



(10) **DE 11 2010 002 232 B4** 2021.12.23

(12)

Patentschrift

(21) Deutsches Aktenzeichen: **11 2010 002 232.1**
(86) PCT-Aktenzeichen: **PCT/US2010/036656**
(87) PCT-Veröffentlichungs-Nr.: **WO 2010/141369**
(86) PCT-Anmeldetag: **28.05.2010**
(87) PCT-Veröffentlichungstag: **09.12.2010**
(43) Veröffentlichungstag der PCT Anmeldung
in deutscher Übersetzung: **05.07.2012**
(45) Veröffentlichungstag
der Patenterteilung: **23.12.2021**

(51) Int Cl.: **G06K 9/40 (2006.01)**
G06K 9/46 (2006.01)
G06K 9/62 (2006.01)

Innerhalb von neun Monaten nach Veröffentlichung der Patenterteilung kann nach § 59 Patentgesetz gegen das Patent Einspruch erhoben werden. Der Einspruch ist schriftlich zu erklären und zu begründen. Innerhalb der Einspruchsfrist ist eine Einspruchsgebühr in Höhe von 200 Euro zu entrichten (§ 6 Patentkostengesetz in Verbindung mit der Anlage zu § 2 Abs. 1 Patentkostengesetz).

(30) Unionspriorität:

61/217,930	04.06.2009	US
12/789,292	27.05.2010	US

(73) Patentinhaber:

Honda Motor Co., Ltd., Tokyo, JP

(74) Vertreter:

**Weickmann & Weickmann Patent- und
Rechtsanwälte PartmbB, 81679 München, DE**

(72) Erfinder:

Ranganathan, Ananth, Raymond, Ohio, US

(56) Ermittelter Stand der Technik:
siehe Folgeseiten

(54) Bezeichnung: **Semantische Szenensegmentierung mittels Random multinomalem Logit (RML)**

(57) Hauptanspruch: Computer implementiertes Verfahren zum Lernen eines Random multinomialen Logit (RML) Klassifizierers (126) zur Szenensegmentierung, wobei das Verfahren umfasst:

Empfangen (210) eines Bildtrainingssatzes, worin der Bildtrainingssatz mehrere digitale Repräsentationen von Bildern (302) aufweist, und ein Objekt eines Bilds (302) des Bildtrainingssatzes eine semantische Kennung hat;

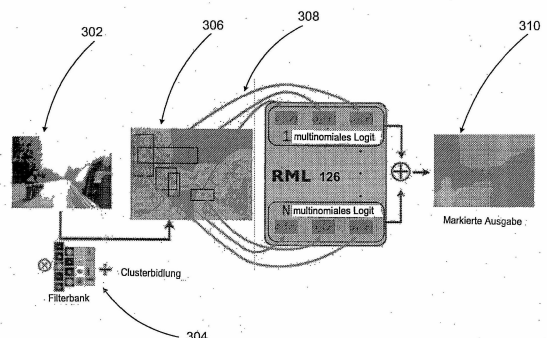
Erzeugen (212) einer Mehrzahl von Texton-Bildern (306) entsprechend den Bildern (302) in dem Bildtrainingssatz, worin ein Texton-Bild (306) seines entsprechenden Bilds (302) in dem Bildtrainingssatz ein Bild von Pixeln ist, und wobei jeder Pixelwert in einem Texton-Bild (306) durch eine Repräsentation des Pixelwerts seines entsprechenden Bilds (302) in dem Bildtrainingssatz ersetzt wird;

Wählen (214) von einem oder mehreren Textur-Layout-Merkmalen (308) aus der Mehrzahl der Texton-Bilder (306), wobei das Wählen von einem oder mehreren Textur-Layout-Merkmalen (308) umfasst: Tauschen eines Merkmals, das von dem RML-Klassifizierer (126) gegenwärtig verwendet wird, gegen ein zufällig gewähltes neues Merkmal basierend auf der statistischen Signifikanz des gegenwärtig verwendeten Merkmals;

Lernen (216) von mehreren multinomial logistischen Regressionsmodellen des RML-Klassifizierers (126) basierend auf den gewählten Textur-Layout-Merkmalen; und

Evaluiere (218) der Leistungsfähigkeit der mehreren mul-

tinomialen logistischen Regressionsmodelle basierend auf den semantischen Kennungen der Objekte in dem Bildtrainingssatz.



(56) Ermittelter Stand der Technik:

US 2008 / 0 075 361 A1

J. Malik, S. Belongie, T. Leung, and J. Shi: Contour and texture analysis for image segmentation. In: *Int'l. Journal of Computer Vision*, 43, June 2001, 1, 7-27.

MATTHEW ALASTAIR JOHNSON: Semantic Segmentation and Image Search. In: *Dissertation University of Cambridge*, April 2008, 1-171.

Prinzie Anita, Dirk Van den Poel: Random Forests for Multiclass classification: Random Multinomial Logit. In: *WORKING PAPER*, 2007/435, D/2006/7012/3X, Januar 2007, 1-31.

Song-Chun Zhu, Cheng-en Guo, Yingnian Wu, and Yizhou Wang: What Are Textons?. In: *A. Heyden et al. (Eds.): ECCV 2002*, 2002, 1-16.

Xiaofeng Ren and Jitendra Malik: Learning a Classification Model for Segmentation. In: *Proceedings of the Ninth IEEE International Conference on Computer Vision (ICCV'03)*, 2003, 1-8.

Beschreibung

Gebiet der Erfindung

[0001] Die vorliegende Erfindung betrifft allgemein Szenensegmentierung in Bildverarbeitungssystemen, und insbesondere die semantische Szenensegmentierung mittels multinomialen logistischen Regressionsmodellen.

Hintergrund der Erfindung

[0002] Bei der Computerbildanalyse, wie etwa bei intelligenten Transportsystemen, ist es eine häufige Aufgabe, Straßenszenen in einem aufgenommenen Bild zu klassifizieren. Diese Aufgabe beinhaltet häufig das Erfassen einer Straße, von anderen Fahrzeugen und Fußgängern, um einen Benutzer des intelligenten Transportsystems auf potentiell gefährliche Situationen aufmerksam zu machen. Eine Detektion von interessierenden Objekten in dem aufgenommenen Bild erfordert allgemein die Segmentierung des Bilds in interessierende Regionen und/oder eine weitere Segmentierung der interessierenden Regionen in interessierende Objekte.

[0003] US 2008/0075361 A1 zeigt eine semantische Segmentierung, wobei ein Bildtrainingsatz, Texton-Bilder und ein multinomialer Klassifizierer verwendet werden.

[0004] Dissertation MATTHEW ALASTAIR JOHNSON „Semantic Segmentation and Image Search“, University of Cambridge, April 2008; J. Malik, S. Belongie, T. Leung and J. Shi. „Contour and texture analysis for image segmentation“ Int'l Journal of Computer Vision, 43(1):7-27, June 2001 und Xiaofeng Ren and Jitendra Malik „Learning a Classification Model for Segmentation“ Proceedings of the Ninth IEEE International Conference on Computer Vision (ICCV'03), 2003, betreffen jeweils das Lernen von Klassifikatoren unter Verwendung von Textonen.

[0005] Prinzie Anita, Dirk Van den Poel „Random Forests for Multiclass classification: Random Multinomial Logit“ WORKING PAPER, 2007/435, D/2006/7012/3X, Ghent University, Ghent, Belgium, Januar 2007, befasst sich mit dem Lernen von Random-Multinomialen Logits (RMNL) an sich.

[0006] Die Szenensegmentierung war ein Gebiet von aktiver Untersuchung und hat einen weiten Anwendungsbereich auf Probleme der realen Welt, wie etwa bei Anwendungen in Roboter- und Kraftfahrzeugsystemen. Ein herkömmliches Szenensegmentierungsverfahren verwendet diskretisierte Repräsentationen, wie etwa Codebooks von Merkmalen oder Texton-Bilder, die ein Modell eines Gesamtbilds oder von spezifischen Regionen des Bilds mit oder ohne räumlichem Kontext des Bilds erstellen. Textone eines eingegebenen Bilds sind diskretisierte Textonwörter, welche erlernt werden durch Anwenden einer Filterbank auf das eingegebene Bild und Clusterbildung der Ausgabe der Filterbank. Das Problem mit diesem Verfahren ist, dass es nur Szenensegmentierung auf Bildebene anspricht. Somit steht man Herausforderungen gegenüber, Objekte, insbesondere kleine Objekte in einem Bild zu detektieren und zu lokalisieren, wo Bildebenen-Merkmale und Statistiken häufig ungenügend sind.

[0007] Ein anderes herkömmliches Szenensegmentierungsverfahren verwendet Textur-Layout Merkmale eines eingegebenen Bilds zum Verstärken von Merkmalsauswahlen, die auf Textone einwirken. Ein Beispiel dieses herkömmlichen Szenensegmentierungsverfahrens verwendet eines semantischen Textonwald sowohl zur Textonerzeugung als auch zur Textonklassifizierung. Da die Anzahl dieser Merkmale sehr groß ist, ist das Training einer für dieses Verfahren verwendete Szenensegmentierungsmaschine sehr langsam, und die Leistungsfähigkeit dieser Szenensegmentierung wird mit zunehmender Größe des Trainingsdatensatzes und Variation in Objektklassen im Trainingsdatensatz schlechter.

[0008] Aufgabe der Erfindung ist es daher, ein Computer implementiertes Verfahren, eine Computer-Vorrichtung und ein Computerprogrammprodukt zum Lernen eines RML-Klassifizierers anzugeben, womit sich auch bei zunehmender Größe eines Trainingsdatensatzes kleine Objekte in einem Bild gut detektieren und lokalisieren lassen.

[0009] Zur Lösung der obigen Aufgabe wird ein Computer implementiertes Verfahren, eine Computer-Vorrichtung und ein Computerprogrammprodukt nach den Ansprüchen 1, 14 und 24 angegeben.

Figurenliste

Fig. 1 stellt ein Computersystem zur semantischen Szenensegmentierung mittels eines Random multinomialen Logit Klassifizierers gemäß einer Ausführung der Erfindung dar.

Fig. 2 ist ein Flussdiagramm von multinomialen logistischen Regressions-Lernmodellen des Random multinomialen Logit Klassifizierers gemäß einer Ausführung der Erfindung.

Fig. 3A ist eine Schwarz- und Weißdarstellung einer semantischen Szenensegmentierung gemäß einer Ausführung der Erfindung.

Fig. 3B ist eine entsprechende Farbdarstellung von **Fig. 3A** der semantischen Szenensegmentierung gemäß einer Ausführung der Erfindung.

Fig. 4 ist ein Blockdiagramm eines Bildtextonisierungsmoduls gemäß einer Ausführung der Erfindung.

Fig. 5A ist eine Schwarz- und Weißdarstellung einer Merkmalsselektion gemäß einer Ausführung der Erfindung.

Fig. 5B ist eine entsprechende Farbdarstellung von **Fig. 5A** der Merkmalsselektion gemäß einer Ausführung der Erfindung.

Fig. 6 ist ein Flussdiagramm der Anwendung von gelernten multinomialen logistischen Regressionsmodellen zur semantischen Szenensegmentierung eines eingegebenen Bilds gemäß einer Ausführung der Erfindung.

Detaillierte Beschreibung der Erfindung

[0010] Nun wird eine Ausführung der vorliegenden Erfindung in Bezug auf die Figuren beschrieben, worin gleiche Bezugszahlen identisch funktionell ähnliche Elemente bezeichnen.

[0011] Der Bezug in der Beschreibung auf „eine Ausführung“ oder auf „Ausführung“ bedeutet, dass ein bestimmtes Merkmal, eine bestimmte Struktur oder eine bestimmte Eigenschaft, die in Verbindung mit den Ausführungen beschrieben ist, in zumindest einer Ausführung der Erfindung enthalten ist. Die Vorkommnisse der Phrase „in einer Ausführung“ an verschiedenen Stellen der Beschreibung beziehen sich nicht notwendiger Weise alle auf die gleiche Ausführung.

[0012] Einige Teile der detaillierten Beschreibung, welche folgt, sind als Algorithmen oder symbolische Repräsentationen von Operationen an Datenbits innerhalb eines Computerspeichers dargestellt. Diese algorithmischen Beschreibungen und Repräsentationen sind die Mittel, die vom Fachkundigen der Datenverarbeitung verwendet werden, um die Substanz ihrer Arbeit anderen Fachkundigen besonders effizient zu übermitteln. Ein Algorithmus versteht sich hier und allgemein als in sich konsistente Sequenz von Schritten (Instruktionen), die zu einem gewünschten Ergebnis führen. Die Schritte sind jene, die physikalische Manipulationen von physikalischen Größen erfordern. Gewöhnlich, gleichwohl nicht notwendig, haben diese Größen die Form von elektrischen, magnetischen oder optischen Signalen, die gespeichert, übertragen, kombiniert, verglichen und anderweitig bearbeitet werden können. Manchmal ist es hauptsächlich aus Gründen der allgemeinen Gebräuchlichkeit angemessen, diese Signale als Bits, Werte, Elemente, Symbole, Schriftzeichen, Terme, Zahlen oder dergleichen zu bezeichnen. Ferner ist manchmal auch ein Bezug auf bestimmte Anordnungen von Schritten angemessen, die physikalische Manipulationen von physikalischen Größen als Module oder Codierungsvorrichtungen erfordern, ohne an Verallgemeinerung zu verlieren.

[0013] Jedoch sind alle diese und ähnliche Terme geeigneten physikalischen Größen zuzuordnen und sind lediglich passende Kennungen (engl. label), die diesen Größen zugeordnet werden. Solange nicht anders ausdrücklich gesagt, wie aus der folgenden Schreibung ersichtlich, versteht es sich, dass in der gesamten Beschreibung Diskussionen zur Verwendung von Begriffen wie etwa „bearbeiten“ oder „berechnen“ oder „errechnen“ oder „bestimmen“ oder „anzeigen“ oder „bestimmen“ oder dergleichen sich auf den Betrieb und Prozesse eines Computersystems beziehen, oder eine ähnliche elektronische Rechenvorrichtung, die Daten manipuliert und überträgt, die als physikalische (elektronische) Größen innerhalb von Speichern oder Registern des Computersystems repräsentiert sind, oder anderen solchen Informationsspeicher-, Übertragungs- oder Anzeigevorrichtungen.

[0014] Verschiedene Aspekte der vorliegenden Erfindung enthalten Prozessschritte und Anweisungen, die hierin in der Form eines Algorithmus beschrieben sind. Es sollte angemerkt werden, dass die Prozessschritte und Anweisungen der vorliegenden Erfindung in Software, Firmware oder Hardware verkörpert sein könnten,

und wenn sie in Software verkörpert sind, könnten sie heruntergeladen werden, um an verschiedenen Plattformen, die durch eine Vielzahl von Betriebssystemen verwendet werden, zu residieren und darauf betrieben zu werden.

[0015] Die vorliegende Erfindung betrifft auch eine Vorrichtung zum Durchführen der Operationen darin. Diese Vorrichtung kann speziell für die erforderlichen Zwecke konstruiert sein oder kann ein Mehrzweckcomputer aufweisen, der durch einen im Computer gespeicherten Computerprogramm selektiv aktiviert oder rekonfiguriert werden. Ein solches Computerprogramm kann in einem Computer lesbaren Speichermedium gespeichert sein, wie etwa, aber nicht beschränkt auf ein beliebigen Plattentyp, einschließlich Floppydisks, optischen Disks, CD-ROMs, magnetoptischen Disks, Nur-Lesespeichern (ROMs), Speichern mit wahlfreiem Zugriff (RAMs), EPROMs, EEPROMs, magnetische oder optische Karten, anwendungsspezifische integrierte Schaltungen (ASICs) oder jeden beliebigen Typ eines Mediums, der zum Speichern von elektronischen Anweisungen geeignet und jeweils mit einem Computersystembus gekoppelt ist. Ferner können die in dieser Beschreibung angesprochenen Computer einen Einzelprozessor enthalten oder können Architekturen sein, die mehrfach Prozessorkonstruktionen verwenden, um die Rechenleistung zu erhöhen.

[0016] Die hierin aufgezeigten Algorithmen und Displays beziehen sich nicht inherent auf jeden bestimmten Computer oder eine andere Vorrichtung. Verschiedene Mehrzwecksysteme können auch mit Programmen gemäß hierin angegebenen Lehren verwendet werden, oder es mag sich als angemessen erweisen, eher spezialisierte Vorrichtung zu konstruieren, um die erforderlichen Verfahrensschritte durchzuführen. Die erforderliche Struktur für eine Vielzahl dieser Systeme wird aus der folgenden Beschreibung ersichtlich. Zusätzlich ist die vorliegende Erfindung nicht in Bezug auf irgendeine bestimmte Programmiersprache beschrieben. Es versteht sich, dass eine Vielzahl von Programmiersprachen dazu verwendet werden können, die hierin beschriebenen Lehren der Erfindung zu implementieren, und nachfolgende Referenzen auf spezifische Sprachen sind vorgesehen, um die Offenbarung und die beste Art der vorliegenden Erfindung zu ermöglichen.

[0017] Zusätzlich ist die in der Beschreibung verwendete Sprache prinzipiell zu Lesbarkeits- und Instruktionzwecken ausgewählt worden, und ist nicht dazu gewählt worden, den erfindungsgemäßen Gegenstand zu skizzieren oder zu umschreiben. Dementsprechend ist die Offenbarung der vorliegenden Erfindung so zu verstehen, dass sie den Umfang der Erfindung, der in den Ansprüchen aufgeführt ist, veranschaulicht, aber nicht beschränkt.

[0018] Ausführungen der Erfindung geben an, einen Random multinomialen Logit (RML) Klassifizierer zur Szenensegmentierung anzulernen und den angelerten RML-Klassifizierer auf ein Ausgangsbild der Szenensegmentierung anzuwenden. Ein Szenensegmentierungssystem unter Verwendung des RML-Klassifizierers lernt den RML-Klassifizierer durch einen vormarkierten (engl. pre-labeled) Bildtrainingssatz an und verbessert die Leistungsfähigkeit des RML-Klassifizierers zur Szenensegmentierung durch zufällig (randomisiert) ausgewählte Textur-Layout-Merkmale. Das Szenensegmentierungssystem vorverarbeitet den Bildtrainingssatz durch Beseitigung von Kontrast- und Helligkeitsänderungen unter den Bildern in dem Bildtrainingssatz, und durch Verfalten (engl. convoluting) und Gruppenbildung (engl. Clustering) der Bilder in dem Bildtrainingssatz. Das System ersetzt die von dem RML-Klassifizierer verwendeten Merkmale durch zufällig ausgewählte Textur-Layout-Merkmale basierend auf einer statistischen Signifikanzmessung, welche den Merkmalen zugeordnet ist. Jedes multinomiale logistische Regressionsmodell des RML-Klassifizierers schätzt die Wahrscheinlichkeitsverteilung eines Texton-Worts eines gewählten Textur-Layout-Merkmals. Das System erzeugt eine letzte Kennung für das Texton-Wort durch Mittelwertbildung der Wahrscheinlichkeitsverteilung jedes multinomialen logistischen Regressionsmodells des RML-Klassifizierers. Der angelernete RML-Klassifizierer kann auf ein Eingangsbild zur Szenensegmentierung angewendet werden. Das offenbarte Szenensegmentierungssystem hat auch einen weiten Anwendungsbereich auf Probleme der realen Welt, und zwar Anwendungen in Roboter- und Kraftfahrzeugsystemen.

[0019] Eine Ausführung eines offenbarten Systems (und Verfahrens) enthält das Anlernen eines Random multinomialen Logit (RML) Klassifizierers und Anwenden des RML-Klassifizierers zur Szenensegmentierung. Das System enthält ein Bildtextonisierungsmodul, ein Merkmalselektionsmodul und einen RML-Klassifizierer. Das Bildtextonisierungsmodul ist konfiguriert, um einen Bildtrainingssatz zu erhalten, wobei die Objekte der Bilder vormarkiert sind. Das Bildtextonisierungsmodul ist ferner konfiguriert, um entsprechende Texton-Bilder aus dem Bildtrainingssatz zu erzeugen. Das Merkmalselektionsmodul ist konfiguriert, um durch Zufall eines oder mehrere Textur-Layout-Merkmale aus den Texton-Bildern auszuwählen. Der RML-Klassifizierer umfasst mehrere multinomiale logistische Regressionsmodelle. Der RML-Klassifizierer ist konfiguriert, um jedes multinomiale logistische Regressionsmodell unter Verwendung der gewählten Textur-Layout-Merkmale anzuler-

nen. Der RML-Klassifizierer ist ferner konfiguriert, um die erlernten Regressionsmodelle auf ein Bild zur Szenensegmentierung anzuwenden.

Systemüberblick

[0020] Fig. 1 zeigt ein Computersystem **100** zur semantischen Szenensegmentierung unter Verwendung eines Random multinomialen Logit (RML) Klassifizierers **126** gemäß einer Ausführung der Erfindung. Das Computersystem **100** umfasst einen Speicher **120**, einen Eingangscontroller **130**, einen Ausgangscontroller **140** und einen Prozessor **150**. In einer Ausführung ist das Computersystem **100** konfiguriert, um einen Bildtrainingssatz **110A** zum Anlernen des RML-Klassifizierers **126** zu empfangen. In einer anderen Ausführung ist das Computersystem **100** konfiguriert, um ein Eingangsbild **110B** zur Szenensegmentierung durch den angelernten RML-Klassifizierer **126** zu empfangen. Der Bildtrainingssatz **110A** umfasst mehrere vormarkierte Bilder. Das Eingangsbild **110B** ist ein beliebiges stehendes Bild oder ein Einzelbild (engl. frame) eines Videos, das von einem Computersystem **100** zur Szenensegmentierung empfangen wird. Das Anlernen des RML-Klassifizierers **126** mittels des Bildtrainingssatzes **110A** ist nachfolgend weiter in Bezug auf Fig. 2 beschrieben. Die Anwendung des RML-Klassifizierers **126** auf das Eingangsbild **110B** ist nachfolgend weiter in Bezug auf Fig. 6 beschrieben.

[0021] In einer Ausführung umfasst der Bildtrainingssatz **110A** Videosequenzen, die durch Anbringen einer Kamera an einem sich bewegendem Fahrzeug erhalten werden, das Motorräder auf Straßen aufnimmt. Die Videos sind ähnlich zu dem, was ein Fahrer im seitlichen Rückspiegel eines sich bewegendem Fahrzeugs sehen würde. Die Videos können sich den Typen der aufgenommenen Motorräder unterscheiden, wobei sie Blickperspektiven und den Störfleckenbetrag und die Beleuchtungsqualität sehen. Z.B. enthält ein Bildtrainingssatz **116A** dreiundsechzig Einzelbilder, die aus sechs Videosequenzen ausgewählt wurden, bei insgesamt angenähert 5.800 Einzelbildern. Die ausgewählten Einzelbilder werden gemäßen dem Typ Objekte, die sie repräsentieren, markiert, wie etwa Fahrrad, Straße, Himmel und andere. Der vormarkierte Bildtrainingssatz **110A** wird von dem Computersystem **100** aufgenommen, um den RML-Klassifizierer **126** anzulernen. Die Lernergebnisse des RML-Klassifizierers **126** werden mit den Markierungen des Bildtrainingssatzes **110A** verglichen, um die Leistungsfähigkeit des RML-Klassifizierers **126** zu evaluieren.

[0022] Der Speicher **120** speichert Daten und/oder Anweisungen, die von dem Prozessor **150** ausgeführt werden können. Die Anweisungen können einen Code zur Durchführung von irgendeinem oder allen der hierin beschriebenen Techniken aufweisen. Der Speicher **120** kann eine DRAM Vorrichtung, ein statischer Speicher mit wahlfreiem Zugriff (SRAM), ein Flash RAM (nichtflüchtiger Speicher), Kombinationen der obigen oder irgendeine andere Speichervorrichtung sein, die in der Technik bekannt ist. In einer Ausführung umfasst der Speicher **120** ein Bildtextonisierungsmodul **122**, ein Merkmalsselektionsmodul **124**, den RML-Klassifizierer **126** und einen Datenspeicher **128**.

[0023] Das Bildtextonisierungsmodul **122** umfasst computerausführbare Anweisungen zum Erzeugen von entsprechenden Texton-Bildern aus den Eingangsbildern. Ein aus einem Eingangsbild erzeugtes Texton-Bild ist ein Bild aus Pixeln, wo jeder Pixelwert in dem Texton-Bild eine Repräsentation seines entsprechenden Pixelwerts in dem Eingangsbild ist. Insbesondere wird jeder Pixelwert des Eingangsbilds durch eine Repräsentierung, z.B. eine Clusteridentifizierung, ersetzt, entsprechend dem Pixelwert des Eingangsbilds, nachdem das Eingangsbild bearbeitet worden ist. Z.B. wird ein Eingangsbild mit einer Filterbank verfalltet, was in 17-dimensionalen Vektoren für jeden Pixel der Eingangsbilder resultiert. Die 17-dimensionalen Vektoren des Eingangsbilds nach Filterung werden geclustert (gruppiert), und jeder der 17-dimensionalen Vektoren wird durch eine Identifizierung des Clusters (die Clusterzahl) repräsentiert, zu der der Vektor gehört. Ein Texton-Bild erhält man in Antwort auf jeden Pixelwert des Eingangsbilds, der durch seine entsprechende Clusteridentifizierung ersetzt wird. Jedes Pixel eines Texton-Bilds ist ein Textonwort, das auch als Texturcluster bezeichnet wird. Das Bildtextonisierungsmodul **122** wird nachfolgend weiter in Bezug auf Fig. 4 beschrieben.

[0024] Das Merkmalsselektionsmodul **124** umfasst computerausführbare Anweisungen zum Ausführen der Zufalls-Textur-Layout-Merkmalsselektion, die nachfolgend in Bezug auf die Fig. 5A und Fig. 5B beschrieben wird. Ein Textur-Layout-Merkmal ist ein ausgewählter Anteil eines Texton-Bilds, z.B. ein gewählter rechteckiger Bereich des Texton-Bilds. Für jedes gewählte Textur-Layout-Merkmal berechnet das Merkmalsselektionsmodul **124** einen Prozentsatz von Pixeln, deren Pixelwert gleich der Clusteridentifizierung eines spezifizierten Texturclusters ist. Der RML-Klassifizierer **126** umfasst coomputerausführbare Anweisungen zum Anlernen und Anwenden von mehreren multinomialen logistischen Regressionsmodellen des RML-Klassifizierers **126**, wie unten in Bezug auf die Fig. 2, Fig. 3A und Fig. 3B beschrieben. In einer Ausführung der

Erfindung speichert der Datenspeicher **128** die Lernergebnisse und/oder semantischen Segmentierungsergebnisse des RML-Klassifizierers **126**.

[0025] **Fig. 2** ist ein Flussdiagramm zum Lernen von multinomialen logistischen Regressionsmodellen des Random multinomialen Logit Klassifizierers **126** gemäß einer Ausführung der Erfindung. Zu Anwendung erhält **210** das Computersystem **100** einen Bildtrainingssatz, der mehrere Bilder erhält, wobei jedes der Bilder entsprechend dem Typ der Objekte, die sie in dem Bild repräsentieren, markiert wird, wie etwa ein Motorrad, eine Straße, Himmel und andere. Das Bildtextonisierungsmodul **122** erzeugt die entsprechenden Texton-Bilder aus den Bildern in dem Trainingssatz. Das Merkmalselektionsmodul **124** wählt **214** zufällig Textur-Layout-Merkmale aus den Texton-Bildern und schickt die gewählten Merkmale zu dem RML-Klassifizierer **126**. Jedes der multinomialen logistischen Regressionsmodelle des RML-Klassifizierers **26** erstellt ein Modell **216** der Wahrscheinlichkeitsverteilung eines gegebenen Texturworts in einem gewählten Merkmal eines Texton-Bilds. Der RML-Klassifizierer **126** evaluiert **218** die Leistungsfähigkeit der multinomialen logistischen Regressionsmodelle. In Antwort auf die zufriedenstellende Leistungsfähigkeit der multinomialen logistischen Regressionsmodelle, beendet **220** der RML-Klassifizierer **126** den Lernprozess; andernfalls wiederholt der RML-Klassifizierer **126** den Lernprozess mit neuerlich gewählten Textur-Layout-Merkmalen aus den Texton-Bildern.

[0026] **Fig. 3A** ist eine Schwarz- und Weißdarstellung einer semantischen Szenensegmentierung gemäß einer Ausführung der Erfindung. Anfänglich wird ein Bildtrainingssatz **202** von dem in **Fig. 1** gezeigten Computersystem **100** empfangen, um die oben in **Fig. 2** beschriebenen multinomialen logistischen Regressionsmodelle anzulernen. Zu Erläuterungszwecken ist in den **Fig. 3A** bis **Fig. 3B** der Bildtrainingssatz, welcher mehrere Bilder enthält, durch ein Einzelbild **302** gezeigt. Ähnlich sind entsprechende Texton-Bilder mit einem Texton-Bild **306** gezeigt, um die markierte Ausgabe (d.h., die szenensegmentierten Bilder entsprechend den Bildern des Bildtrainingssatzes) wird durch ein markiertes Bild **310** repräsentiert. Der Bildtrainingssatz wird vorbearbeitet, gefaltet und geclustert **304**, um Texton-Bilder **306** zu erzeugen, entsprechend den Bildern des Trainingsdatensatzes. Das Merkmalselektionsmodul **124** wählt zufällig mehrere Textur-Layout-Merkmale **308** (z.B. Rechtecke, die die gewählten Textur-Layout-Merkmale repräsentieren) und schickt diese zu dem RML-Klassifizierer **126** zum Lernen der multinomialen logistischen Regressionsmodelle des RML-Klassifizierers **126**. Eine Darstellung in **Fig. 3A** hat der RML-Klassifizierer insgesamt N multinomiale logistische Regressionsmodelle, wobei N ein Konstruktionsparameter ist. Jedes der multinomialen logistischen Regressionsmodelle berechnet die Wahrscheinlichkeitsverteilung eines gegebenen Texturworts in einem gewählten Textur-Layout-Merkmal und erzeugt die markierte Ausgabe **310** der Eingangsbilder **302**. Die Markierungen der Ausgabe bezeichnet das Ergebnis der Szenensegmentierung der Eingangsbilder. **Fig. 3B** ist eine entsprechende Farbdarstellung von **Fig. 3A** der semantischen Szenensegmentierung gemäß einer Ausführung der Erfindung.

RML Bildtextonisierung

[0027] **Fig. 4** ist ein Blockdiagramm des Bildtextonisierungsmoduls **122** gemäß einer Ausführung der Erfindung. In der in **Fig. 4** dargestellten Ausführung enthält das Bildtextonisierungsmodul **122** ein Bildvorverarbeitungsmodul **410**, ein Bildfaltungsmodul **420** und ein Bildclustermodule **430**. Das Bildtextonisierungsmodul **122** ist konfiguriert, um entsprechende Texton-Bilder aus einem Bildtrainingssatz zu erzeugen, durch Vorverarbeitung des Bildtrainingssatzes, Faltung und Clusterbildung des vorbearbeiteten Bildtrainingssatzes.

[0028] Allgemein enthält ein Bildsatz von mehreren Bildern Kontrast- und Helligkeitsvariationen unter den Bildern in dem Bildtrainingssatz. Kontrast- und Helligkeitsvariationen können die Texturrechnung nachteilig beeinflussen. Das Bildvorverarbeitungsmodul **410** ist konfiguriert, um die Kontrast- und Helligkeitsveränderung in dem Bildsatz vor der Texturberechnung zu entfernen. Insbesondere bearbeitet das Bildvorverarbeitungsmodul **410** die Bilder in einem Bildsatz derart, dass die bearbeiteten Bilder Histogramm-angeglichen sind und eine mittlere Standardeinheitsabweichung von Null haben.

[0029] Das Bildfaltungsmodul **420** ist konfiguriert, um den vorbearbeiteten Bildtrainingssatz mit einer Filterbank zu falten. In einer Ausführung verwendet das Bild Faltungsmodul **420** eine 17-dimensionale Filterbank, bestehend aus Gauss-Verteilungen bei Skalen k , $2k$ und $4k$, Ableitungen der Gauss-Verteilungen entlang den x - und y -Achsen bei den Skalen $2k$ und $4k$, und Laplace-Transformationen von Gauss-Verteilungen bei Skalen k , $2k$, $4k$ und $8k$, wobei kein Konstruktionsparameter ist. Die Gauss-Verteilungen werden auf allen drei Kanälen eines CIELab Farbraums berechnet, und der Rest der Filter wird nur auf den Luminanzkanal angewendet.

[0030] Das Bildclusteringsmodul **430** ist konfiguriert, um den gefalteten Bildtrainingssatz zu clustern. In einer Ausführung clustert das Bildclusteringsmodul **430** die von dem gefalteten Bildsatz erzeugten 17-dimensionalen Vektoren unter Verwendung einer hierarchischen k-Mittelclusterungsmethode. Jeder der 17-dimensionalen Vektoren eines Bilds des Bildtrainingssatzes nach Faltung wird durch einen Identifizierer des Clusters (z.B. der Clusterzahl) repräsentiert, zu dem der Vektor gehört. Man erhält ein Texton-Bild in Antwort auf jeden Pixelwert des Bilds, der durch seine entsprechende Clusteridentifizierung ersetzt wird. Jedes Pixel eines Texton-Bilds ist ein Textonwort, das auch als Texturcluster bezeichnet wird. Das Bildclusteringsmodul **430** kann ferner konfiguriert sein, um eine Dreiecks-Ungleichheit zu verwenden, um die k-Mittelclusterung zu beschleunigen. Andere Clusterverfahren, die den Fachkundigen in der Technik bekannt sind, sind für die Ausführungen der Erfindung leicht verfügbar.

Random multinomialer Logit (RML) Klassifizierer

[0031] Ein RML-Klassifizierer besteht aus N multinomialen logistischen Regressionsmodellen, deren jedes Modell die Wahrscheinlichkeitsverteilung der Kennung y erstellt, wobei der Eingangsvektor x wie in der folgenden Gleichung (1) gegeben ist:

$$\pi_{il} = p_t(y = l | x, \beta_l) = \begin{cases} \frac{\exp\left(\beta_{l0} + \sum_{f=1:M} \beta_{lf} \phi_f(x)\right)}{Z}, & l = 1:L-1 \\ 1/Z, & l = L \end{cases} \quad (1)$$

wobei i und l jeweils Indizes in das Modell und den Markierungssatz sind, und Z die Normalisierungskonstante ist, die die Verteilungssumme einheitlich macht. Das $\Phi(\cdot)$ repräsentiert Merkmalsfunktionen, die an dem Eingangsvektor x berechnet sind, und β_{il} ist der Vektor von Längenkoeffizienten (L-1), die die Detektionsfunktion für die Objektkategorie l definieren. Stapelt man jeden dieser Vektoren, erhalten wir die (L-1)x(M+1) Matrix β_l aller Koeffizienten für das multinomiale Regressionsmodell.

[0032] Das Training für den RML Klassifizierer **126** involviert das Lernen der β -Koeffizienten aus dem Bildtrainingssatz. Der Bildtrainingssatz wird mit Austausch zum Erlangen von N kleineren Sätzen abgetastet (gesampelt), womit die einzelnen Regressionsmodelle des RML Klassifizierers **126** gelernt werden. Die Merkmale für die einzelnen Modelle werden auch nach Zufall ausgewählt, M Merkmale pro Modell, wobei M gewöhnlich eine kleine Zahl (z.B. 10 bis 20) ist. Eine Vergrößerung der Anzahl von Merkmalen über dies hinaus resultiert allgemein in einer schlechten Leistungsfähigkeit, da die Ausgangsvarianz mit der Anzahl der verwendeten Merkmale zunimmt. Die letztendliche Ausgangsmarkierungsverteilung des RML wird durch Aufmitteln über die Ausgabe der einzelnen Modelle so berechnet, wie in der folgenden Gleichung (2):

$$\hat{\pi}_l = \sum_{i=1:N} \pi_{il} \quad (2)$$

[0033] Die Koeffizienten β für die einzelnen Regressionsmodelle werden in einem maximalen Wahrscheinlichkeitsrahmenwerk erlernt, was die Minimierung des Fehlers der in Gleichung (1) beschriebenen Verteilung über alle Trainingsdaten beinhaltet. Wenn man die log-Wahrscheinlichkeit betrachtet, wird die zu maximierende Funktion (Verwerfen des Index für die Modellzahl) in Gleichung (3) beschrieben:

$$L(\beta | \{x, y\}) = \sum_{\{x, y\}} \log \pi_y \quad (3)$$

wobei $\{x, y\}$ die Bildtrainingsdaten sind und π_y die in Gleichung (1) definierte Wahrscheinlichkeitsverteilung ist.

[0034] Die in Gleichung (3) beschriebene log-Wahrscheinlichkeit kann durch Gradientendeszenz-Verfahren oder solche zweiter Ordnung optimiert werden, die dem normalen Fachkundigen bekannt sind. Z.B. wird der Gradient von Gleichung (3) für einen spezifischen Koeffizient gegeben als:

$$\frac{\partial L}{\partial \beta_{lf}} = \sum_{\{x, y\}} \phi_f(x) (I(l = y) - \pi_{il}) \quad (4)$$

wobei $I(\cdot)$ die Indikatorfunktion ist, die bis zur Einheitlichkeit führt, wenn ihr Argument wahr ist.

RML Textur-Layout-Merkmalss Selektion

[0035] Die Leistungsfähigkeit des RML-Klassifizierers, der durch die Gleichungen (1) - (4) beschrieben ist, kann durch Textur-Layout-Merkmalss Selektion verbessert werden. Ein Textur-Layout-Merkmal ist ein ausgewählter Abschnitt eines Texton-Bilds, z.B. ein gewählter rechteckiger Bereich des Texton-Bilds. Die RML Merkmalsselektion erfolgt durch Tauschen (engl. swapping) eines Merkmals, das gegenwärtig in dem RML Klassifizierer verwendet wird, mit einem zufällig gewählten Merkmal basierend auf der statistischen Signifikanz des gegenwärtig verwendeten Merkmals, und das gewählte Merkmal wird von dem RML Klassifizierer dazu benutzt, seine mehreren multinomialen logistischen Regressionsmodelle zu erlernen. Wenn die in dem multinomialen logistischen Regressionsmodell verwendeten Merkmale statistisch signifikant sind, wird das Modell durch randomisierte Merkmalsselektion verbessert.

[0036] In einer Ausführung verwendet das Merkmalsselektionsmodul **124** einen einfachen Skalen-unabhängigen Test zur Bestimmung der statistischen Signifikanz eines gewählten Merkmals. Ein Merkmal trägt nicht zu dem in Gleichung (1) beschriebenen Regressionsmodell bei, wenn die Spalten der ihm entsprechenden Koeffizienten alle extrem klein sind. Der Skalenunabhängige Test zur Bestimmung des Merkmals-Beitrags dient zur Feststellung der statistischen Signifikanz der β -Werte, indem diese mit ihrer Standardabweichung verglichen werden. Wenn $|\beta_{if}| \leq 2\sigma_{if}, \forall i \in [1: L-I]$, wobei σ_{if} die entsprechenden Standardabweichungen repräsentiert, fällt das Merkmal Φ_f aus dem Modell heraus, und an dessen Stelle wird ein anderes Merkmal zufällig gewählt. Das Regressionsmodell wird dann mit den gegenwärtigen Koeffizienten als Anfangswerten für die Optimierung neu gelernt. Da das verworfene Merkmal nicht statistisch signifikant ist, ändern sich die Koeffizientenwerte für die anderen Merkmale allgemein nicht signifikant, und das Nachlernen findet effizient statt.

[0037] Die Standardabweichung der Koeffizientenschätzungen kann aus der Hessschen der log-Wahrscheinlichkeitsfunktion wie in der folgenden Gleichung (5) berechnet werden:

$$\frac{\partial^2 L}{\partial \beta_{chl f}^2} = \sum_{\{x, y\}} -\phi_h(x) \phi_f(x) \pi_c (I(c = l) - \pi_l) \quad (5)$$

wobei c, l und h, f jeweils Indizes in die Markierungs- und Merkmalssätze sind. Der Kehrwert der Hessschen ist die Kovarianzmatrix von β , woraus man die Standardabweichungen erhalten kann. Wenn alle Merkmale in einem multinomialen logistischen Regressionsmodell statistisch signifikant sind, wird das Modell durch die randomisierte Merkmalsauswahl basierend auf der maximalen Wahrscheinlichkeit verbessert. Die Größe $-2 \log L$, wobei L die log-Wahrscheinlichkeit des Modells ist, folgt der chi-Quadrat Statistik und ist für das am besten passende Modell am kleinsten. Daher verbleibt für zwei Modelle, die sich durch ein Einzelmerkmal unterscheiden, das Modell mit der geringeren Statistik.

[0038] Um die Textur-Layout-Merkmalss Selektion weiter zu veranschaulichen, folgt ein Pseudocode eines Beispiels mit RML-Merkmalss Selektion.

Algorithm 1 RML Feature Selection

Input: Current multinomial logistic regression model $ML = \{\phi_{1:M}, \beta\}$, coefficient standard deviations σ , rounds of feature selection S

$LL \leftarrow$ log-likelihood of ML

for $i = 1$ to S do

$B \leftarrow$ set of features $\phi_{\{f\}}$ s.t. $|\beta_{lf}| \leq 2\sigma_{lf} \forall l \in 1:L-1$

 if B is empty then

 Swap a randomly selected feature from current model for a randomly selected new feature ϕ'

 else

 Swap a randomly selected feature from B for a randomly selected new feature ϕ'

 end if

$ML' \leftarrow$ new model learnt using maximum likelihood

$LL' \leftarrow$ log-likelihood of ML'

 if $LL' > LL$ then

$ML \leftarrow ML', LL \leftarrow LL'$

 end if

end for

return ML

(in Übersetzung):

Algorithmus 1 RML Merkmalsselektion

Eingabe: gegenwärtiges multinomiales logistisches Regressionsmodell

$ML = \{\phi_{1:M}, \beta\}$, Koeffizienten-Standardabweichungen σ , Rundungen der Merkmalsselektion S

$LL \leftarrow$ log-Wahrscheinlichkeit von ML

für $i = 1$ bis S mache

$B \leftarrow$ Satz von Merkmalen

wenn B leer ist, dann

 Tausche ein zufällig gewähltes Merkmal aus gegenwärtigem Modell gegen ein zufällig gewähltes neues Merkmal ϕ'

sonst

 Tausche ein zufällig gewähltes Merkmal von B gegen ein zufällig gewähltes neues Merkmal ϕ'

Ende wenn

$ML' \leftarrow$ neues Modell, das anhand maximaler Wahrscheinlichkeit erlernt ist

$LL' \leftarrow$ log-Wahrscheinlichkeit von ML'

wenn $LL' > LL$, **dann**

$ML \leftarrow ML', LL \leftarrow LL'$

Ende wenn

Ende für

zurück ML

[0039] In einer Ausführung bestehen die Textur-Layout-Merkmale aus einem Rechteck r und einem Textonwort t . Ein Textur-Wort wird oben auch als Textur-Cluster bezeichnet. Für jedes Pixel p berechnet das Merkmalsselektionsmodul **124** den Anteil des Textonworts t innerhalb des Rechtecks r , wobei r in das Koordinatensystem mit p als dem Ursprung verschoben worden ist. Zum Beispiel berechnet, für jedes gewählte Textur-Layout-Merkmal, das Merkmalsselektionsmodul **124** einen Prozentsatz von Pixeln innerhalb des Rechtecks r , dessen Pixelwert gleich der Clusteridentifikation des Textonworts t ist. Man kann sagen, dass die Textur-Layout-Merkmale den örtlichen Textur-Kontext in dem Bild erfassen, zum Beispiel die Beziehung, dass ein Boot normalerweise von Wasser umgeben ist. Zusätzlich ist diese kontextuale Beziehung, ausgedrückt als lineare Kombination von mehreren Textur-Layout-Merkmalswerten ausreichend, um die pixelweise Szenenkenennung (engl. labeling) durchzuführen.

[0040] Um den RML-Klassifizierer basierend auf den Textur-Layout-Merkmalen zu lernen, wählt das Merkmalsselektionsmodul **124** zuerst N rechteckige Regionen nach Zufall vor, so dass die Gesamtanzahl der möglichen Merkmale $N_r \times N_t$ ist, wobei N_t die Anzahl der Textonworte in dem Codebuch ist, das in dem Lernprozess verwendet wird. Anschließend werden, für jedes multinomiale Regressionsmodul im RML-Klassifizierer, ein Satz von $M \ll N_r \times N_t$ nach Zufall gewählt, um die in Gleichung (1) angegebenen Verteilungen zu erzeugen.

[0041] Das RML wird unter Verwendung von pixelweise markierten Daten überwacht angelern. Die Merkmalswerte, die an einem Pixel zusammen mit dessen Kennung evaluiert sind, stellen eine Trainingsinstanz dar. Zufällig gewählte Teilsätze dieser Trainingsdaten werden benutzt, um die einzelnen Regressionsmodelle anzulernen. Während der Laufzeit werden die Merkmale in den Regressionsmodellen an einem Eingangsbild evaluiert und durch laufende Regressionsmodelle, um die Ausgangskennung in Gleichung (2) zu bekommen.

[0042] Nun in Bezug auf **Fig. 5A**, ist **Fig. 5A** eine Schwarz- und Weißdarstellung einer Merkmalsselektion gemäß einer Ausführung der Erfindung. Bild **510a** ist eines der Bilder eines Bildtrainingssatzes, das in dem RML-Klassifizierer-Lernprozess verwendet wird, und das Bild **510b** ist das entsprechende Texton-Bild des Bilds **510a**. Die Koeffizienten β **512**, die in Gleichung (1) für die einzelnen Regressionsmodelle des RML-Klassifizierers definiert sind, werden in einem maximalen Wahrscheinlichkeitsrahmenwerk erlernt. Im in **Fig. 5A** dargestellten Beispiel werden insgesamt vier Textur-Layout-Merkmale (**520a**, **530a**, **540a** und **550a**) nach Zufall ausgewählt. Jedes gewählte Textur-Layout-Merkmal besteht aus einer rechteckigen Form, die ein Teil des Texton-Bilds **510b** ist, und einem Textonwort (z.B. **520b**, **530b**, **540b** und **550b**). Für jedes gewählte Textur-Layout-Merkmal berechnet das Merkmalsselektionsmodul **124** einen Prozentsatz von Pixeln innerhalb des Rechtecks des Textur-Layout-Merkmals, dessen Pixelwerte gleich dem Wert des Textonworts ist, das dem gewählten Merkmal zugeordnet ist. Wenn das Textonwort durch die Clusteridentifizierung nach Textonisierung repräsentiert wird, berechnet das Merkmalsselektionsmodul **124** ein Verhältnis der Anzahl von Pixeln in dem Rechteck eines gewählten Merkmals, dessen Pixelwert gleich der Clusteridentifikation ist, und der Gesamtzahl der Pixel innerhalb des Rechtecks.

[0043] **Fig. 5B** ist eine entsprechende Farbdarstellung von **Fig. 5A** der Merkmalsselektion gemäß einer Ausführung der Erfindung. Jedes colorierte Rechteck repräsentiert ein Textur-Layout-Merkmal über das rechteckige Fenster und das Textonwort (das zur Darstellung mit der Farbe angegeben ist), das dem Textur-Layout-Merkmal zugeordnet ist. Merkmale sind zur Markierung von Pixeln in einer Szene statistisch insignifikant, wenn ihre Werte keine Indizes der Pixelmarkierung sind. Ein Textur-Layout-Merkmal, das statistisch insignifikant ist, kann sein, weil die Form des Merkmals zu groß oder zu klein ist. Wenn man **Fig. 5** als Beispiel nimmt, ist das vierte Merkmale für den Koeffizienten β_4 in dem zu lernenden multinomialen Regressionsmodell statistisch insignifikant und wird durch die Merkmalsselektion verworfen.

RML-Klassifizieranwendung für Szenensegmentierung

[0044] Der oben beschriebene angelernete RML-Klassifizierer **126** kann zur Textur basierenden Szenensegmentierung mit verbesserter Genauigkeit und Leistungsfähigkeit verwendet werden, im Vergleich zu herkömmlichen Szenensegmentierungsverfahren. **Fig. 6** ist ein Flussdiagramm der Anwendung des gelernten RML-Klassifizierers **126** auf ein Eingangsbild zur semantischen Szenensegmentierung gemäß einer Ausführung der Erfindung. Anfänglich empfängt **610** das Computersystem **100**, das einen angelerneten RML-Klassifizierer **126** hat, ein Eingangsbild zur Szenensegmentierung. Das Bildtextonisierungsmodul **122** erzeugt **612** ein entsprechendes Texton-Bild unter Verwendung der Cluster, die während des RML-Klassifizierer-Lernprozesses erlernt wurden. Der RML-Klassifizierer **126** wendet **614** die erlernten multinomial logistischen Regressionsmodelle auf das Eingangsbild an. Zum Beispiel berechnet der RML-Klassifizierer **126** die Wahrscheinlichkeitsverteilung einer Pixelkennung, von der durch jedes multinomial logistischer Regressionsmodell

des RML-Klassifizierers **126** mittels der Gleichung (1) ein Modell erstellt wurde. Der RML-Klassifizierer **126** erzeugt **616** die semantische Bildsegmentierung für das Eingangsbild, wobei die semantische Bildsegmentierung durch Segmentierungsmarkierung repräsentiert wird. Zum Beispiel berechnet der RML-Klassifizierer **126** die letztendliche Ausgangskennungsverteilung des Eingangsbilds durch Aufmittlung über die Ausgabe der einzelnen multinomialen logistischen Regressionsmodelle des RML-Klassifizierers **126** mittels der Gleichung (2). Schließlich gibt **618** das Computersystem **100** die semantische Bildsegmentierung des Bilds aus.

[0045] Um Straßenszenen zur Verwendung mit intelligenten Transportsystemen und anderen Anwendungen zu klassifizieren, wird ein RML-Klassifizierer über einen Bildtrainingssatz mit vormarkierten Bildern angeleitet. Der RML-Klassifizierer besteht aus mehreren multinomialen logistischen Regressionsmodellen. Jedes der multinomialen logistischen Regressionsmodelle arbeitet an einem zufällig gewählten Teilsatz von Merkmalen aus dem Bildtrainingssatz und gibt eine Wahrscheinlichkeitsverteilung der Marke des Pixels aus, das den gewählten Merkmalen entspricht. Die Verwendung eines maximalen Wahrscheinlichkeitsrahmenwerks erlaubt, dass die multinomiale logistische Regression in größeren Merkmalsräumen mit verbesserter Leistungsfähigkeit arbeitet. Der angeleitete RML-Klassifizierer kann effizient an einem Eingangsbild zur Szenensegmentierung verwendet werden. Das offenbarte RML-basierte Szenensegmentierungssystem hat einen weiten Anwendungsbereich auf Probleme der realen Welt, wie etwa Anwendungen in Roboter- und Kraftfahrzeugsystemen.

Patentansprüche

1. Computer implementiertes Verfahren zum Lernen eines Random multinomialen Logit (RML) Klassifizierers (126) zur Szenensegmentierung, wobei das Verfahren umfasst:

Empfangen (210) eines Bildtrainingssatzes, worin der Bildtrainingssatz mehrere digitale Repräsentationen von Bildern (302) aufweist, und ein Objekt eines Bilds (302) des Bildtrainingssatzes eine semantische Kennung hat;

Erzeugen (212) einer Mehrzahl von Texton-Bildern (306) entsprechend den Bildern (302) in dem Bildtrainingssatz, worin ein Texton-Bild (306) seines entsprechenden Bilds (302) in dem Bildtrainingssatz ein Bild von Pixeln ist, und wobei jeder Pixelwert in einem Texton-Bild (306) durch eine Repräsentation des Pixelwerts seines entsprechenden Bilds (302) in dem Bildtrainingssatz ersetzt wird;

Wählen (214) von einem oder mehreren Textur-Layout-Merkmalen (308) aus der Mehrzahl der Texton-Bilder (306), wobei das Wählen von einem oder mehreren Textur-Layout-Merkmalen (308) umfasst: Tauschen eines Merkmals, das von dem RML-Klassifizierer (126) gegenwärtig verwendet wird, gegen ein zufällig gewähltes neues Merkmal basierend auf der statistischen Signifikanz des gegenwärtig verwendeten Merkmals;

Lernen (216) von mehreren multinomial logistischen Regressionsmodellen des RML-Klassifizierers (126) basierend auf den gewählten Textur-Layout-Merkmalen; und

Evaluieren (218) der Leistungsfähigkeit der mehreren multinomialen logistischen Regressionsmodelle basierend auf den semantischen Kennungen der Objekte in dem Bildtrainingssatz.

2. Verfahren von Anspruch 1, wobei das Erzeugen (212) der Mehrzahl von Texton-Bildern (306) entsprechend den Bildern (302) in dem Bildtrainingssatz umfasst:

Beseitigen von Kontrast- und Helligkeitsvariationen unter den Bildern (302) in dem Bildtrainingssatz;

Falten der Bilder (302) im Bildtrainingssatz mit einer Filterbank zum Erzeugen von gefalteten Bildern; und

Gruppieren (304) der gefalteten Bilder.

3. Verfahren von Anspruch 2, wobei die Filterbank eine 17-dimensionale Filterbank ist, bestehend aus Gauss-Verteilungen mit verschiedenen Skalen, Ableitungen von Gauss-Verteilungen mit verschiedenen Skalen und Laplace-Transformationen von Gauss-Verteilungen mit verschiedenen Skalen.

4. Verfahren von Anspruch 2, wobei das Gruppieren (304) der gefalteten Bilder umfasst, die gefalteten Bilder unter Verwendung eines hierarchischen k-Mittel-Clusterungsverfahrens (304) zu gruppieren.

5. Verfahren von Anspruch 4, wobei jeder von 17-dimensionalen Vektoren eines gefalteten Bilds durch eine Identifizierung des Clusters repräsentiert wird, zu dem der Vektor gehört.

6. Verfahren von Anspruch 1, wobei die statische Signifikanz eines Merkmals durch einen Vergleich der Koeffizientenwerte, die dem Merkmal zugeordnet sind, und der Standardabweichung der Koeffizientenwerte bestimmt wird.

7. Verfahren von Anspruch 1, wobei ein gewähltes Textur-Layout-Merkmal (308) aus einem Texton-Bild (306) einen rechteckigen Bereich des Texton-Bilds (306) und ein Textonwort aufweist.

8. Verfahren von Anspruch 7, das ferner umfasst: Berechnen eines Prozentsatzes von Pixeln innerhalb des rechteckigen Bereichs, dessen Pixelwert gleich dem Wert des Textonworts ist, über die Gesamtzahl von Pixeln innerhalb des rechteckigen Bereichs.

9. Verfahren von Anspruch 8, wobei der Wert des Textonworts gleich der Clusteridentifikation des Textonworts ist.

10. Verfahren von Anspruch 1, wobei das Lernen (216) von mehreren multinomialen logistischen Regressionsmodellen des RML-Klassifizierers umfasst:

für jedes der multinomialen logistischen Regressionsmodelle des RML-Klassifizierers:

Schätzen der Wahrscheinlichkeitsverteilung eines Textonworts eines gewählten Textur-Layout-Merkmals;

Minimieren des Fehlers der geschätzten Wahrscheinlichkeitsverteilung des Textonworts; und

Optimieren der geschätzten Wahrscheinlichkeitsverteilung des Textonworts.

11. Verfahren von Anspruch 10, wobei das Lernen (216) von mehreren multinomialen logistischen Regressionsmodellen des RML-Klassifizierers (126) ferner umfasst: Aufmitteln der Wahrscheinlichkeitsverteilung des Textonworts von jedem der mehreren multinomialen logistischen Regressionsmodelle zum Erzeugen einer letztendlichen Ausgabekennung für das Textonwort.

12. Verfahren von Anspruch 1, welches ferner umfasst: Nachlernen eines multinomialen logistischen Regressionsmodells des RML-Klassifizierers (126) in Antwort darauf, dass die Leistungsfähigkeit des multinomialen logistischen Regressionsmodells unterhalb eines Leistungsfähigkeitsschwellenwerts liegt.

13. Verfahren von Anspruch 1, welches ferner umfasst: Anwenden des RML-Klassifizierers (126) auf ein Eingangsbild (302) zur Szenensegmentierung.

14. Computer-Vorrichtung für ein Verfahren zum Lernen eines Random multinomialen Logit (RML) Klassifizierers (126) zur Szenensegmentierung, wobei die Vorrichtung umfasst:

ein Bildtextonisierungsmodul (122), konfiguriert zum:

Empfangen (210) eines Bildtrainingssatzes, wobei der Bildtrainingssatz mehrere digitale Repräsentationen von Bildern (302) aufweist und ein Objekt eines Bilds (302) des Bildtrainingssatzes eine semantische Kennung hat;

Erzeugen (212) einer Mehrzahl von Texton-Bildern (306) entsprechend den Bildern (302) in dem Bildtrainingssatz, wobei ein Texton-Bild (306) seines entsprechenden Bilds (302) in dem Bildtrainingssatz ein Bild von Pixeln ist, und wobei jeder Pixelwert in einem Texton-Bild (306) durch eine Repräsentation des Pixelwerts seines entsprechenden Bilds (302) in dem Bildtrainingssatz ersetzt wird; und

ein Merkmalsselektionsmodul (124), konfiguriert zum Wählen (214) von einem oder mehreren Textur-Layout-Merkmalen (308) aus der Mehrzahl der Texton-Bilder (306), wobei das Wählen (214) von einem oder mehreren Textur-Layout-Merkmalen (208) umfasst: Tauschen eines Merkmals, das von dem RML-Klassifizierer (126) gegenwärtig verwendet wird, gegen ein zufällig gewähltes neues Merkmal basierend auf der statistischen Signifikanz des gegenwärtig verwendeten Merkmals; wobei

der RML-Klassifizierer (126) konfiguriert ist zum:

Lernen (216) von mehreren multinomialen logistischen Regressionsmodellen des RML-Klassifizierers (126) basierend auf den gewählten Textur-Layout-Merkmalen; und

Evaluieren (218) der Leistungsfähigkeit der mehreren multinomialen logistischen Regressionsmodelle basierend auf den semantischen Kennungen der Objekte in dem Bildtrainingssatz.

15. Vorrichtung von Anspruch 14, wobei das Bildtextonisierungsmodul (122) ferner konfiguriert ist zum: Beseitigen von Kontrast- und Helligkeitsvariationen unter den Bildern in dem Bildtrainingssatz; Falten der Bilder (302) im Bildtrainingssatz mit einer Filterbank zum Erzeugen von gefalteten Bildern; und Gruppieren (304) der gefalteten Bilder.

16. Vorrichtung von Anspruch 15, wobei die Filterbank eine 17-dimensionale Filterbank ist, bestehend aus Gauss-Verteilungen mit verschiedenen Skalen, Ableitungen von Gauss-Verteilungen mit verschiedenen Skalen und Laplace-Transformationen von Gauss-Verteilungen mit verschiedenen Skalen.

17. Vorrichtung von Anspruch 16, wobei jeder von 17-dimensionalen Vektoren eines gefalteten Bilds durch eine Identifizierung des Clusters repräsentiert wird, zu dem der Vektor gehört.

18. Vorrichtung von Anspruch 14, wobei ein gewähltes Textur-Layout-Merkmal (308) aus einem Texton--Bild (306) einen rechteckigen Bereich des Texton-Bilds (306) und ein Textonwort aufweist.

19. Vorrichtung von Anspruch 18, wobei das Merkmalsselektionsmodul (124) ferner konfiguriert ist zum Berechnen eines Prozentsatzes von Pixeln innerhalb des rechteckigen Bereichs, dessen Pixelwert gleich dem Wert des Textonworts ist, über die Gesamtzahl von Pixeln innerhalb des rechteckigen Bereichs.

20. Vorrichtung von Anspruch 14, wobei der RML-Klassifizierer (126) ferner konfiguriert ist zum:
Konfigurieren von jedem der multinomialen logistischen Regressionsmodelle des RML-Klassifizierers zum:
Schätzen der Wahrscheinlichkeitsverteilung eines Textonworts eines gewählten Textur-Layout-Merkmals;
Minimieren des Fehlers der geschätzten Wahrscheinlichkeitsverteilung des Textonworts; und
Optimieren der geschätzten Wahrscheinlichkeitsverteilung des Textonworts.

21. Vorrichtung von Anspruch 20, wobei der RML-Klassifizierer (126) ferner konfiguriert ist zum Aufmitteln der Wahrscheinlichkeitsverteilung des Textonworts von jedem der mehreren multinomialen logistischen Regressionsmodelle zum Erzeugen einer letztendlichen Ausgabekennung für das Textonwort.

22. Vorrichtung von Anspruch 14, wobei der RML-Klassifizierer (126) ferner konfiguriert ist zum Nachlernen eines multinomialen logistischen Regressionsmodells des RML-Klassifizierers (126) in Antwort darauf, dass die Leistungsfähigkeit des multinomialen logistischen Regressionsmodells unterhalb eines Leistungsfähigkeitsschwellenwerts liegt.

23. Vorrichtung von Anspruch 14, wobei der RML-Klassifizierer (126) ferner konfiguriert ist zum Anwenden des RML-Klassifizierers (126) auf ein Eingangsbild (302) zur Szenensegmentierung.

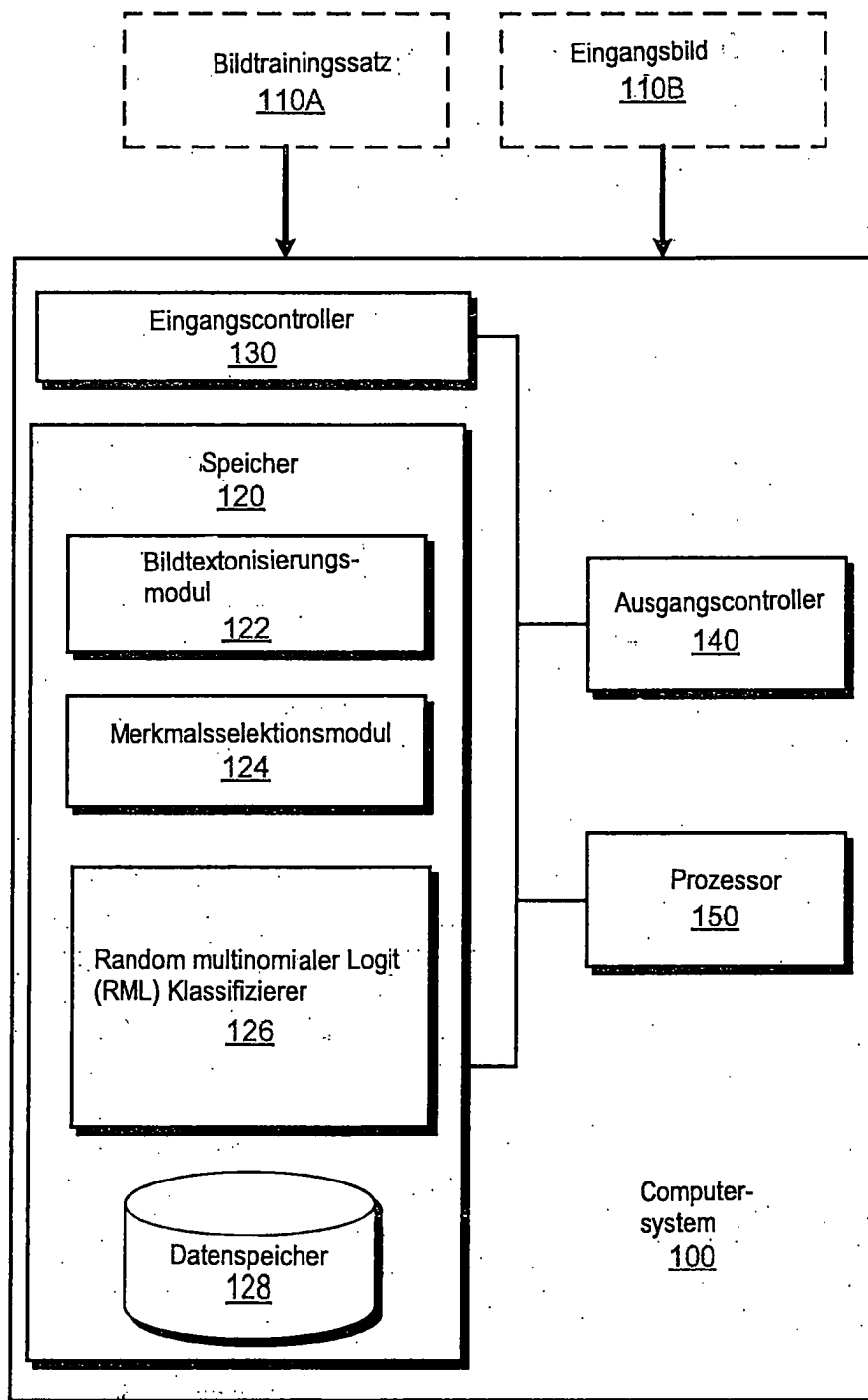
24. Computerprogrammprodukt zum Lernen eines Random multinomial Logit (RML) Klassifizierers (126) zur Szenensegmentierung, wobei das Computerprogrammprodukt ein computerlesbares Medium aufweist, das einen Computerprogrammcode zur Durchführung der Operationen enthält:
Empfangen (210) eines Bildtrainingssatzes, worin der Bildtrainingssatz mehrere digitale Repräsentationen von Bildern (302) aufweist, und ein Objekt eines Bilds (302) des Bildtrainingssatzes eine semantische Kennung hat;
Erzeugen (212) einer Mehrzahl von Texton-Bildern (306) entsprechend den Bildern (302) in dem Bildtrainingssatz, worin ein Texton-Bild (306) seines entsprechenden Bilds (302) in dem Bildtrainingssatz ein Bild von Pixeln ist, und wobei jeder Pixelwert in einem Texton-Bild (306) durch eine Repräsentation des Pixelwerts seines entsprechenden Bilds (302) in dem Bildtrainingssatz ersetzt wird;
Wählen (214) von einem oder mehreren Textur-Layout-Merkmalen (308) aus der Mehrzahl der Texton-Bilder (306), wobei das Wählen (214) von einem oder mehreren Textur-Layout-Merkmalen umfasst: Tauschen eines Merkmals, das von dem RML-Klassifizierer (126) gegenwärtig verwendet wird, gegen ein zufällig gewähltes neues Merkmal basierend auf der statistischen Signifikanz des gegenwärtig verwendeten Merkmals;
Lernen (216) von mehreren multinomial logistischen Regressionsmodellen des RML-Klassifizierers (126) basierend auf den gewählten Textur-Layout-Merkmalen; und
Evaluieren (218) der Leistungsfähigkeit der mehreren multinomialen logistischen Regressionsmodelle basierend auf den semantischen Kennungen der Objekte in dem Bildtrainingssatz.

25. Computerprogrammprodukt von Anspruch 24, wobei der Computerprogrammcode zum Lernen von mehreren multinomialen logistischen Regressionsmodellen des RML-Klassifizierers (126) einen Computerprogrammcode umfasst für:
jedes der multinomialen logistischen Regressionsmodelle des RML-Klassifizierers:
Schätzen der Wahrscheinlichkeitsverteilung eines Textonworts eines gewählten Textur-Layout-Merkmals;
Minimieren des Fehlers der geschätzten Wahrscheinlichkeitsverteilung des Textonworts; und
Optimieren der geschätzten Wahrscheinlichkeitsverteilung des Textonworts.

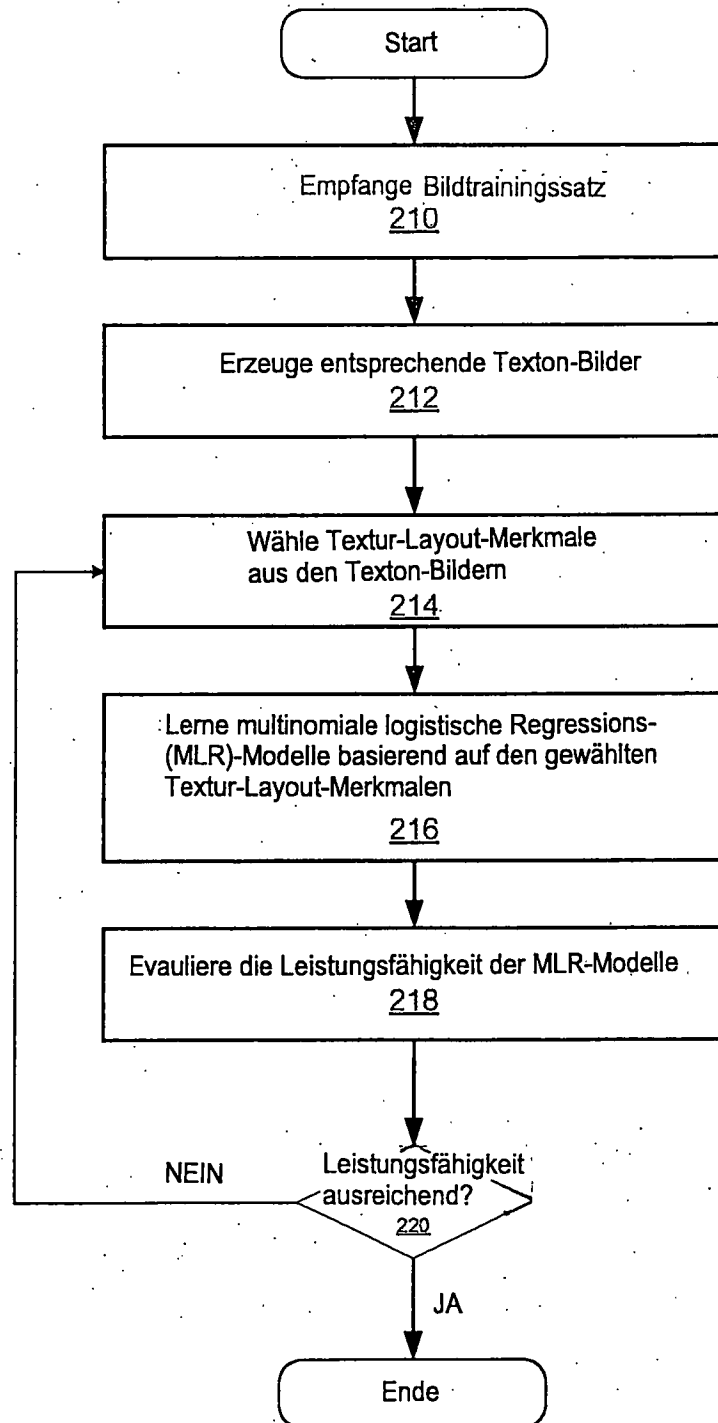
26. Computerprogrammprodukt von Anspruch 24, das ferner einen Computerprogrammcode aufweist zum Anwenden des RML-Klassifizierers (126) auf ein Eingangsbild (302) zur Szenensegmentierung.

Es folgen 8 Seiten Zeichnungen

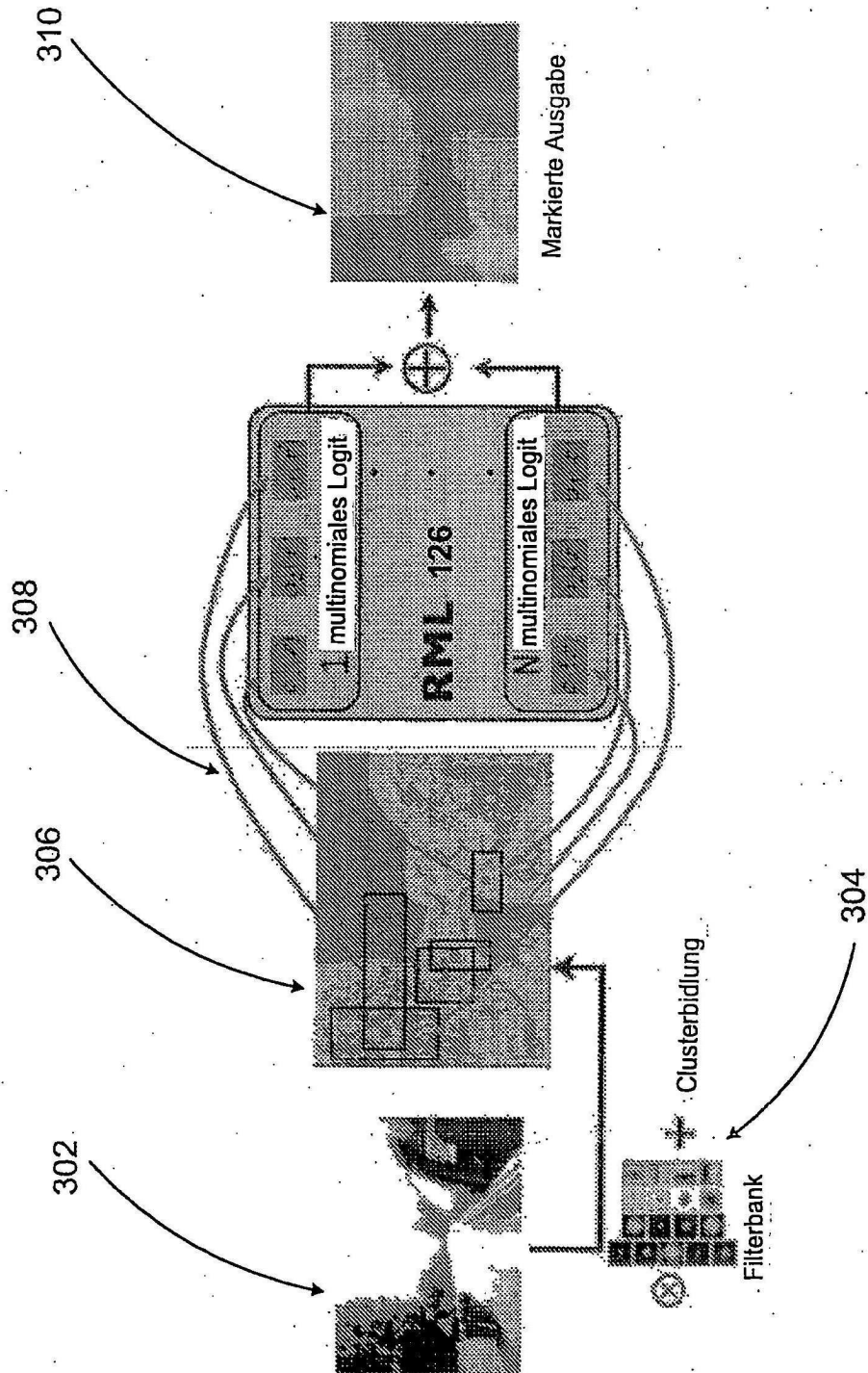
Anhängende Zeichnungen



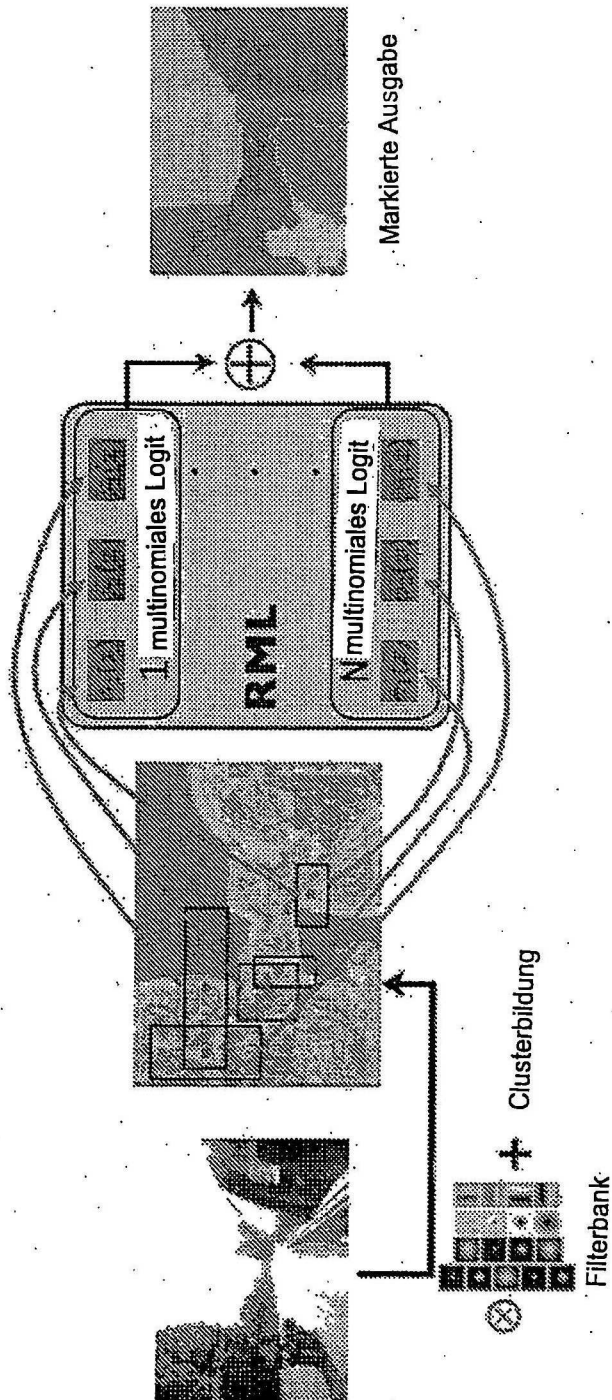
Figur 1



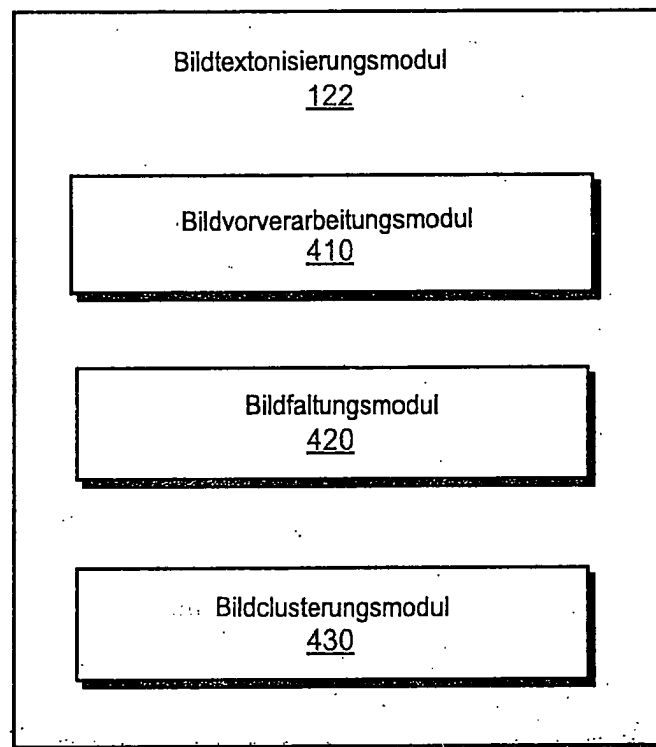
Figur 2



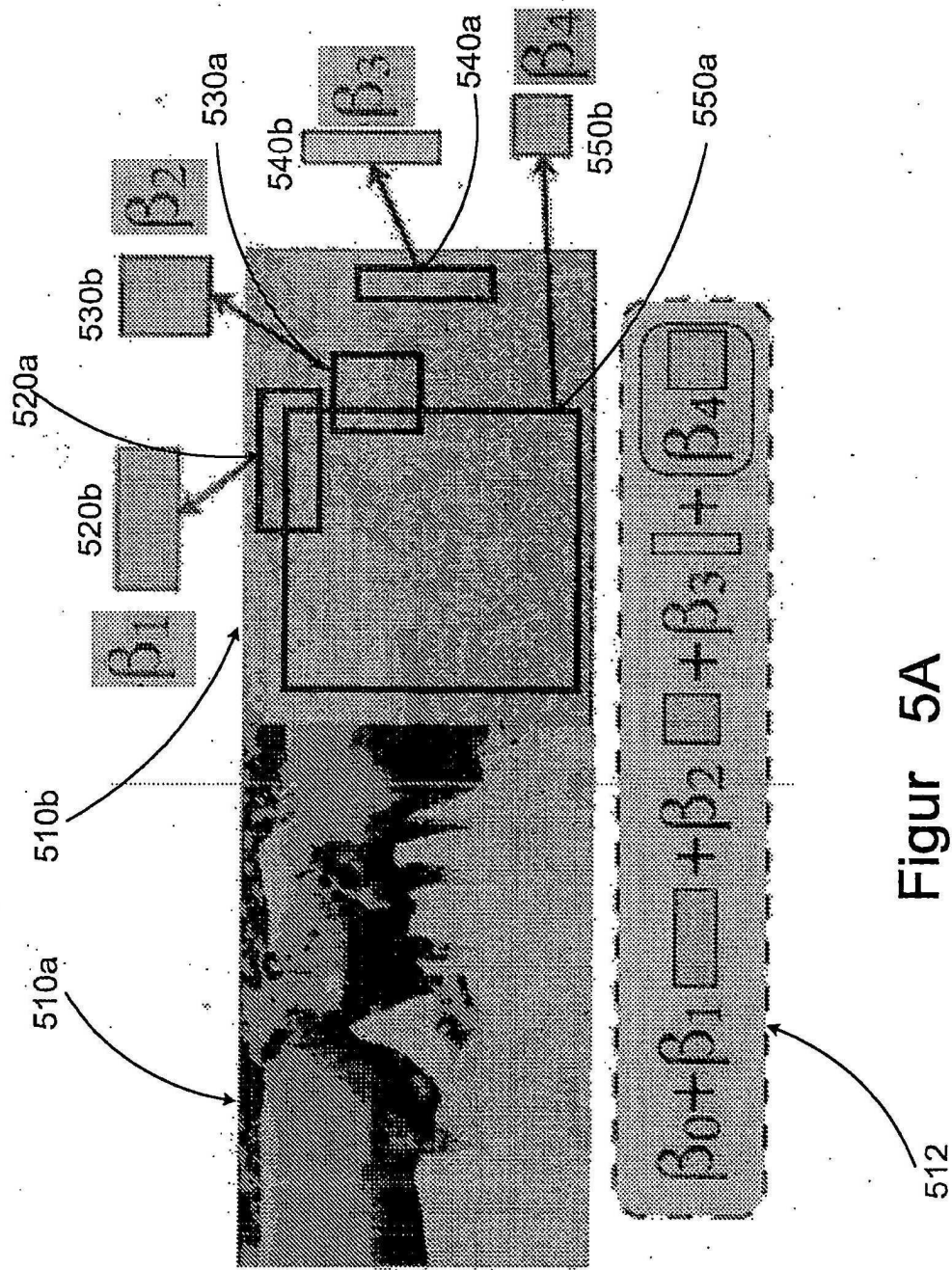
Figur 3A



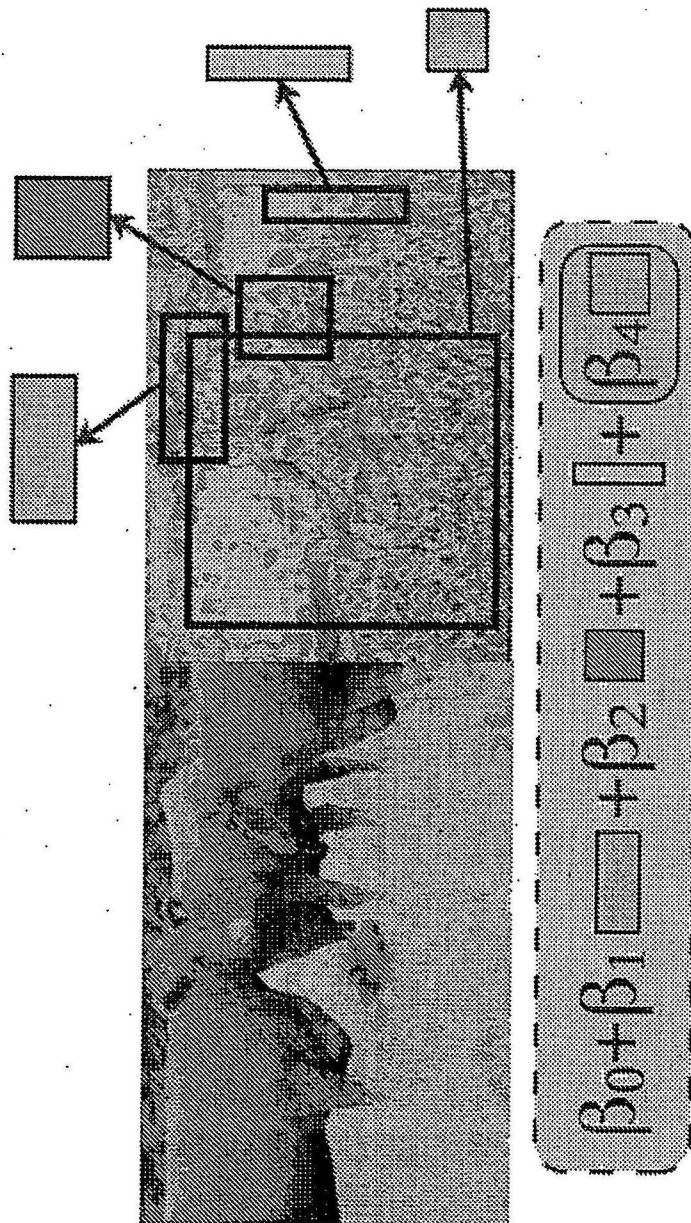
Figur 3B



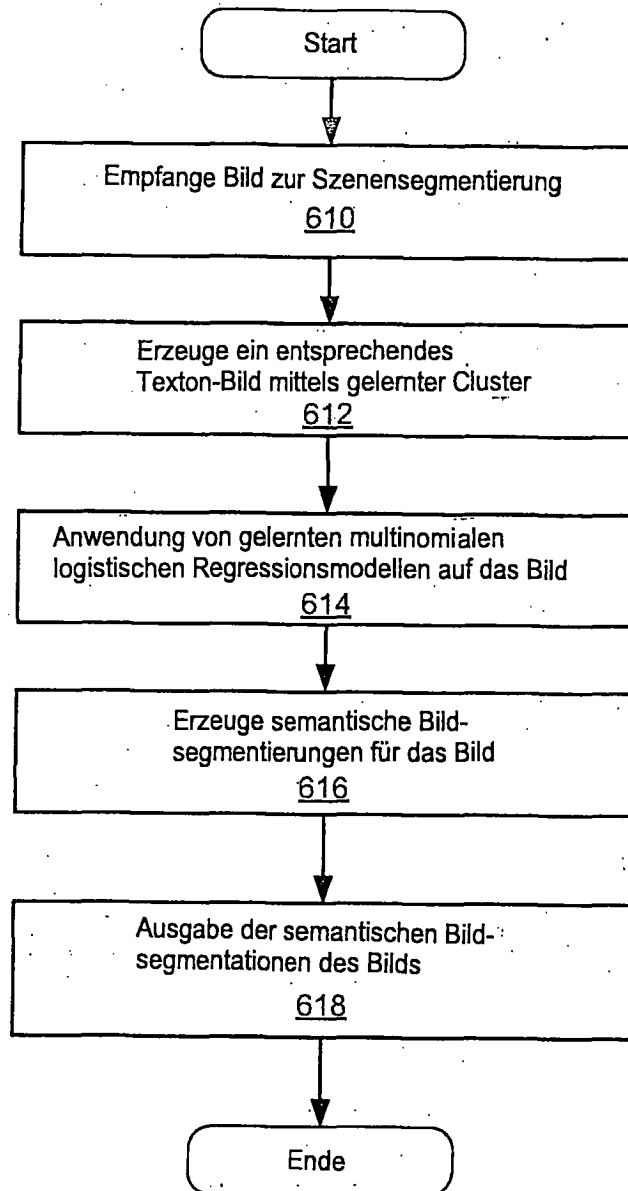
Figur 4



Figur 5A



Figur 5B



Figur 6