Objektdetektionsmodul **404** zur Detektion von Objekten, die im Sichtfeld des Videoerfassungsgerätes **108** erscheinen. Das Objektdetektionsmodul **404** kann alle bekannten Objektdetektionsverfahren wie z.B. Bewegungsdetektion und Blob-Detektion verwenden. Das Objektdetektionsmodul **404** kann die Systeme enthalten und die Detektionsmethoden verwenden, die in US Pat. Nr. 7.627.171 mit dem Titel „Methoden und Systeme zur Detektion von Objekten von Interesse in raumzeitlichen Signalen“, deren gesamter Inhalt hier durch Verweis aufgenommen wird.

[0084] Das Videoanalysemodul **400** kann auch ein Objektverfolgungsmodul **408** enthalten, das an das Objektdetektionsmodul **404** angeschlossen ist. Das Objektverfolgungsmodul **408** kann Umstände eines vom Objektdetektionsmodul **404** detektierten Objekts zeitlich zuordnen. Das Objektverfolgungsmodul **408** kann die Systeme enthalten und die Methoden verwenden, die in der gemeinsamen U.S. Pat. Nr. 8.224.029 mit dem Titel „Object Matching for Tracking, Indexing, and Search“, deren gesamter Inhalt hier durch Verweis aufgenommen wird. Das Objektverfolgungsmodul **408** generiert Metadaten, die den visuellen Objekten entsprechen, die es verfolgt. Die Metadaten können Signaturen des visuellen Objekts entsprechen, die das Aussehen des Objekts oder andere Merkmale darstellen. Die Metadaten können zur Speicherung in die Metadatenbank **256** übertragen werden.

[0085] Das Videoanalysemodul **400** kann auch ein zeitliches Objektklassifikationsmodul **412** enthalten, das an das Objektverfolgungsmodul **408** angeschlossen ist. Das zeitliche Objektklassifikationsmodul **412** ist in der Lage, ein Objekt nach seinem Typ (z.B. Mensch, Fahrzeug, Tier) zu klassifizieren, indem es das Erscheinen des Objekts über die Zeit betrachtet. Mit anderen Worten, das Objektverfolgungsmodul **408** verfolgt ein Objekt für mehrere Frames, und das zeitliche Objektklassifikationsmodul **412** bestimmt den Objekttyp basierend auf seinem Auftreten in den mehreren Frames. Zum Beispiel kann eine Ganganalyse der Art und Weise, wie eine

Person geht, nützlich sein, um eine Person zu klassifizieren, oder eine Analyse der Beine einer Person kann nützlich sein, um einen Fahrradfahrer zu klassifizieren. Das zeitliche Objektklassifikationsmodul **412** kann Informationen über die Flugbahn eines Objekts (z.B. ob die Flugbahn glatt oder chaotisch ist, ob sich das Objekt bewegt oder nicht) und die Sicherheit von Klassifikationen, die von einem Objektklassifikationsmodul **416** (im Folgenden detailliert beschrieben) über mehrere Frames gemittelt werden, kombinieren. Beispielsweise können die vom Objektklassifikationsmodul **416** ermittelten Klassifikationsvertrauenswerte auf der Grundlage der Glätte der Flugbahn des Objekts angepasst werden. Das zeitliche Objektklassifikationsmodul **412** kann ein Objekt einer unbekannten Klasse zuordnen, bis das visuelle Objekt durch das Objektklassifikationsmodul ausreichend oft klassifiziert und eine vorbestimmte Anzahl von Statistiken gesammelt wurde. Bei der Klassifizierung eines Objekts kann das zeitliche Objektklassifikationsmodul **412** auch berücksichtigen, wie lange sich das Objekt im Sichtfeld befindet. Das zeitliche Objektklassifizierungsmodul kann anhand der oben beschriebenen Informationen eine endgültige Entscheidung über die Klasse eines Objekts treffen. Das zeitliche Objektklassifikationsmodul **412** kann auch einen Hysterese-Ansatz zur Änderung der Klasse eines Objekts verwenden. Genauer gesagt, kann ein Schwellenwert für den Übergang der Klassifizierung eines Objekts von einer unbekannten zu einer bestimmten Klasse festgelegt werden, und dieser Schwellenwert kann größer sein als ein Schwellenwert für den umgekehrten Übergang (z.B. von einem Menschen zu einer unbekannten Klasse). Das temporäre Objektklassifikationsmodul **412** kann Metadaten erzeugen, die sich auf die Klasse eines Objekts beziehen, und die Metadaten können in der Metadatenbank **256** gespeichert werden. Das zeitliche Objektklassifikationsmodul **412** kann die Klassifikationen des Objektklassifikationsmoduls **416** aggregieren.

[0086] Das Videoanalysemodul **400** beinhaltet auch das Objektklassifikationsmodul **416**, vorzugsweise direkt oder indirekt mit dem Objekterkennungsmodul **404** verbunden. Im Gegensatz zum zeitlichen Objektklassifikationsmodul **412** kann das Objektklassifikationsmodul **416** den Typ eines visuellen Objekts anhand eines einzelnen Vorkommnisses (z.B. Einzelbild) des Objekts bestimmen. Die Eingabe in das Objektklassifikationsmodul **416** ist vorzugsweise ein Teilbereich eines Bildrahmens, in dem sich das gewünschte visuelle Objekt befindet und nicht der gesamte Bildrahmen. Ein Vorteil der Eingabe eines Teilbereichs des Bildrahmens in das Objektklassifikationsmodul **416** ist, dass nicht die gesamte Szene zur Klassifikation analysiert werden muss und somit weniger Rechenleistung benötigt wird. Um die Komplexität des Objektklassifikationsmoduls **416** weiter zu vereinfachen, können auch andere Vmodule, wie z.B. ein Heuristik-basiertes Modul zur Erfassung offensichtlicher Klassifikationen, eingebunden werden.

[0087] In einer alternativen Anordnung wird das Objektklassifikationsmodul **416** nach dem Objektdetektionsmodul **404** und vor dem Objektverfolgungsmodul **408** platziert, so dass die Objektklassifizierung vor der Objektverfolgung erfolgt. In einer anderen alternativen Anordnung sind die Module Objektdetektion, Tracking, zeitliche Klassifikation und Klassifikation **404**, **408** und **416** wie oben beschrieben miteinander verknüpft.

[0088] Das Objektklassifikationsmodul **416** enthält eine Reihe von Objektklassifikatoren, wie im Blockschaltbild von **Fig. 3B** dargestellt. Beispielsweise kann das Objektklassifikationsmodul **416** einen Menschen-Ganzkörper-Klassifikator **424** enthalten, der bestimmt, ob ein Bild eines detektierten Objekts einem menschlichen Ganzkörper entspricht, einen Menschen-Torsoklassifikator **428**, der bestimmt, ob ein Bild eines detektierten Objekts einem menschlichen Torso entspricht, und einen Fahrzeug-Klassifikator **432**, der bestimmt, ob ein Bild eines detektierten Objekts einem Fahrzeug entspricht. Das Objektklassifikationsmodul **416** kann beliebig viele verschiedene Klassifikatoren enthalten, und, wie im Folgenden näher beschrieben, kann ein Benutzer neue Objektklassen für das Objektklassifikationsmodul **416** anlegen, auch wenn das Kamerasystem im Einsatz ist und funktioniert. Mit anderen Worten, das Objektklassifikationsmodul **416** ist vor Ort trainierbar.

[0089] Die Objektklassifikatoren sind in der Lage, ein Objekt anhand seiner Eigenschaften (z.B. Erscheinungscharakteristika) zu klassifizieren. Beispielsweise empfängt der Menschen-Ganzkörper-Klassifikator **424** Daten (z.B. ein Eingabemuster X), die den Merkmalen eines Objekts entsprechen und bestimmt, ob das Objekt einem vollständigen menschlichen Körper entspricht oder nicht. Nachdem das Objektklassifikationsmodul **416** ein Objekt klassifiziert hat, können Metadaten, die die Klasse des Objekts und die Merkmale des Objekts repräsentieren, in der Metadatenbank **256** gespeichert werden.

[0090] Merkmale, die vom Objektklassifikationsmodul **416** genutzt werden können, werden nun näher beschrieben. Ein nachfolgend beschriebener Trainingsalgorithmus wählt eine Teilmenge von Merkmalen aus $F = \{f_{k1}, f_{k2}, \dots, f_{km}\}$ aus einer Reihe von Merkmalen $F = \{f_1, f_2, \dots, f_n\}$. Das Eingabemuster X besteht aus den Elementen von F . Die Elemente von F kann als eine Art Transformation eines Bildbereichs R eines Objekts angesehen werden. Also, X kann die folgende Form annehmen:

$$X = \begin{pmatrix} \hat{f}_1 = f_{k1}(R) \\ \hat{f}_2 = f_{k2}(R) \\ \vdots \\ \hat{f}_m = f_{km}(R) \end{pmatrix}$$

[0091] Die Eigenschaften $\hat{f}_1, \hat{f}_2, \dots, \hat{f}_m$ eines Objekts kann einer Reihe von Erscheinungscharakteristika entsprechen, ist aber nicht darauf eingeschränkt, wie z.B. Seitenverhältnis, Farbe, Kantenausrichtung und normierte Sättigung. Darüber hinaus können die Merkmale $\hat{f}_1, \hat{f}_2, \dots, \hat{f}_m$ Merkmalsvektoren (z.B. Histogramme, in denen die Histogramm-Bins den Vektorkomponenten entsprechen) der Erscheinungscharakteristika darstellen und von einem oder mehreren Objektklassifikatoren zur Bestimmung der Objektklasse (z.B. Typ) verwendet werden. Beispielsweise können Histogramme der Kantenausrichtung eines Objekts für verschiedene Bereiche (z.B. Unterfenster) des Objektbildes konstruiert werden. Mit anderen Worten, ein Bild eines Objekts kann in Unterfenster unterteilt werden, und die Kantenausrichtung kann für jedes Pixel der Unterfenster berechnet werden. Die Kantenorientierung eines Pixels kann mit einem steuerbaren Filter (z.B. mit einem Gaußschen Ableitungsfilter in mehrere Richtungen) abgeleitet werden. Anhand eines steuerbaren Filters können den Pixeln eines Unterfensters dominante Richtungen zugeordnet und ein Histogramm der Richtungen für das Unterfenster erstellt werden. Beispielsweise kann für ein bestimmtes Pixel ein steuerbarer Filter in mehrere Richtungen verwendet werden, um mehrere Antworten zu erzeugen, und die Richtung, die der maximalen Richtungsableitung entspricht, wird als Richtung des Pixels zugewiesen.

[0092] Das Klassifizierungsproblem für einen der Objektklassifikatoren kann generell durch eine Klassifizierungsfunktion $\Gamma(X)$ definiert werden, in dem ein visuelles Objekt, dargestellt durch das Eingabemuster X , als Mitglied der Objektklasse deklariert wird, wenn $\Gamma(X) > 0$ oder als Nicht-Mitglied der Objektklasse wenn $\Gamma(X) < 0$. Generell wird die Klassifikatorfunktion $\Gamma(X)$ mit einem Satz von Parametern parametrisiert und das Eingabemuster X setzt sich aus den oben beschriebenen Merkmalen zusammen. Ein bestimmter Klassifikator $\Gamma_c(X)$ wird für jede Objektklasse von Interesse trainiert. Das durch das Objektklassifikationsmodul **416** in **Fig. 3A** dargestellte Mehrklassenklassifikationsmodell kann wie folgt mathematisch definiert werden:

$$\Omega = \{\omega_1, \omega_2, \dots, \omega_c\}$$

$$\omega = \omega_c : (\Gamma_c(X) > 0 \text{ und } \Gamma_u(X) \leq 0 \forall u \in \{1, 2, \dots, C\}, u \neq c)$$

wobei ω eine Objektklasse darstellt und Ω die Menge aller Objektklassen repräsentiert.

[0093] Eine Klassifizierungsfunktion $R(X)$ für eine bestimmte visuelle Objektklasse kann durch die Definition von Regeln (z.B. Größe und Seitenverhältnis von visuellen Objekten) erstellt werden. Die Klassifikatorfunktion kann durch maschinelles Lernen mit Hilfe von Trainingsdaten weiter trainiert werden. Wie in der Technik bekannt ist, versucht das Training eines Klassifikators, die Regeln dieses Klassifikators weiter zu verfeinern, damit er ein bestimmtes visuelles Objekt genauer klassifizieren kann. Die Trainingsdaten können positive Trainingsbeispiele und/oder negative Trainingsbeispiele enthalten. Ein positives Trainingsbeispiel bezieht sich auf ein Vorkommen eines visuellen Objekts, das als zu einer bestimmten Klasse von Objekten gehörig bestätigt wurde. Das positive Trainingsbeispiel dient dazu, einen Klassifikator zu trainieren, seine Regeln zu verfeinern, um ein bestimmtes visuelles Objekt genauer als in die Klasse dieses positiven Trainingsbeispiels einzuordnen. Ein negatives Trainingsbeispiel bezieht sich auf ein Vorkommen eines visuellen Objekts oder einer anderen visuellen Darstellung, die nicht zu einer bestimmten Klasse von Objekten gehört. Das negative Trainingsbeispiel kann ein Beispiel für ein visuelles Objekt sein, das von einem Klassifikator falsch als zu einer bestimmten Klasse von Objekten gehörend klassifiziert wurde. Das negative Trainingsbeispiel dient zum Trainieren eines Klassifikators.

[0094] Das maschinelle Lernen für das Training des Objektklassifikators kann jede geeignete maschinelle Lerntechnik sein, die in der Technik bekannt ist, wie z.B. Faltungsneuronale Netze, Support-Vektor-Maschinen, Entscheidungsbäume, Random Forests, Kaskaden-Klassifikatoren.

[0095] Das Training eines Objektklassifikators kann überwacht werden. Im überwachten Training wurden die positiven und/oder negativen Trainingsbeispiele von einem Menschen bestätigt. Beispielsweise, unter einem

großen Satz von Bildern, prüft und kennzeichnet ein oder mehrere menschliche Benutzer jedes Bild einzeln als ein visuelles Objekt, das zu einer Klasse gehört (z.B. Person, Fahrzeug, Tier) oder kein visuelles Objekt enthält.

[0096] Das Training eines Objektklassifikators kann auch unbeaufsichtigt sein. Beim unbeaufsichtigten Training wird ein Basisklassifikator verwendet, um zunächst ein oder mehrere visuelle Objekte zu klassifizieren, z.B. Objekte, die vom Objektdetektionsmodul **404** detektiert werden. Das visuelle Objekt und das Ergebnis der vom Basisklassifikator bestimmten Klassifizierung (z.B. eine positive Bestimmung, dass das visuelle Objekt zu einer spezifischen Objektklasse gehört) kann als positives Trainingsbeispiel für die weitere Ausbildung des Basisklassifikators verwendet werden. Bilddaten, in denen keine Objekte detektiert wurden, können auch als negative Trainingsbeispiele für das Training des Objektklassifikators verwendet werden. Im unbeaufsichtigten Training werden die Bilddaten, die als positive Trainingsbeispiele oder als negative Trainingsbeispiele verwendet werden, nicht von einem menschlichen Benutzer überprüft.

[0097] Ein Basisklassifikator bezieht sich hierin auf einen Objektklassifikator, der durch Definition von Regeln und/oder Training durch maschinelles Lernen für einen bestimmten Grad der Objektklassifikation konfiguriert wurde, der aber durch eine weitere Schulung mit Hilfe einer computerimplementierten visuellen Maschinsprache weiter optimiert werden kann.

[0098] Es soll nun auf **Fig. 4** verwiesen werden, worin ein Ablaufdiagramm eines Verfahrens **500** zur Weiterbildung eines Basisklassifikators dargestellt ist. Es wird davon ausgegangen, dass während die Methode **500** für das Training eines einzelnen Basisklassifikators dargestellt wird, die Methode **500** für das Training einer Vielzahl von Basisklassifikatoren parallel angewendet werden kann. Zum Beispiel kann ein Objektklassifikationsmodul **416**, wie an anderer Stelle beschrieben, eine Vielzahl von Objektklassifikatoren enthalten, wobei jeder Klassifikator betreibbar ist, um festzustellen, ob ein visuelles Objekt zu einem bestimmten Typ von Klasse gehört. Dementsprechend können die Vielzahl von Objektklassifikatoren des Objektklassifikationsmoduls **416** anhand von Trainingsbeispielen gemeinsam trainiert werden. Beispielsweise kann ein Trainingsbeispiel, das ein visuelles Vordergrund-Objekt einer bestimmten Klasse ist, als positives Trainingsbeispiel für einen Klassifikator verwendet werden, der zur gleichen Klasse gehört.

[0099] Bei **504** ist ein Basisklassifikator vorgesehen. Der Basisklassifikator kann ein beliebiger Objektklassifikator sein, der durch maschinelles Lernen anhand von visuellen Trainingsbeispielen weiter trainiert werden kann.

[0100] Bei **508** können ein oder mehrere Trainingsbeispiele empfangen werden. Die Trainingsbeispiele können positive Trainingsbeispiele und/oder negative Trainingsbeispiele sein, die automatisch oder unter Aufsicht erstellt werden können.

[0101] Bei **512** wird der Basisklassifikator durch Anwendung von maschinelles Lernen am Basisklassifikator anhand der bei **508** erhaltenen Trainingsbeispiele weiter trainiert.

[0102] Bei einigen Ausführungsformen werden die Schritte **508** und **512** so wiederholt, dass die Aktualisierung des Basisklassifikators einem iterativen Prozess folgt. Das heißt, ein erster Satz aus einer Vielzahl von Trainingsbeispielen kann für das Training des Basisklassifikators durch maschinelles Lernen in einer ersten Iteration verwendet werden. Ein zweiter Satz aus einer Vielzahl von Trainingsbeispielen kann zum weiteren Trainieren des Klassifikators durch maschinelles Lernen in einer nachfolgenden zweiten Iteration angewendet werden.

[0103] Bei **516** wird der nach den Schritten **508** und **512** trainierte Basisklassifikator auf dem Gebiet zur Klassifizierung von visuellen Vordergrund-Objekten eingesetzt.

[0104] In einigen Beispielen kann das Training des Basisklassifikators aus den Schritten **508** und **512** vor dem Einsatz des trainierten Klassifikators bei **516** durchgeführt werden.

[0105] In anderen Beispielen kann das Training eines Basisklassifikators in den Schritten **508** und **512** durchgeführt werden, während der Objektklassifikator bereits auf dem Gebiet eingesetzt wird. Die Trainingsbeispiele können visuelle Darstellungen von realen Objekten sein, die sich im Sichtfeld eines Videoerfassungsgerätes befinden, wenn dieses Gerät auf dem Gebiet eingesetzt wird. Beispielsweise kann der Basisklassifikator anfänglich eingesetzt und während des Einsatzes schrittweise von detektierten visuellen Vordergrund-Objekten von Feld **508** trainiert werden.

[0106] Visuelle Objekte, die als Trainingsbeispiele verwendet werden, können in einer beaufsichtigten Weise (z.B. durch einen menschlichen Benutzer visuell überprüft) oder in einer unbeaufsichtigten Weise (z.B. durch einen computerimplementierten Objektklassifikator) als zu einer Klasse gehörig identifiziert werden.

[0107] Es soll nun auf **Fig. 5** verwiesen werden, worin ein Flussdiagramm einer verbesserten computerimplementierten Methode **540** zum weiteren Training eines Basisklassifikators nach einer exemplarischen Ausführungsform dargestellt ist. Es wird davon ausgegangen, dass während die Methode **540** für das Training eines einzelnen Basisklassifikators dargestellt wird, die Methode **500** auch für das Training einer Vielzahl von Basisklassifikatoren zeitgleich angewendet werden kann. Zum Beispiel kann ein Objektklassifikationsmodul **416**, wie an anderer Stelle beschrieben, eine Vielzahl von Objektklassifikatoren enthalten, wobei jeder Klassifikator betreibbar ist, um zu bestimmen, ob ein visuelles Objekt zu einer spezifischen Klasse gehört. Dementsprechend können die mehreren Objektklassifikator des Objektklassifikationsmoduls **416** anhand von Trainingsbeispielen, die ihm bereitgestellt sind, gemeinsam trainiert werden. Beispielsweise kann ein Trainingsbeispiel, das ein visuelles Objekt einer bestimmten Klasse ist, als positives Trainingsbeispiel für einen Klassifikator verwendet werden, der zur gleichen Klasse gehört.

[0108] Bei **504** ist ein Basisklassifikator vorgesehen. Der Basisklassifikator kann jeder beliebige Objektklassifikator sein, der durch Anwenden von maschinellem Lernen anhand von Trainingsbeispielen visueller Objekte weiter optimiert werden kann.

[0109] Bei **544** wird ein visuelles Vordergrund-Objekt innerhalb von Bilddaten, die eine Szene darstellen, detektiert. Eine Szene hierin bezieht sich auf die visuelle Darstellung, die im Sichtfeld eines Videoerfassungsgerätes über ein bestimmtes Zeitintervall erfasst wird. Das Videoerfassungsgerät ist in diesem Zeitintervall statisch, so dass sein Sichtfeld unverändert bleibt. Dementsprechend bleibt auch die Szene, die über das Zeitintervall aufgenommen wird, unverändert, aber Objekte (z.B. Menschen, Fahrzeuge, andere Objekte) innerhalb der Szene können sich über dieses Zeitintervall verändern. Die visuelle Darstellung der Szene können die Bilderrahmen der Bilddaten sein, die vom Videoerfassungsgerät über das Zeitintervall erzeugt werden.

[0110] Das visuelle Vordergrund-Objekt kann auch von einem menschlichen Operator oder von einem computerimplementierten Modul als zu einer bestimmten Klasse gehörig klassifiziert werden. Das detektierte visuelle Vordergrund-Objekt befindet sich in einem Teilbereich der Szene. Beispielsweise kann der Teilbereich der Szene einem Teil eines Bildrahmens der Bilddaten entsprechen, in dem sich das detektierte visuelle Vordergrund-Objekt befindet. Beispielsweise kann der Teilbereich der Szene dem Teilbereich des Bildrahmens entsprechen, der durch den vom Objekterkennungsmodul **404** gezeichneten Begrenzungsrahmen zur visuellen Identifizierung des detektierten visuellen Vordergrund-Objekts begrenzt wird.

[0111] Bei **548** wird ein Hintergrundmodell des detektierten visuellen Objekts bestimmt. Ein Hintergrundmodell ist eine visuelle Darstellung der Szene oder einer Teilregion davon, wobei jedoch jedes visuelle Vordergrund-Objekt nicht in der Szene oder der Teilregion vorhanden ist. Das Hintergrundmodell des detektierten visuellen Vordergrund-Objekts ist das Hintergrundmodell des Teilbereichs der Szene, in dem sich das detektierte visuelle Vordergrund-Objekt befindet.

[0112] Wenn beispielsweise das bei **544** detektierte visuelle Vordergrund-Objekt ein Mensch ist und der Teilbereich der Szene einem Bereich eines Raumes entspricht, in dem sich dieser Mensch befindet, stellt das Hintergrundmodell dieses Teilbereichs diesen Bereich des Raumes dar, in dem dieser Mensch oder ein anderer Mensch nicht anwesend ist.

[0113] Wenn beispielsweise das bei **544** detektierte Vordergrund-Objekt ein Fahrzeug ist und der Teilbereich der Szene einem Teil eines Parkplatzes entspricht, auf dem sich dieses Fahrzeug befindet, stellt das Hintergrundmodell dieses Teilbereichs diesen Teil des Parkplatzes dar, in dem dieses Fahrzeug oder ein anderes Fahrzeug nicht vorhanden ist.

[0114] Bei **552** wird der Basisklassifikator optional durch Anwendung maschinellen Lernens auf den Basisklassifikator weiter trainiert, wobei das bei **544** detektierte visuelle Vordergrund-Objekt als positives Trainingsbeispiel verwendet wird.

[0115] Bei **556** wird der Basisklassifikator weiter trainiert, indem das maschinelle Lernen auf den Basisklassifikator angewendet wird, wobei das Hintergrundmodell des detektierten visuellen Vordergrund-Objekts als negatives Trainingsbeispiel verwendet wird.

[0116] Die Schritte **544** bis **556** können für eine Vielzahl von visuellen Objekte wiederholt werden, die detektiert und/oder klassifiziert sind. Für jedes bei **544** detektierte visuelle Objekt wird ein Hintergrundmodell, das spezifisch für den Teilbereich einer Szene ist, in dem sich dieses visuelle Objekt befindet, bei **548** bestimmt und für das Training des Basisklassifikators bei **556** verwendet.

[0117] In anderen Beispielen kann der Basisklassifikator durch Anwenden maschinellen Lernens auf den Basisklassifikator unter Verwendung eines Sets von einer Vielzahl von Trainingsbeispielen trainiert werden. Dieses Set enthält eine Vielzahl von Hintergrundmodellen von visuellen Vordergrund-Objekten, die in den Teilbereichen einer Vielzahl von verschiedenen Szenen detektiert sind.

[0118] Es wird verständlich sein, dass in einigen Ausführungsformen die Schritte **544** und **556** wiederholt werden, so dass die Aktualisierung des Basisklassifikators einem iterativen Prozess folgt. Das heißt, ein erstes Set von einem oder mehreren Trainingsbeispielen kann für das Training des Basisklassifikators durch maschinelles Lernen in einer ersten Iteration verwendet werden. Ein zweites Set aus einer Vielzahl von Trainingsbeispielen kann zum weiteren Trainieren des Basisklassifikators weiter verwendet werden, wie er nach der ersten Iteration durch maschinelles Lernen in einer nachfolgenden zweiten Iteration trainiert wurde.

[0119] Bei **516** wird der nach Schritt **556**, und optional Schritt **552**, trainierte Basisklassifikator auf dem Gebiet zur Klassifizierung weiterer visueller Vordergrund-Objekte eingesetzt.

[0120] Wie an anderer Stelle herein beschrieben, kann das Training des Basisklassifikators vor dem Einsatz des trainierten Klassifikators oder während der Objektklassifikators schon auf dem Gebiet eingesetzt wird.

[0121] Die **Fig. 6A** bis **Fig. 6F** zeigen visuelle Vordergrund-Objekte, die in Teilbereichen von Szenen und deren Hintergrundmodell detektiert wurden. **Fig. 6A** zeigt beispielsweise eine Person, die auf einem Abschnitt eines Bürgersteigs läuft. Die gehende Person ist das visuelle Objekt im Vordergrund, das detektiert wird. **Fig. 6B** zeigt das Hintergrundmodell des visuellen Objekts von **Fig. 6A**. Es wird anerkannt, dass das Hintergrundmodell das gleiche Segment des Bürgersteigs zeigt, ohne dass die Person geht oder ein anderes visuelles Vordergrund-Objekt vorhanden ist.

[0122] **Fig. 6C** zeigt eine Person, die eine Treppe hinabsteigt. Die Person ist das visuelle Vordergrund-Objekt, das detektiert wird. **Fig. 6D** zeigt das Hintergrundmodell des visuellen Vordergrund-Objekts von **Fig. 6C**. Es wird anerkannt, dass das Hintergrundmodell den gleichen Treppenlauf zeigt, ohne dass die Person oder ein anderes visuelles Objekt im Vordergrund vorhanden ist.

[0123] **Fig. 6E** zeigt ein Fahrzeug, das über ein Straßensegment fährt. Das Fahrzeug ist das visuelle Vordergrund-Objekt, das detektiert wird. **Fig. 6F** zeigt das Hintergrundmodell des visuellen Vordergrund-Objekts von **Fig. 6E**. Es wird anerkannt, dass das Hintergrundmodell das gleiche Straßensegment zeigt, ohne dass das Fahrzeug oder ein anderes visuelles Vordergrund-Objekt vorhanden ist.

[0124] Nach verschiedenen Beispielausführungsformen wird von einem historischen Bildrahmen ein Hintergrundmodell eines detektierten visuellen Objekts bestimmt. Ein visuelles Vordergrund-Objekt wird in einem bestimmten Teilbereich eines aktuellen Bildrahmens eines Sequenzbildes von Bilddaten detektiert, die ein vom Videoerfassungsgerät erfasstes Video bilden. Ein historischer Bildrahmen ist ein vorhergehender Bildrahmen in der Folge von Bildrahmen, in denen das visuelle Vordergrund-Objekt und jedes andere visuelle Vordergrund-Objekt in diesem vorherigen Bildrahmen fehlen. In diesem Fall stellen der aktuelle Bildrahmen und der historische Bildrahmen die gleiche Szene dar. Das heißt, das Videoerfassungsgerät ist statisch (d.h. unbewegt) zwischen dem Zeitpunkt des historischen Bildrahmens und dem aktuellen Bildrahmen, so dass das Videoerfassungsgerät die gleiche Szene aufnimmt. Ein bestimmter Teilbereich des historischen Bildrahmens, der dem Teilbereich des aktuellen Bildrahmens entspricht, in dem sich das visuelle Vordergrund-Objekt befindet, wird aus dem historischen Bildrahmen ausgeschnitten. Der auf diese Weise ausgeschnittene historische Bildrahmen ist das Hintergrundmodell des detektierten visuellen Vordergrund-Objekts. Dieser ausgeschnittene historische Bildrahmen wird bei **556** als negatives Trainingsbeispiel für die Weiterbildung des Basisklassifikators zur bereitgestellt.

[0125] Nach verschiedenen Beispielausführungsformen kann zunächst ein komplettes Hintergrundmodell der gesamten Szene erstellt werden. Das Hintergrundmodell einer bestimmten Teilbereichs der Szene kann dann aus dem kompletten Hintergrundmodell extrahiert werden.

[0126] Zum Beispiel in einer weniger belebten Szene, in der es nur wenige visuelle Vordergrund-Objekte gibt, kann ein einzelner historischer Bildrahmen, der völlig frei von Vordergrund-Objekten ist, als vollständiges Hintergrundmodell verwendet werden.

[0127] In einer geschäftigen Szene kann es immer mindestens ein visuelles Vordergrund-Objekt geben. Für solche Szenen kann das komplette Hintergrundmodell konstruiert werden, indem verschiedene Teilbereiche aus einer Vielzahl von historischen Bildrahmen zu einem vollständigen Hintergrundmodell aggregiert werden.

[0128] Gemäß einem Beispiel werden mehrere historische Bilderrahmen ausgewählt. Jeder dieser historischen Bildrahmen enthält mindestens einen Teilbereich des Bildrahmens, der frei von Vordergrund-objekten ist.

[0129] Es werden die Koordinaten eines oder mehrerer Teilbereiche bestimmt, die frei von Vordergrund-Objekten jedes ausgewählten historischen Bildrahmens sind. Diese Teilbereiche können aus ihrem jeweiligen historischen Bildrahmen ausgeschnitten werden.

[0130] Die Teilbereiche, wie sie aus der Vielzahl der historischen Bilder ausgeschnitten wurden, werden dann zu einem aggregierten Bild aggregiert. Ein aggregiertes Bild, das die gesamte Szene darstellt, kann durch geeignete Auswahl der Vielzahl von historischen Bildrahmen erhalten werden, so dass die Teilbereiche dieser Rahmen, die frei von jeglichen Vordergrund-Objekten sind, gemeinsam die gesamte Szene abdecken. Dementsprechend bildet das aggregierte Bild ein vollständiges Hintergrundmodell der Szene. Beispielsweise können die Bildunterbereiche, wie sie aus der Vielzahl der historischen Bilder ausgeschnitten wurden, zu einem aggregierten Bild nach in der Technik bekannten Zusammenfügungsverfahren zusammengefügt werden.

[0131] Dementsprechend, nachdem ein visuelles Vordergrund-Objekt innerhalb eines bestimmten Teilbereichs einer Szene detektiert wurde, kann das Hintergrundmodell dieses Teilbereichs durch Ausschneiden eines Teilbereichs des aggregierten Bildes erhalten werden, der dem gegebenen Teilbereich entspricht, in dem das visuelle Objekt detektiert wird.

[0132] **Fig. 7A** zeigt einen ersten vollständigen historischen Bilderrahmen, der eine Beispielszene darstellt, der ein Platz ist. Es wird anerkannt, dass der erste Teilbereich **700**, der einen Teil des Essbereiches und der Grasfläche abdeckt, frei von visuellen Vordergrund-Objekten ist. Dementsprechend kann der erste Teilbereich **700** als eine der zu aggregierenden Teilbereiche zur Bildung des vollständigen Hintergrundmodells verwendet werden. In dem zweiten Teilbereich **708**, der die Stufen umfasst, befindet sich jedoch eine Person. Da dieser zweite Teilbereich **708** im ersten vollen historischen Bildrahmen ein visuelles Vordergrund-Objekt enthält, kann er nicht für den Aufbau des kompletten Hintergrundmodells verwendet werden.

[0133] **Fig. 7B** zeigt einen zweiten vollständigen historischen Bilderrahmen, der die gleiche Szene des Platzes darstellt. Der zweite vollständige historische Bilderrahmen wurde zu einem späteren Zeitpunkt erfasst als der erste vollständige historische Bilderrahmen. Es wird anerkannt, dass der zweite Teilbereich **708** im zweiten vollständigen historischen Bildrahmen frei von einem visuellen Vordergrund-Objekt ist. Die Person, die sich in den Schritten in dem ersten vollständigen historischen Bilderrahmen befand, ist nun vollständig abgestiegen. Dementsprechend kann dieser zweite Teilbereich **708** im zweiten vollständigen historischen Bildrahmen als eine der zu aggregierenden Teilbereiche zur Bildung des vollständigen Hintergrundmodells verwendet werden. Andere Teilbereiche der Szene, die zur Bildung des kompletten Hintergrundmodells geeignet sind, können auf die gleiche Weise bestimmt werden.

[0134] Es soll nun auf **Fig. 8** verwiesen werden, worin ein Flussdiagramm eines verbesserten computerimplementierten Verfahrens **558** zum weiteren Training eines Basisklassifikators nach einer alternativen exemplarischen Ausführungsform dargestellt ist. Das alternative Beispielverfahren **558** enthält die gleichen Schritte wie das Verfahren **540**, aber auch die zusätzlichen Schritte **560** und **564**.

[0135] Bei **560** wird ein falsch klassifizierter Teilbereich einer Szene bereitgestellt. Ein falsch klassifizierter Teilbereich einer Szene bezieht sich auf einen Teilbereich, in der ein Objektklassifikator den Teilbereich fälschlicherweise als zu einer bestimmten Klasse gehörig klassifiziert hat, wenn der Teilbereich tatsächlich keine Objekte dieser Klasse enthält.

[0136] Ein falsch klassifizierter Teilbereich kann in einer überwachten Umgebung bestimmt werden, in der von einem Objektklassifikator klassifizierte Objekte von einem Menschen überprüft werden, der alle vom Objektklassifikator vorgenommenen Fehlklassifikationen identifiziert.

[0137] Ein falsch klassifizierter Teilbereich kann in einer teilweise überwachten oder völlig unbeaufsichtigten Umgebung bestimmt werden. In einigen Beispielen können Teilbereiche von Bildrahmen, in denen keine Objekte vorhanden sind, einem Objektklassifikator zugeführt werden. Jede Klassifizierung durch den Objektklassifikator, dass der Teilbereich ein Objekt enthält, das zu einer bestimmten Klasse gehört (anderen als dem Hintergrund), wird fehlerhaft sein und ist als falsch klassifizierter Teilbereich identifiziert.

[0138] Die Szene, in der ein falsch klassifizierter Teilbereich identifiziert wird, kann dieselbe Szene sein wie die Szene, in der ein visuelles Vordergrund-Objekt bei 544 detektiert wird. Alternativ kann sich die Szene des falsch klassifizierten Teilbereichs von der Szene unterscheiden, in der das visuelle Vordergrund-Objekt detektiert wird.

[0139] Bei **564** wird der Basisklassifikator durch Anwenden von maschinellen Lernen auf den Basisklassifikator unter Verwendung des falsch klassifizierten Teilbereichs als negatives Trainingsbeispiel weiter trainiert.

[0140] Bei **516**, dem Klassifikator, der wie aus dem Hintergrundmodell des detektierten visuellen Objekts trainiert ist, wird der falsch klassifizierte Teilbereich, und, optional, das detektierte visuelle Objekt zur Klassifizierung weiterer detektierter visueller Objekte eingesetzt.

[0141] Es soll nun auf **Fig. 9** verwiesen werden, worin nun ein Flussdiagramm eines verbesserten computerimplementierten Verfahrens **600** zum szenenspezifischen Training eines Basisklassifikators nach einer beispielhaften Ausführungsform dargestellt ist. Es wird davon ausgegangen, dass zahlreiche Schritte des Beispielverfahrens **600** ähnlich oder gleich den Schritten des Beispielverfahrens **540** sind und dass die Beschreibung, die in Bezug auf das Beispielverfahrens **540** vorhanden ist, auch auf das Beispielverfahren **600** anwendbar ist. Es wird davon ausgegangen, dass das szenenspezifische Verfahren **600** auch nach dem alternativen Beispielverfahren **560** angewendet werden kann.

[0142] Bei **504** ist ein Basisklassifikator vorgesehen.

[0143] Nach der Bereitstellung des Basisklassifikators beginnt das Training des Basisklassifikators. Der Basisklassifikator wird speziell für eine aktuelle reale Szene trainiert. Die aktuelle Szene kann dem Sichtfeld einer bestimmten Kamera entsprechen, die an einem bestimmten Ort positioniert und in eine bestimmte Richtung ausgerichtet ist.

[0144] Bei **544** wird ein visuelles Vordergrund-Objekt in den Bilddaten der aktuellen Szene erkannt.

[0145] Bei **548** wird ein Hintergrundmodell des erkannten Objekts bestimmt.

[0146] Bei **552** wird der Basisklassifikator optional durch Anwenden maschinellen Lernens auf den Basisklassifikator trainiert, wobei die aus der aktuellen Szene bei **544** detektierten visuellen Vordergrund-Objekte als positives Trainingsbeispiel verwendet werden.

[0147] Bei **556** wird der Basisklassifikator durch Anwenden maschinellen Lernens auf den Basisklassifikator unter Verwendung des Hintergrundmodells des visuellen Vordergrund-Objekts, das bei 548 als negatives Trainingsbeispiel bestimmt wurde, trainiert.

[0148] Bei **516**, der Basisklassifikator, der auf Basis von visuellen Vordergrund-Objekten und/oder Hintergrundmodellen der aktuellen Szene trainiert wurde, wird zur Klassifizierung der in der aktuellen Szene gefundenen Objekte eingesetzt.

[0149] Es wird verständlich sein, dass solange die aktuelle Szene unverändert bleibt, können die Schritte **544** bis **556** wiederholt werden, um den Basisklassifikator durch Anwenden maschinellen Lernens anhand einer Vielzahl von Trainingsbeispielen aus der aktuellen Szene weiter zu trainieren. Wie an anderer Stelle beschrieben, können die Schritte **544** bis **556** wiederholt werden, so dass die Aktualisierung des Basisklassifikators einem iterativen Prozess folgt.

[0150] Bei **608** wird bestimmt, ob sich die aktuelle Szene geändert hat. Eine solche Änderung in der aktuellen Szene kann aufgrund einer Änderung der Position der Kamera, die die Szene erfasst hat, auftreten. Eine solche Änderung kann auch aufgrund einer Änderung der Ausrichtung der Kamera, die die Szene erfasst hat, auftreten. Eine solche Änderung kann auch durch eine Änderung einer Einstellung der Kamera, die die

Szene erfasst hat, wie z.B. eine signifikante Änderung des von der Kamera verwendeten Zooms oder eines Betriebsmodus der Kamera (z.B. Umschalten von Normallicht auf Schwachlicht), hervorgerufen werden.

[0151] Wenn die Szene unverändert bei **608** bleibt, kann das Verfahren **600** zu **544** zurückkehren, um weitere visuelle Objekte innerhalb der Szene zu detektieren und zu klassifizieren. Alternativ kann die Verfahren **600** zu **516** zurückkehren, um den Objektklassifikator wie in den Schritten **544** bis **556** für die aktuelle Szene trainiert weiter zu verwenden.

[0152] Wird die Szene bei **608** geändert, fährt das Verfahren mit Schritt **616** fort, um zumindest teilweise zum Basisklassifikator zurückzukehren. In einigen Beispielen wird der Objektklassifikator, der derzeit bei **516** eingesetzt wird, bei einem Szenenwechsel vollständig auf den Basisklassifikator zurückgesetzt.

[0153] Nach der Rückkehr zum Basisklassifikator bei Schritt **616** kann die neue Szene, die sich aus dem Szenenwechsel ergibt, als aktuelle Szene eingestellt werden. Das Verfahren **600** kann dann zu **544** zurückkehren, um visuelle Vordergrund-Objekte zu erkennen und zu klassifizieren, die in der „neuen“ aktuellen Szene gefunden wurden. Diese Objekte und/oder Hintergrundmodelle, die diesen Objekten entsprechen, können nach dem Zurücksetzen von Schritt **616** zur Aktualisierung des Basisklassifikators verwendet werden.

[0154] Es wird anerkannt, dass der Rückgriff auf den Basisklassifikator in Situationen nützlich sein kann, in denen die Charakteristika einer Ausgangsszene und einer nachfolgenden Szene signifikant unterschiedlich sind, so dass das Training des Basisklassifikators gemäß den Charakteristika der Ausgangsszene auf die nachfolgende Szene nicht anwendbar ist. Die Rückkehr zum Basisklassifikator ermöglicht es, den Klassifikator gezielt für die Eigenschaften der nachfolgenden Szene umzuschulen.

EXPERIMENT

[0155] Einem Experiment zufolge wurde die Leistung eines Basisklassifikators (eine spezifische Architektur eines neuronalen Netzwerks mit tiefer Faltung, bekannt als „AlexNet“, beschrieben in Alex Krizhevsky, Ilya Sutskever, Geoffrey Hinton, „ImageNet Classification with deep convolution neural networks“, NIPS 2012) anhand verschiedener Trainingsbeispiele bewertet.

[0156] Schulungsbeispiele stammen aus dem VIRAT-Datensatz (<http://www.viratdata.org>). Dieser Datensatz enthält mehr als 300 Videos von verschiedenen statischen Kameras. Visuelle Objekte der menschlichen Klasse und visuelle Objekte der Fahrzeugklasse wurden aus dem VIRAT-Datensatz extrahiert und als erste Trainingsbeispiele verwendet. Für jedes visuelle Objekt wurde ein Hintergrundmodell bestimmt, das als Trainingsbeispiel dient. Diese Hintergrundmodelle werden als zweiter Satz von Trainingsbeispielen verwendet.

[0157] Neben der Menschenklasse und der Fahrzeugklasse wurden auch Trainingsbeispiele der Hintergrundklasse aus dem VIRAT-Datensatz extrahiert. Zur Generierung der Schulungsbeispiele der Hintergrundklasse wurden Bildmuster erstellt, die kein visuelles Vordergrund-Objekt der menschlichen Klasse oder ein visuelles Vordergrund-Objekt der Fahrzeugklasse enthalten. Jedes Bildbeispiel ist ein Ausschnitt aus einem Bildrahmen der im VIRAT-Datensatz enthaltenen Videos. Ein einfacher Objektklassifikator, der nicht auf dem neuronalen Netzwerkklassifikator basiert, wird verwendet, um diese Bildproben zu klassifizieren. Eine Fehlklassifizierung liegt vor, wenn der einfache Klassifikator eine der Bildproben als ein visuelles Objekt klassifiziert, das in die menschliche Klasse oder die Fahrzeugklasse fällt. Diese falsch klassifizierten Bildbeispiele sind in einem dritten Satz von Trainingsbeispielen enthalten.

[0158] Der AlexNet-Klassifikator wird als Basisklassifikator zur Verfügung gestellt, der durch die aus dem VTRAT-Datensatz extrahierten Trainingsbeispiele trainiert werden soll. Die positiven und negativen Trainingsbeispiele werden verwendet, um den Basisklassifikator mit dem Caffe Deep Learning Framework des Berkeley Vision and Learning Center zu trainieren (verfügbar unter: caffe.berkeleyvision.org). Die Aktualisierung des Basisklassifikators erfolgte auf einer Tesla K80 GPU.

[0159] In einem ersten Teil des Experiments wurde der Basisklassifikator trainiert, indem der erste Satz von Beispielen als positive Trainingsbeispiele (100 positive Trainingsbeispiele) und der zweite Satz von Trainingsbeispielen als negative Trainingsbeispiele (100 negative Trainingsbeispiele) verwendet wurde. Dieses Training des Basisklassifikators führte zu einem ersten geschulten Testklassifikator.

[0160] In einem zweiten Teil des Experiments wurde der Basisklassifikator trainiert, indem der erste Satz von Beispielen als positive Trainingsbeispiele (100 positive Trainingsbeispiele) und der dritte Satz von Trainings-

beispielen als negative Trainingsbeispiele (100 negative Trainingsbeispiele) angewendet wurden. Dieses Training des Basisklassifikators führte zu einem zweiten trainierten Testklassifikator.

[0161] In einem dritten Teil des Experiments wurde der Basisklassifikator trainiert, indem der erste Satz von Beispielen als positive Trainingsbeispiele (100 positive Trainingsbeispiele) und eine Mischung aus dem zweiten Satz von Trainingsbeispielen und dem dritten Satz von Trainingsbeispielen als negative Trainingsbeispiele verwendet wurde. Genauer gesagt wurden 50 Trainingsbeispiele aus dem zweiten Satz und 50 Trainingsbeispiele aus dem dritten Satz als negative Trainingsbeispiele für das Training des Basisklassifikators verwendet. Dieses Training des Basisklassifikators führte zu einem dritten trainierten Testklassifikator.

[0162] Jeder der ersten trainierten Testklassifikator, der zweite trainierten Testklassifikator und der dritte trainierten Testklassifikator wurden für die Objektklassifizierung auf einem Testsatz von Videos aus einem internen Videodatensatz eingesetzt. Die Fehlerquote beim Einsatz der einzelnen Klassifikatoren wurde gemessen. Ein Fehler liegt vor, wenn ein visuelles Objekt falsch klassifiziert wurde oder wenn ein Hintergrundbild (z.B. kein visuelles Objekt im Vordergrund) als visuelles Objekt der Klasse Mensch oder der Klasse Fahrzeug klassifiziert wurde.

[0163] Tabelle 1 ist eine Verwirrungsmatrix, die die Leistung des ersten trainierten Testklassifikators bei der Klassifizierung von visuellen Vordergrund-Objekten im zeigt, die im Testset der Videos enthalten sind.

Tabelle 1:

		KLASSIFIZIERT (%)		
		Menschlich	Fahrzeug	Hintergrund
IST	Menschlich	36.26	9.58	3.26
	Fahrzeug	0.94	26.78	0.29
	Hintergrund	0.14	0.15	22.60

[0164] Tabelle 2 ist eine Verwirrungsmatrix, die die Leistung des zweiten trainierten Testklassifikators bei der Klassifizierung von visuellen Vordergrund-Objekten zeigt, die im Testset der Videos enthalten sind.

Tabelle 2:

		KLASSIFIZIERT (%)		
		Menschlich	Fahrzeug	Hintergrund
IST	Menschlich	38.80	8.56	1.74
	Fahrzeug	4.49	23.36	0.16
	Hintergrund	0.32	0.14	22.42

[0165] Tabelle 3 ist eine Verwirrungsmatrix, die die Leistung des dritten trainierten Testklassifikators bei der Klassifizierung von visuellen Vordergrund-Objekten zeigt, die im Testset der Videos enthalten sind.

Tabelle 3:

		KLASSIFIZIERT (%)		
		Menschlich	Fahrzeug	Hintergrund
IST	Menschlich	43.26	4.72	1.11
	Fahrzeug	3.49	24.39	0.14
	Hintergrund	0.34	0.12	22.43

[0166] Die Fehlerquote des ersten trainierten Testklassifikators beträgt 14,36%, die des zweiten trainierten Testklassifikators 15,42% und der Fehler des dritten trainierten Testklassifikators 9,92%.

[0167] Es wird anerkannt, dass die Verwendung von Hintergrundmodellen von visuellen Vordergrund-Objekten für das Training des Basisklassifikators (der erste trainierte Testklassifikator und der dritte trainierte Testklassifikator) geringere Fehlerraten als der zweite trainierte Testklassifikator aufwies, bei dem Hintergrundmodelle nicht als Trainingsbeispiele verwendet wurden. Die geringere Fehlerquote ist ein Indikator für eine verbesserte Performance. Noch wichtiger wird anerkannt, dass die Verwendung einer Kombination von Hintergrundmodellen von visuellen Vordergrund- und Hintergrundklassenobjekten zusammen als negative Trainingsbeispiele eine deutlich verbesserte Leistung zeigen (35,6% niedrigere Fehlerrate im Vergleich zum zweiten trainierten Testklassifikator).

[0168] Ohne an eine bestimmte Theorie gebunden zu sein, kann die Verwendung von Hintergrundmodellen erkannter visueller Objekte als negative Trainingsbeispiele für das Training eines Klassifikators die Häufigkeit der Fehlklassifizierung von Objekten einer Szene reduzieren, die ansonsten zum Hintergrund der Szene gehören.

[0169] In Anlehnung an **Fig. 6C** wird der Teilbereich der im Bild gezeigten Szene mit der Person und einem Laternenpfahl anerkannt. Die Person ist ein visuelles Vordergrund-Objekt und der Laternenpfahl bildet den Teil des Hintergrundes der Szene. Wenn dieser Teilbereich jedoch als positives Trainingsbeispiel verwendet wird, kann der Basisklassifikator dazu veranlasst werden, den Laternenpfahl als visuelles Vordergrund-Objekt der Personenklasse zu erkennen. Entspricht dieser Teilbereich der Szene beispielsweise einem realen Ort, der oft ein Objekt von Interesse ist (z.B. ein häufig genutzter Flur, Weg oder eine Straße), kann der Laternenpfahl in mehreren Teilbereichen erscheinen, die jeweils als positive Trainingsbeispiele verwendet werden. Dies kann die Wahrscheinlichkeit erhöhen, dass der Klassifikator trainiert wird, den Laternenpfahl als Vorkommnis eines Objekts der Personenklasse zu erkennen. Die Verwendung des Hintergrundmodells des Teilbereichs als negatives Trainingsbeispiel kann diesem Effekt zumindest teilweise entgegenwirken, indem der Klassifikator trainiert wird, dass der Laternenpfahl Teil des Hintergrundes der Szene ist.

[0170] Ebenso wird durch das Training eines Klassifikators anhand des in **Fig. 6F** gezeigten Hintergrundmodells der Klassifikator trainiert, den vertikalen Strahl als Teil des Hintergrunds zu erkennen, wodurch die Möglichkeit, den vertikalen Strahl oder ähnliche Objekte als einer menschlichen Klasse oder Fahrzeugklasse zuzuordnen, reduziert wird.

[0171] Ganz allgemein und ohne an eine bestimmte Theorie gebunden zu sein, führt das Training eines Klassifikators mit Hilfe eines Hintergrundmodells dazu, dass ein Klassifikator trainiert wird, reale Objekte, die Teil des Hintergrundes einer Szene sind, korrekt als Hintergrundobjekte zu erkennen. In einem Teilbereich einer Szene, in dem häufig ein visuelles Vordergrund-Objekt erkannt wird, kann die Verwendung des Hintergrundmodells dieses Teilbereichs beispielsweise als negatives Trainingsbeispiel die Wahrscheinlichkeit verringern, dass der Klassifikator Objekte, die Teil des Hintergrunds sind, fälschlicherweise als visuelle Vordergrund-Objekte einer bestimmten Klasse klassifiziert.

[0172] Während die obige Beschreibung Beispiele für die Ausführungsformen enthält, wird anerkannt, dass einige Merkmale und/oder Funktionen der beschriebenen Ausführungsformen modifizierbar sind, ohne vom Geist und den Funktionsprinzipien der beschriebenen Ausführungsformen abzuweichen. Dementsprechend soll das oben Beschriebene nicht einschränkend dargestellt werden und es wird von Fachleuten verstanden, dass andere Varianten und Änderungen vorgenommen werden können, ohne vom Anwendungsbereich der Erfindung, wie er in den beigefügten Ansprüchen definiert ist, abzuweichen.

ZITATE ENTHALTEN IN DER BESCHREIBUNG

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

Zitierte Patentliteratur

- US 8934709 [0081]
- US 7627171 [0083]
- US 8224029 [0084]

Patentansprüche

1. Verfahren zum Trainieren eines computerimplementierten Objektklassifikators, wobei das Verfahren umfasst:

Erkennen eines visuellen Vordergrund-Objekts innerhalb eines Teilbereichs einer Szene;

Bestimmen eines Hintergrundmodells des Teilbereichs der Szene, wobei das Hintergrundmodell den Teilbereich darstellt, wenn ein visuelles Vordergrund-Objekt davon abwesend ist; und

Trainieren des Objektklassifikators durch computerimplementiertes maschinelles Lernen unter Verwendung des Hintergrundmodells des Teilbereichs als ein negatives Trainingsbeispiel.

2. Verfahren nach Anspruch 1, ferner umfassend Trainieren des Objektklassifikators durch maschinelles Lernen unter Verwendung des detektierten visuellen Vordergrund-Objekts als ein positives Trainingsbeispiel.

3. Verfahren nach einem der Ansprüche 1 oder 2, wobei Bestimmen des Hintergrundmodells des Teilbereichs der Szene umfasst:

Auswählen eines historischen Bildrahmens, der aufgenommen ist, als irgendein Vordergrund-Objekt in einem Teilbereich des historischen Bildrahmens fehlt, der dem Teilbereich der Szene entspricht; und

Ausschneiden aus dem historischen Bildrahmen des Teilbereichs, der dem Teilbereich der Szene entspricht, wobei der ausgeschnittene Bildrahmen das Hintergrundmodell des Teilbereichs der Szene ist.

4. Verfahren nach einem der Ansprüche 1 oder 2, wobei Bestimmen des Hintergrundmodells des Teilbereichs der Szene umfasst:

Bestimmen, innerhalb jedes von einer Vielzahl von historischen Bildrahmen, eines oder mehrerer Teilbereiche, die frei von jeglichen Vordergrundobjekten sind;

Aggregieren der einen oder mehreren Teilbereiche aus der Vielzahl der historischen Bilder um ein vollständiges Hintergrundbild zu bilden, das die gesamte Szene darstellt; und

aus dem gesamten Hintergrundbild einen Teilbereich entsprechend des Teilbereichs der Szene ausschneiden, wobei das ausgeschnittene komplette Hintergrundbild das Hintergrundmodell des Teilbereichs der Szene ist.

5. Verfahren nach Anspruch 4, wobei Aggregieren des einen oder der mehreren Teilbereiche aus der Vielzahl der historischen Bilder Zusammenfügen des einen oder der mehreren Teilbereiche umfasst, um ein Bild zu bilden, das die gesamte Szene darstellt.

6. Verfahren nach einem der Ansprüche 1 bis 5, wobei der Objektklassifikator speziell für eine aktuelle Szene trainiert wird.

7. Verfahren nach Anspruch 6, wobei bei einem Wechsel der aktuellen Szene zu einer neuen Szene auf den Objektklassifikator ohne das für die aktuelle Szene spezifische Training zurückgegriffen wird; und Trainieren des Objektklassifikators durch maschinelles Lernen mit Hilfe von Hintergrundmodellen aus der neuen Szene.

8. Verfahren nach einem der Ansprüche 1 bis 7, wobei der Objektklassifikator teilweise unter Verwendung von überwachtem Lernen vorbereitet wird.

9. Verfahren nach einem der Ansprüche 1 bis 8, wobei das computerimplementierte maschinelle Lernen aus mindestens einem von Faltungsneuronalen Netzen, Support-Vektor-Maschinen, Entscheidungsbäumen, Random Forests und Kaskaden-Klassifikatoren ausgewählt wird.

10. Verfahren nach einem der Ansprüche 1 bis 9, bei dem der Objektklassifikator durch computerimplementiertes maschinelles Lernen unter Verwendung eines falsch klassifizierten Teilbereichs einer Szene als negatives Trainingsbeispiel trainiert wird.

11. Ein computerimplementierter Objektklassifikator, der nach einem der Verfahren der Ansprüche 1 bis 10 trainiert ist.

12. System zur Segmentierung eines computerimplementierten Objektklassifikators, wobei das System umfasst:

einen Prozessor;

eine computerlesbare Speichervorrichtung, die Programmbefehle speichert, die, wenn sie vom Prozessor ausgeführt werden, das System veranlassen, Operationen durchzuführen, die Folgendes umfassen:

Erkennen eines visuellen Vordergrund-Objekts innerhalb eines Teilbereichs einer Szene;
Bestimmen eines Hintergrundmodells des Teilbereichs der Szene, wobei das Hintergrundmodell den Teilbereich darstellt, wenn ein visuelles Vordergrund-Objekt davon abwesend ist; und
Trainieren des Objektklassifikators durch computerimplementiertes maschinelles Lernen unter Verwendung des Hintergrundmodells des Teilbereichs als negatives Trainingsbeispiel.

13. System nach Anspruch 12, wobei die Operationen ferner das Training des Objektklassifikators durch maschinelles Lernen unter Verwendung des erkannten visuellen Vordergrundobjekts als positives Trainingsbeispiel umfassen.

14. System nach einem der Ansprüche 12 oder 13, wobei Bestimmen des Hintergrundmodells des Teilbereichs der Szene umfasst:

Auswählen eines historischen Bildrahmens, der aufgenommen ist, als irgendein beliebiges Vordergrund-Objekt in einem Teilbereich des historischen Bildrahmens fehlt, der dem Teilbereich der Szene entspricht;

Ausschneiden aus dem historischen Bildrahmen des Teilbereichs, der dem Teilbereich der Szene entspricht, wobei der ausgeschnittene Bildrahmen das Hintergrundmodell des Teilbereichs der Szene ist.

15. System eines der Ansprüche 12 oder 13, wobei Bestimmen des Hintergrundmodells des Teilbereichs der Szene umfasst:

Bestimmen, innerhalb jedes einer Vielzahl von historischen Bildrahmen, eines oder mehrerer Teilbereiche, die frei von jeglichen Vordergrundobjekten sind;

Aggregieren der einen oder mehreren Teilbereiche aus der Vielzahl der historischen Bilder zu einem vollständigen Hintergrundbild, das die gesamte Szene darstellt; und

Ausschneiden aus dem gesamten Hintergrundbild eines Teilbereichs entsprechend des Teilbereichs der Szene, wobei das ausgeschnittene vollständige Hintergrundbild das Hintergrundmodell des Teilbereichs der Szene ist.

16. System nach Anspruch 15, wobei das Aggregieren der einen oder mehreren Teilbereiche aus der Vielzahl von historischen Bildern Zusammenfügen der einen oder mehreren Teilbereiche umfasst, um ein Bild zu bilden, das die gesamte Szene darstellt.

17. Das System eines der Ansprüche 12 bis 16, wobei der Objektklassifikator speziell für eine aktuelle Szene trainiert wird.

18. Das System von Anspruch 17, wobei die Operationen weiterhin umfassen:

wenn die aktuelle Szene in eine neue Szene geändert wird, Zurückgreifen auf den Objektklassifikator ohne das für die aktuelle Szene spezifische Training; und

Trainieren des Objektklassifikators durch maschinelles Lernen unter Verwendung von Hintergrundmodellen aus der neuen Szene.

19. Das System eines der Ansprüche 12 bis 18, wobei der Objektklassifikator teilweise unter Verwendung von überwachtem Lernen vorbereitet wird.

20. Das System eines der Ansprüche 12 bis 19, wobei das computerimplementierte maschinelle Lernen aus Faltungsneuronalen Netzen, Support-Vektor-Maschinen, Entscheidungsbäumen, Random Forests und Kaskaden-Klassifikatoren ausgewählt ist.

21. Verfahren nach einem der Ansprüche 12 bis 20, wobei die Operationen ferner Trainieren des Objektklassifikators durch computerimplementiertes maschinelles Lernen unter Verwendung eines falsch klassifizierten Teilbereichs einer Szene als negatives Trainingsbeispiel umfassen.

Es folgen 13 Seiten Zeichnungen

Anhängende Zeichnungen

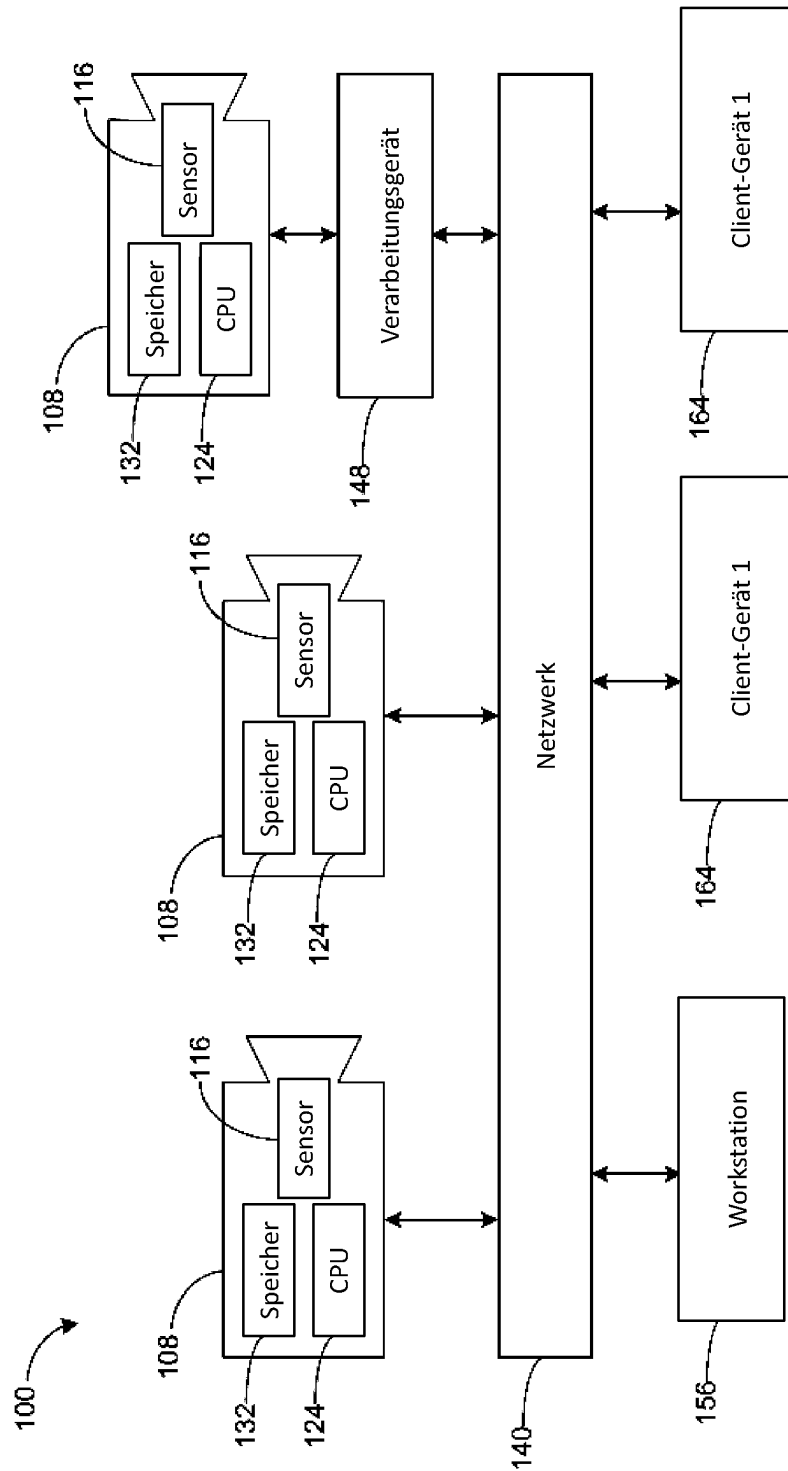


Fig. 1A

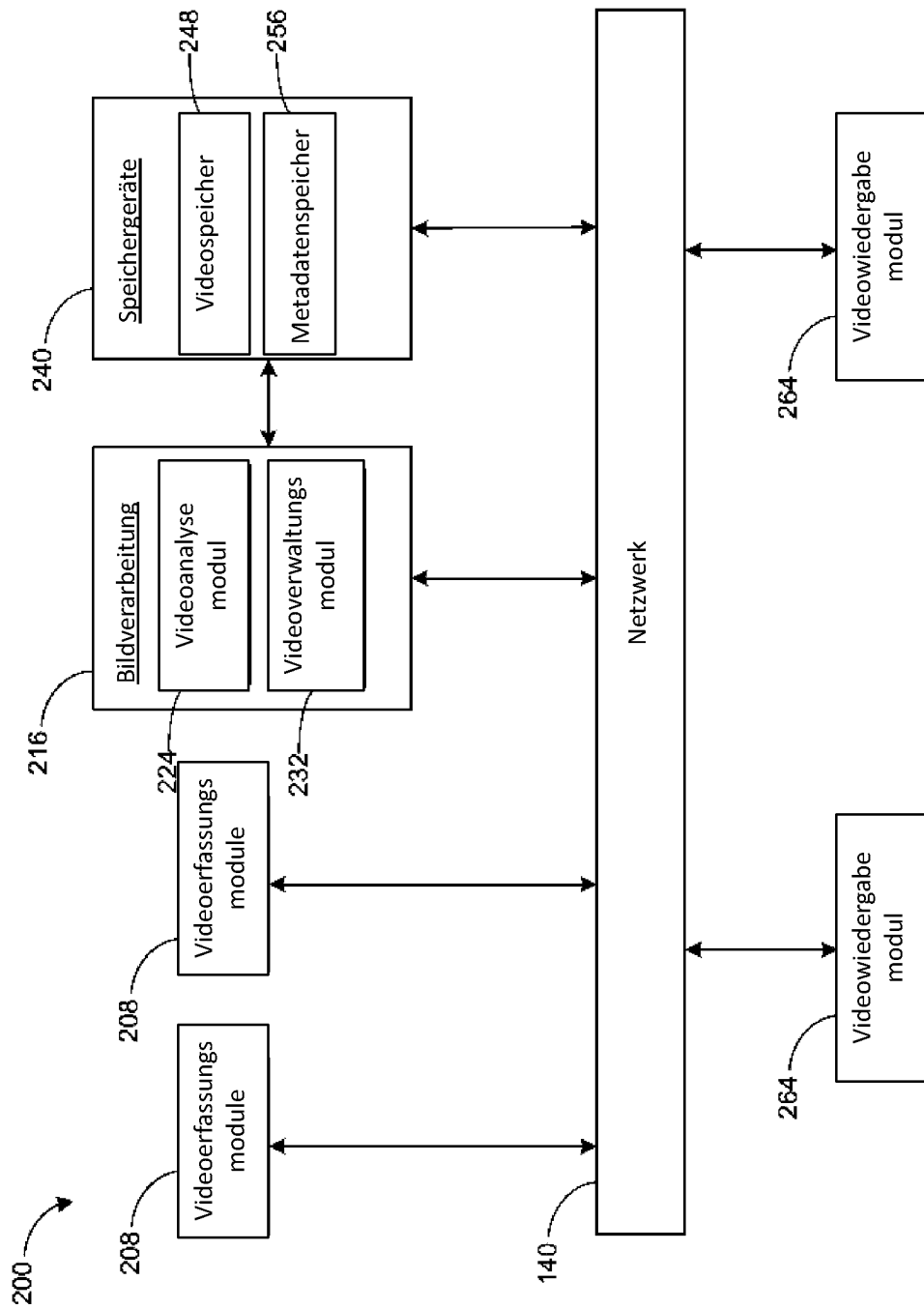


Fig. 1B

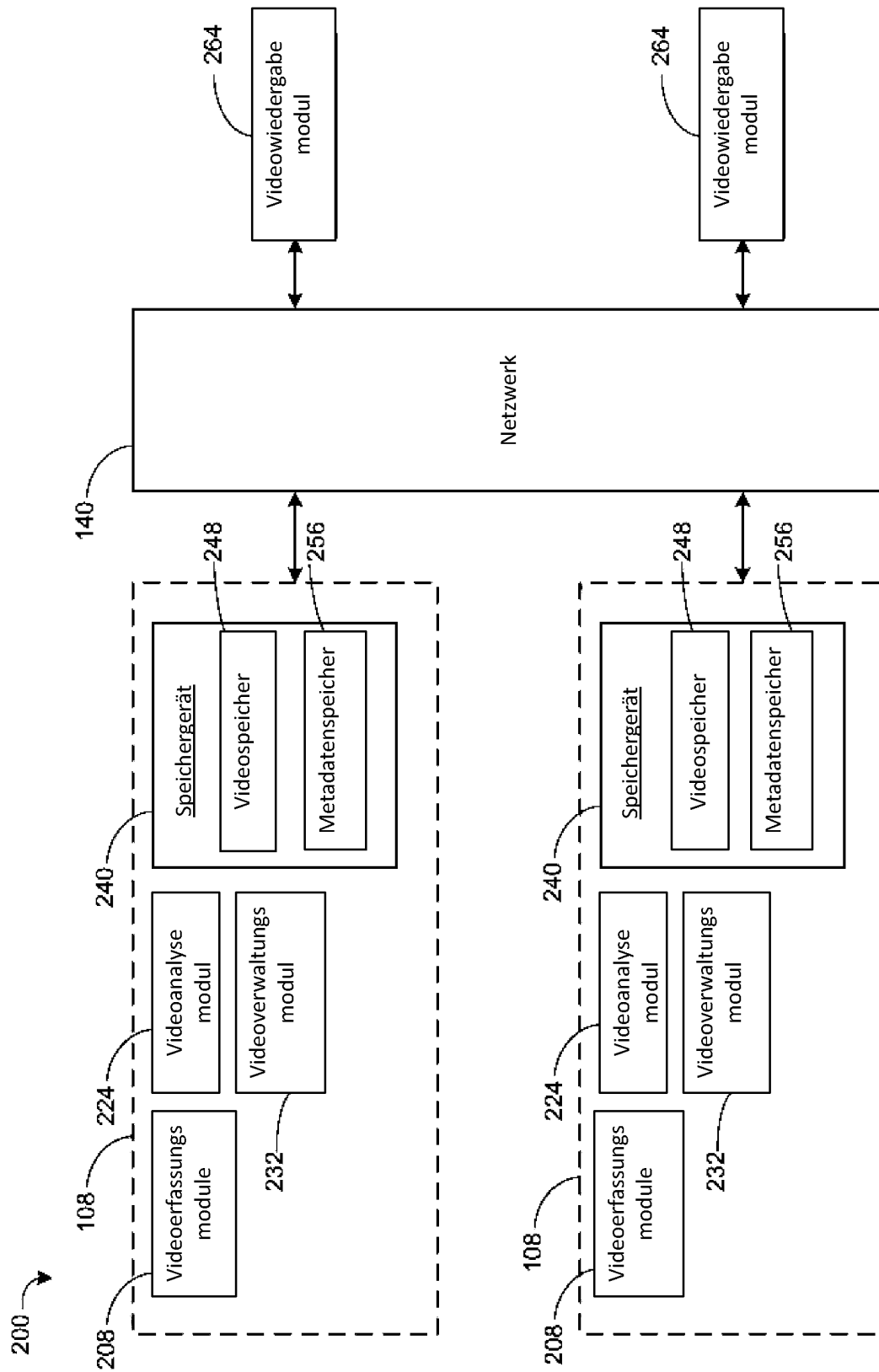


Fig. 1C

272 →

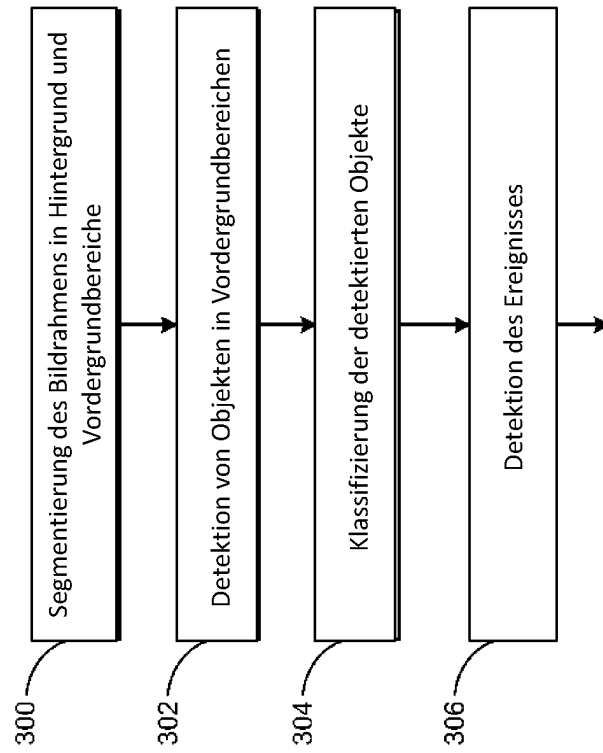


Fig. 2

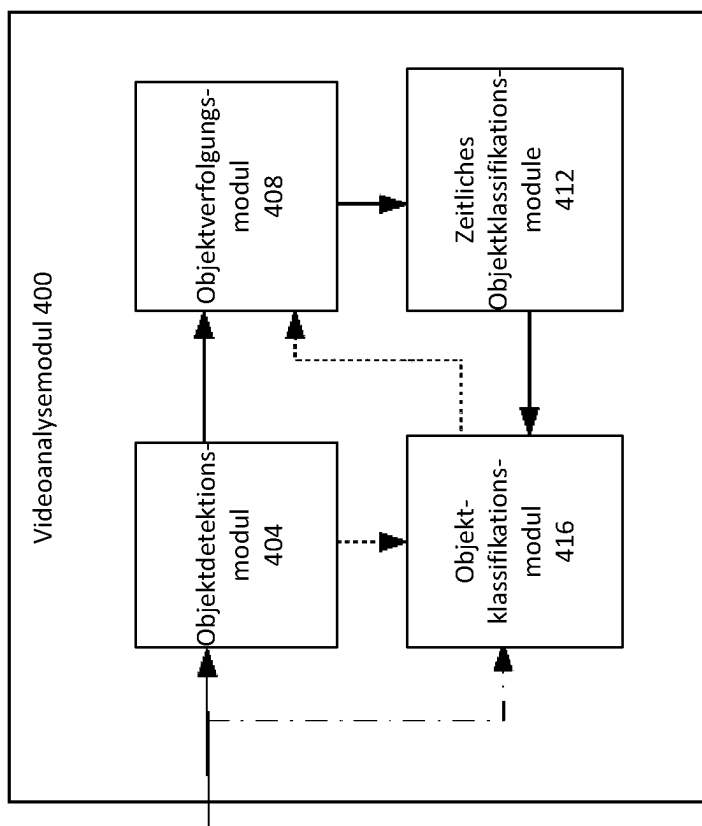


Fig. 3A

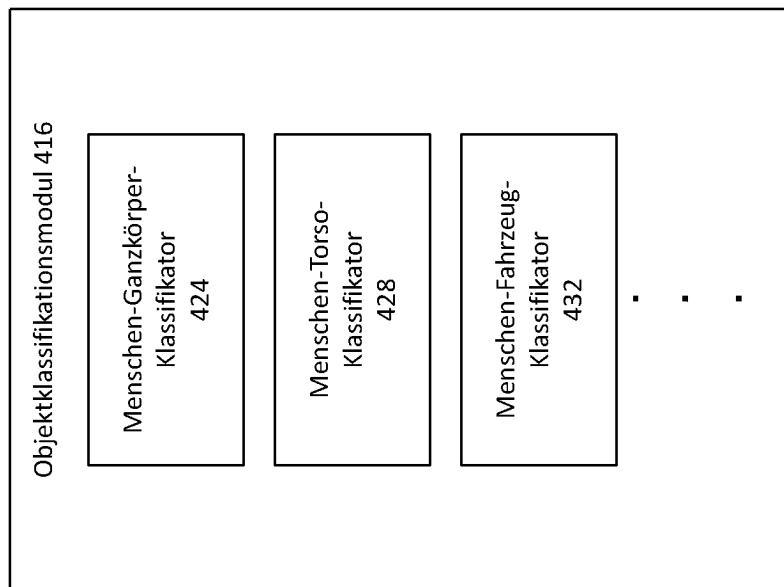


Fig. 3B

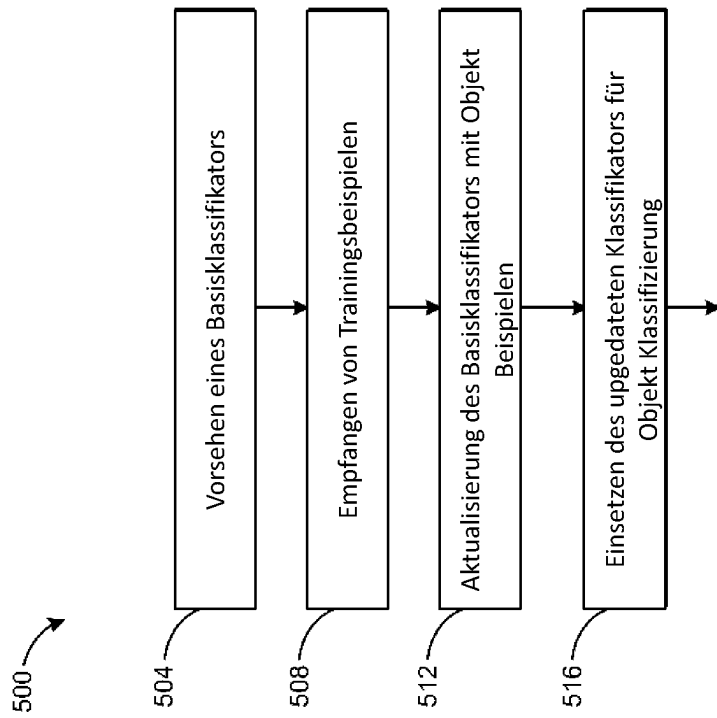


Fig. 4

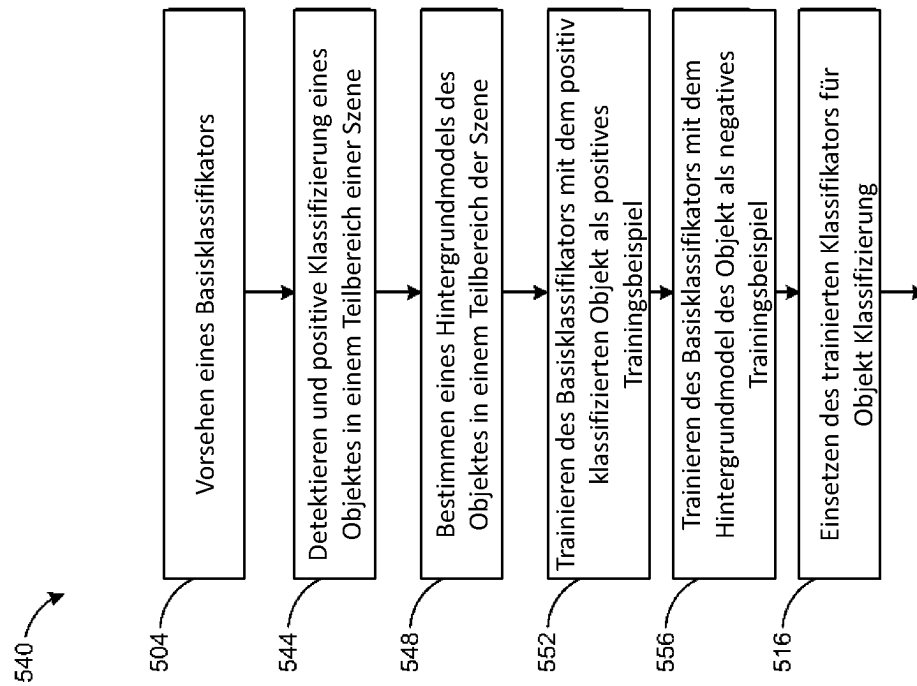


Fig. 5



Fig. 6A



Fig. 6B



Fig. 6C



Fig. 6D



Fig. 6E



Fig. 6F



Fig. 7A



Fig. 7B

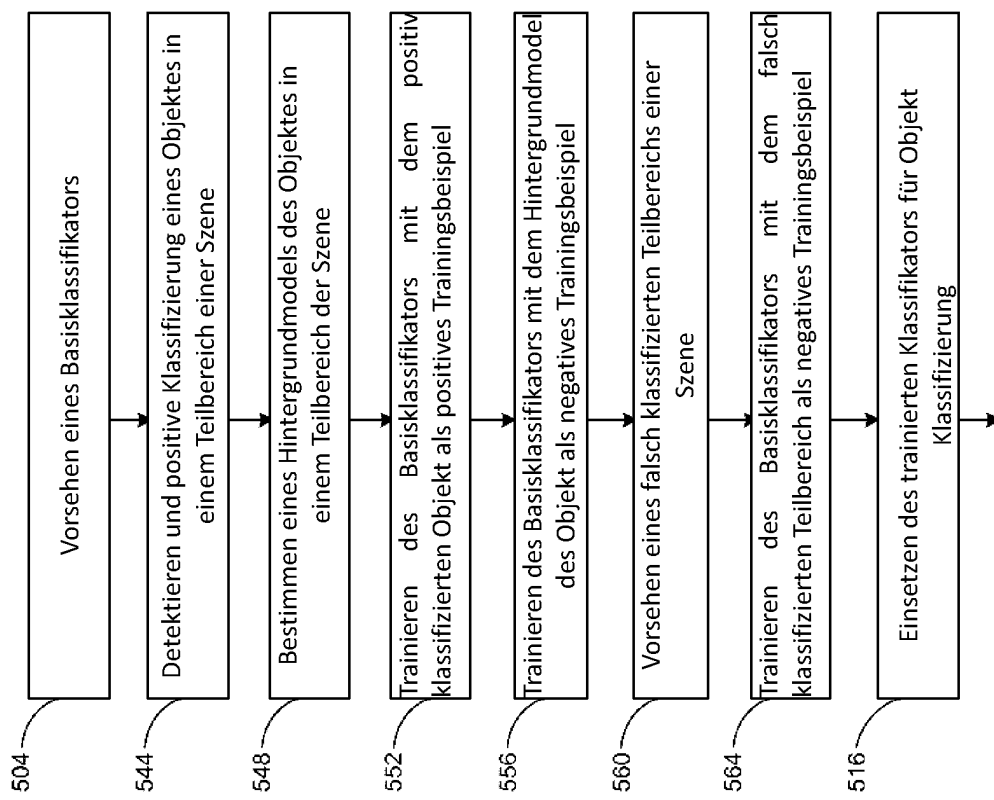


Fig. 8

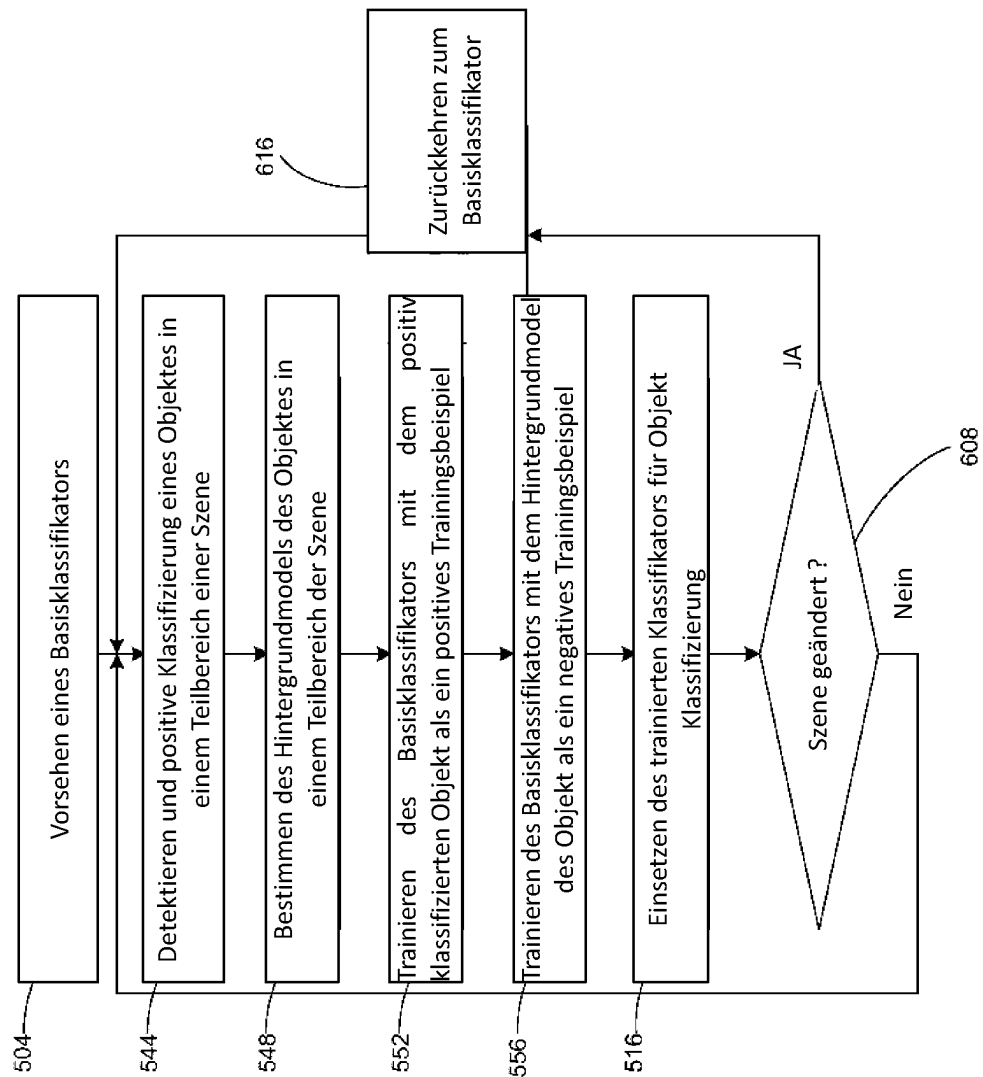


Fig. 9