



(10) **DE 10 2013 210 153 B4** 2021.05.06

(12)

## Patentschrift

(21) Aktenzeichen: **10 2013 210 153.8**  
(22) Anmeldetag: **31.05.2013**  
(43) Offenlegungstag: **20.02.2014**  
(45) Veröffentlichungstag  
der Patenterteilung: **06.05.2021**

(51) Int Cl.: **G06T 15/00** (2006.01)  
**G06T 7/593** (2017.01)  
**G06K 9/60** (2006.01)  
**H04N 13/239** (2018.01)  
**H04N 13/271** (2018.01)

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:  
**13/490,991 07.06.2012 US**

(73) Patentinhaber:  
**Nvidia Corporation, Santa Clara, Calif., US**

(74) Vertreter:  
**Kraus & Weisert Patentanwälte PartGmbB, 80539 München, DE**

(72) Erfinder:  
**Pulli, Kari, Palo Alto, Calif., US; Troccoli, Alejandro, San Jose, Calif., US; Zhou, Changyin, Santa Ana, Calif., US**

(56) Ermittelter Stand der Technik:

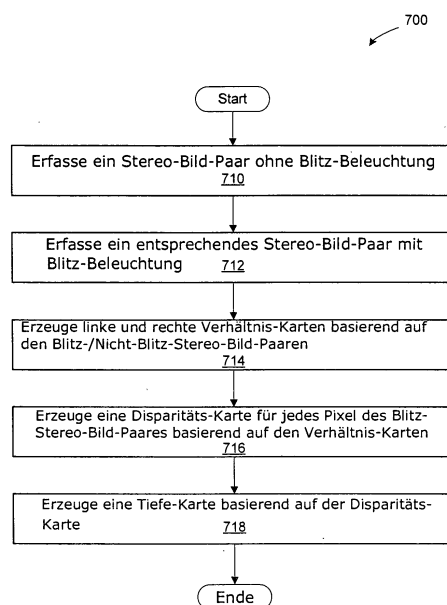
|    |                  |    |
|----|------------------|----|
| US | 2008 / 0 297 621 | A1 |
| US | 2009 / 0 324 059 | A1 |
| US | 2010 / 0 020 078 | A1 |
| US | 2011 / 0 044 531 | A1 |
| US | 2012 / 0 039 525 | A1 |

**FERIS, Rogerio [et al.]: Discontinuity preserving stereo with small baseline multi-flash illumination. In: Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1. IEEE, 2005. S. 412-419**

**SCHARSTEIN, Daniel; SZELISKI, Richard: A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. International journal of computer vision, 2002, 47. Jg., Nr. 1-3, S. 7-42**

(54) Bezeichnung: **Techniken zum Erzeugen von robusten Stereo-Bildern**

(57) Zusammenfassung: Techniken zum Erzeugen von robusten Tiefe-Karten von Stereo-Bildern sind beschrieben. Eine robuste Tiefe-Karte ist von einem Satz von Stereo-Bildern erzeugt, welche mit und ohne Blitz-Beleuchtung erfasst sind. Die Tiefe-Karte ist robuster als Tiefe-Karten, welche unter Benutzung von herkömmlichen Techniken erzeugt sind, weil ein Pixel-Anpassungs-Algorithmus implementiert ist, welcher Pixel in einem Anpassungs-Fenster gemäß des Verhältnisses von Licht-Intensität gewichtet, welche unter Benutzung von verschiedenen Blitz-Beleuchtungs-Pegeln erfasst ist. Die Verhältnis-Karte stellt eine grobe Schätzung von Tiefe relativ zu benachbarten Pixeln bereit, was dem Blitz-/Nicht-Blitz-Pixel-Anpassungs-Algorithmus erlaubt, Pixel abzuwerten, welche bei verschiedenen Tiefen als das zentrale Pixel in dem Anpassungs-Fenster lokalisiert zu sein erscheinen. Zusätzlich kann die Verhältnis-Karte benutzt werden, um die erzeugte Tiefe-Karte zu filtern, um eine glatte Schätzung für die Tiefe von Objekten innerhalb des Stereo-Bildes zu erzeugen.



**Beschreibung****HINTERGRUND DER ERFINDUNG****GEBIET DER ERFINDUNG**

**[0001]** Die Erfindung betrifft im Allgemeinen Stereo-Bilder und insbesondere Techniken zum Erzeugen von robusten Stereo-Bildern von einem Paar von entsprechenden Stereo-Bildern, welche mit und ohne einer Lichtquelle, wie etwa einem Blitz-Gerät, aufgenommen sind.

**BESCHREIBUNG DER BETREFFENDEN TECHNIK**

**[0002]** Neuerdings haben dreidimensionale (3D) Stereo-Bilder und Video in dem Verbraucher-Markt Popularität erlangt. Die Einführung einer größeren Auswahl von 3D-Inhalt zusammen mit der Vermarktung von relativ günstigen 3D-HDTV (Hoch-Definitions-Fernsehen)-Sätzen hat das Betrachten von Bildern und Video in 3D gewöhnlicher gemacht. Ausstattungen von Computern mit Sätzen von Kameras und ein Berechnen von Tiefen und räumlichen Beziehungen von Stereo-Bild-Paaren ist wohl dokumentiert worden, mit Anwendungen in 3D-Modellieren, Roboter-Navigation, neue Bild-Synthese, angereicherte Realität und Spielen, neben anderen. Neuerdings ist Stereo-Bild-Gebung in Anwendungen so gewöhnlich wie handgehaltene Video-Kameras aufgetaucht, wie etwa die Fuji® Finepix 3D-Kamera und Sony® Bloggie-3D-Kamera.

**[0003]** Herkömmlicher Weise erzeugen Anwendungen Tiefe-Karten (depth maps) von erfassten Stereo-Bildern unter Benutzung eines grundsätzlichen Stereo-Rekonstruktions-Algorithmus, um einen Tiefenwert (depth value) für jedes Pixel dadurch zu erzeugen, dass Projektionen von Szene-Punkten über zwei oder mehr Bilder verglichen werden, welche von versetzten (offset) Stellen aufgenommen sind. Anders ausgedrückt, ist der Stereo-Rekonstruktions-Algorithmus im Wesentlichen eine Pixel-Anpassungs- bzw. -Anpassungs-Operation (pixel matching operation). Die Pixel-Übereinstimmung bzw. -Anpassung (pixel matching) wird typischerweise dadurch durchgeführt, dass die Summe von Quadraten minimiert wird, eine Pixel-Korrelation maximiert wird oder eine Rank-Transformierte oder Zensus-Transformierte angewendet wird und dann die Ränge oder Bit-Zeichen-Ketten (bit strings) angepasst werden bzw. zur Übereinstimmung gebracht werden (matching). Diese Algorithmen funktionieren ziemlich gut mit texturierten Oberflächen aber die Algorithmen haben Schwierigkeiten, exakte Übereinstimmungen auf Oberflächen mit einer uniformen Farbe zu machen. Zusätzlich können Pixel-Anpassungs-Algorithmen nahe an Verdeckungs-Grenzen (occlusion boundaries) scheitern, weil die Tiefen-Diskontinuität

bei den Grenzen dazu führen wird, dass lokale Pixel über die verschiedenen Bilder verschieden sind (d.h. die nahe Oberfläche verdeckt Pixel in dem Hintergrund in einem Bild aber ein verschiedenes Hintergrund-Objekt in dem anderen und daher haben diese Pixel keine entsprechende Übereinstimmung). Ein gutes Beispiel dafür, wo der herkömmliche Algorithmus scheitern kann, ist, wenn ein Bild durch ein feines Gitter aufgenommen wird, wie etwa einen Korbstuhl oder einen Kettenglied-Zaun bzw. Maschendrahtzaun (chain link fence).

**[0004]** Aus der Druckschrift US 2011 / 0 044 531 A1 ist ein System und Verfahren zur Extraktion von Tiefeninformationen aus mindestens zwei Bildern bekannt, die auf Regionen basierende Filterung zur Reduzierung von Artefakten verwenden.

**[0005]** Aus der Druckschrift US 2009 / 0 324 059 A1 ist ein fensterbasiertes Matching bekannt, um aus Bildern, die aus verschiedenen Orientierungen gewonnen wurden, eine Tiefenkarte zu bestimmen.

**[0006]** Aus der Druckschrift US 2012 / 0 039 525 A1 ist eine Einrichtung zum Erzeugen einer Disparitätskarte basierend auf einer Tiefenkarte, und eine Einrichtung zum Bestimmen der Genauigkeit von Pixeln in der Tiefenkarte bekannt, wobei die Bestimmungseinrichtung die Pixel entweder als genau oder ungenau identifiziert, basierend auf einer Vertrauenskarte und der Disparitätskarte. Aus der Druckschrift US 2008 / 0 297 621 A1 ist ein Verfahren zum Extrahieren von Vordergrundinformationen aus Bildinformationen bekannt, umfassend ein Empfangen eines blitzfreien Bildes eines Motivs, ein Empfangen eines Blitzbildes des Motivs, und ein Extrahieren von Vordergrundinformationen aus den Bildinformationen basierend auf einer Betrachtung von Blitzinformationen, Bewegungsinformationen und Farbinformationen.

**[0007]** Aus der Druckschrift US 2010 / 0 020 078 A1 ist ein Verfahren zum Zuordnen eines Objekts bekannt, umfassend ein Beleuchten des Objekts mit mindestens zwei Strahlen mit unterschiedlichen Strahlungscharakteristiken, wobei mindestens ein Bild des Objekts wird unter Beleuchtung mit jedem der mindestens zwei Strahlen aufgenommen wird.

**[0008]** Aus der Druckschrift Scharstein „A taxonomy and evaluation of dense two-frame stereo correspondence algorithms“ von Daniel Scharstein und Richard Szeliski, in International journal of computer vision 47.1-3 (2002): 7-42, ist eine Taxonomie von dichten, Zwei-Frame-Stereo-Verfahren bekannt, die darauf ausgelegt ist, die verschiedenen Komponenten und Design-Entscheidungen zu bewerten, die in einzelnen Stereo-Algorithmen getroffen werden.

**[0009]** Aus der Druckschrift „Discontinuity preserving stereo with small baseline multiframe illumination“ von Rogerio Feris et al. in Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1. Vol. 1. IEEE, 2005, ist ein Verfahren zur Berechnung einer qualitativen Tiefenkarte in einem Multiview-Aufbau bekannt, wobei neben den Tiefenrändern auch binokulare, halb verdeckte Pixel beschriftet werden.

**[0010]** Demgemäß ist, was in der Technik gebraucht ist, eine effektivere Zugangsweise zum Erzeugen von exakten, Pro-Pixel-Tiefe-Karten, welche mit Stereo-Bildern assoziiert sind.

#### ZUSAMMENFASSUNG DER ERFINDUNG

**[0011]** Eine Ausführungsform der vorliegenden Erfindung führt ein Verfahren zum Erzeugen von Tiefe-Karten von einem Satz von Stereo-Bildern aus. Das Verfahren umfasst die Schritte eines Erzeugens einer Verhältnis-Karte (ratio map), welche mit einem ersten Bild in dem Satz von Stereo-Bildern und mit einem entsprechenden zweiten Bild in dem Satz von Stereo-Bildern assoziiert ist, wobei das erste Bild mittels eines ersten Bild-Sensors unter Benutzung einer Umgebungslicht-Beleuchtung aufgenommen ist und wobei das zweite Bild mittels des ersten Bild-Sensors unter Benutzung einer Blitzbeleuchtung aufgenommen ist. Das erste Bild entspricht einem dritten Bild, welches mittels eines zweiten Bild-Sensors aufgenommen ist und welches im Wesentlichen simultan mit dem ersten Bild mittels des ersten Bild-Sensors aufgenommen ist. Ähnlich entspricht das zweite Bild einem vierten Bild, welches mittels des zweiten Bild-Sensors aufgenommen ist, im Wesentlichen simultan damit, dass das zweite Bild mittels des ersten Bild-Sensors aufgenommen ist. Das Verfahren umfasst ferner die Schritte eines Erzeugens einer Disparitäts-Karte, welche mit dem zweiten Bild assoziiert ist, basierend auf einem Vergleich des zweiten Bildes und des vierten Bildes, und eines Erzeugens einer Tiefe-Karte, welche mit dem zweiten Bild assoziiert ist. Das Verfahren kann auch die Schritte eines Erzeugens einer Verhältnis-Karte umfassen, welche mit dem dritten Bild und dem vierten Bild assoziiert ist, eines Erzeugens einer Disparitäts-Karte, welche mit dem vierten Bild assoziiert ist, und eines Erzeugens einer Tiefe-Karte, welche mit dem vierten Bild assoziiert ist. Für jedes Pixel in dem zweiten Bild assoziiert die Verhältnis-Karte das Pixel mit einem Verhältnis eines ersten Intensitäts-Wertes für das Pixel basierend auf dem zweiten Bild zu einem zweiten Intensitäts-Wert für ein entsprechenden Pixel basierend auf einem ersten Bild, wobei die Disparitäts-Karte das Pixel mit einem Disparitäts-Wert gemäß eines Pixel-Anpassungs-Algorithmus assoziiert, welcher Pixel in dem zweiten Bild mit entsprechenden Pixeln in dem vierten Bild vergleicht, wobei der Pixel-Anpassungs-Algorithmus zumindest eine Berechnung basierend auf dem Verhältnis umfasst, welches mit dem Pixel

assoziiert ist, und wobei die Tiefe-Karte das Pixel mit einem Tiefe-Wert assoziiert, welcher basierend auf dem Disparitäts-Wert berechnet ist, welcher mit dem Pixel assoziiert ist, basierend auf einer Fokusslänge und eines Grundlinie-Versatz-Abstands (baseline offset distance).

**[0012]** Eine andere Ausführungsform der vorliegenden Erfindung führt ein Computer-lesbares Speicher-Medium aus, welches Anweisungen umfasst, welche, wenn mittels einer Verarbeitungs-Einheit ausgeführt, die Verarbeitungs-Einheit veranlassen, Operationen zum Erzeugen von Tiefe-Karten von einem Satz von Stereo-Bildern gemäß des oben ausgeführten Verfahrens auszuführen. Noch eine andere Ausführungsform der vorliegenden Erfindung führt ein System zum Erzeugen von Tiefe-Karten von einem Satz von Stereo-Bildern aus. Das System umfasst einen Speicher und einen Prozessor. Der Speicher ist konfiguriert, den Satz von Stereo-Bildern zu speichern, welcher mittels zwei oder mehr Bild-Sensoren erfasst ist, welche ein erstes Bild, welches mittels eines ersten Bild-Sensors unter Benutzung einer Umgebungs-Lichtbeleuchtung erfasst ist, ein zweites Bild, welches mittels des ersten Bild-Sensors unter Benutzung einer Blitz-Beleuchtung erfasst ist, ein drittes Bild, welches mittels eines zweiten Bild-Sensors im Wesentlichen simultan damit aufgenommen bzw. erfasst ist, dass das erste Bild mittels des ersten Bild-Sensors erfasst ist, und ein viertes Bild umfassen, welches mittels des zweiten Bild-Sensors im Wesentlichen simultan damit erfasst ist, dass das zweite Bild mittels des ersten Bild-Sensors aufgenommen bzw. erfasst ist. Der Prozessor ist mit dem Speicher gekoppelt und konfiguriert, Operationen zum Erzeugen von Tiefe-Karten von einem Satz von Stereo-Bildern gemäß dem oben ausgeführten Verfahren durchzuführen.

**[0013]** Ein Vorteil des offenbarten Systems ist, dass die Stereo-Visions-Kamera keinen hohen Grad an Kalibrierung der Lichtquellen-Position oder - Intensitäts-Profiles erfordert, um robuste Tiefe-Karten zu erzeugen. Die hierin beschriebene Technik kann in vielen kostengünstigen handgehaltenen Geräten implementiert werden. Ein einfaches Vergleichen des Verhältnisses von erfasster Licht-Intensität unter Benutzung von verschiedenem Ausmaß von Beleuchtung stellt Information bereit, welche benutzt werden kann, um effektiv Pixel anzupassen bzw. in Übereinstimmung zu bringen, welche den denselben Oberflächen entsprechen, und um zwischen Pixeln zu unterscheiden, welche bei verschiedenen Tiefen in den Pixel-Anpassungs-Algorithmen lokalisiert sind.

#### Figurenliste

**[0014]** So dass die Weise, in welcher die oben rezierten Merkmale der vorliegenden Erfindung im Detail verstanden werden können, kann eine bestimm-

tere Beschreibung der Erfindung, welche oben kurz zusammengefasst ist, mittels Bezugnahme auf Ausführungsformen genommen werden, von denen einige in den angehängten Zeichnungen illustriert sind. Es ist jedoch zu bemerken, dass die angehängten Zeichnungen nur typische Ausführungsformen dieser Erfindung illustrieren und daher nicht zu betrachten sind, ihren Geltungsbereich zu begrenzen, da die Erfindung andere Äquivalente oder gleich effektive Ausführungsformen erlauben kann.

**Fig. 1** ist ein Blockdiagramm, welches ein Computer-System illustriert, welches konfiguriert ist, einen oder mehrere Aspekte der vorliegenden Erfindung zu implementieren;

**Fig. 2** detailliert die grundlegende Geometrie einer Stereo-Visions-Kamera gemäß einer Ausführungsform der vorliegenden Erfindung;

**Fig. 3** illustriert eine Stereo-Visions-Kamera zum Erfassen von Blitz-/Nicht-Blitz-Stereo-Bild-Paaren gemäß einer Ausführungsform der vorliegenden Erfindung;

**Fig. 4** illustriert ein Beispiel eines Blitz-/Nicht-Blitz-Stereo-Bild-Paares und der entsprechenden Verhältnis-Karten gemäß einer Ausführungsform der vorliegenden Erfindung;

**Fig. 5A** und **Fig. 5B** illustrieren einen Pixel-Anpassungs-Algorithmus zum Berechnen eines Disparitäts-Raum-Bildes (disparity space image) (DSI), welches mit einem Pixel assoziiert ist, gemäß einer Ausführungsform der vorliegenden Erfindung;

**Fig. 6A** und **Fig. 6B** illustrieren einen iterativen Filterungs-Algorithmus zum Erzeugen mehrerer robuster Tiefe-Karten, gemäß einer Ausführungsform der vorliegenden Erfindung;

**Fig. 7** ist ein Flussdiagramm von Verfahrensschritten zum Erzeugen einer Disparitäts-Karte, welche mit einem Stereo-Bild assoziiert ist, gemäß einer Ausführungsform der vorliegenden Erfindung; und

**Fig. 8** ist ein Flussdiagramm von Verfahrensschritten zum Erzeugen einer gefilterten Disparitäts-Karte, welche mit einem Blitz-Stereo-Bild assoziiert ist, gemäß einer Ausführungsform der vorliegenden Erfindung.

**[0015]** Zur Klarheit sind identische Bezugsnummern benutzt worden, wo anwendbar, um identische Elemente zu bezeichnen, welche zwischen Figuren gemeinsam sind. Es wird betrachtet, dass Merkmale der Ausführungsform in anderen Ausführungsformen ohne weitere Erwähnung inkorporiert werden können.

## DETAILLIERTE BESCHREIBUNG

**[0016]** In der folgenden Beschreibung werden zahlreiche spezifische Details ausgeführt, um ein durchgängigeres Verständnis der vorliegenden Erfindung bereitzustellen. Es wird jedoch für den gewöhnlichen Fachmann in der Technik ersichtlich sein, dass die vorliegende Erfindung ohne eine oder mehrere dieser spezifischen Details praktiziert werden kann.

### System-Übersicht

**[0017]** **Fig. 1** ist ein Blockdiagramm, welches ein Computer-System **100** illustriert, welches konfiguriert ist, einen oder mehrere Aspekte der vorliegenden Erfindung zu implementieren. Das Computer-System **100** kann ein Schreibtisch-Computer, Laptop-Computer, handgehaltenes Gerät, Mobiltelefon, PDA (persönlicher digitaler Assistent), Tablett-Computer, Kamera oder irgendein anderer oder andere bekannte Typen von Verbraucher-elektronischen Geräten sein.

**[0018]** Wie in **Fig. 1** gezeigt ist, umfasst Computersystem **100**, ohne Beschränkung, eine Zentralverarbeitungseinheit (CPU) **102** und einen Systemspeicher **104**, welcher über einen Zwischenverbindungspfad (interconnection path) kommuniziert, welcher eine Speicherbrücke **105** umfassen kann. Speicherbrücke **105**, welche z.B. ein Northbridge-Chip sein kann, ist über einen Bus oder einen anderen Kommunikationspfad **106** (z.B. HyperTransport-Link) mit einer I/O- (Eingabe/Ausgabe)-Brücke **107** verbunden. I/O-Brücke **107**, welche z.B. ein Southbridge-Chip sein kann, empfängt Benutzereingabe von einem oder mehreren Benutzer-Eingabegeräten **108** (z.B. Tastatur, Maus) und leitet die Eingabe an CPU **102** über Kommunikationspfad **106** und Speicherbrücke **105** weiter. Ein Parallel-Verarbeitungs-Subsystem **112** ist mit der Speicherbrücke **105** über einen Bus oder einen zweiten Kommunikationspfad **113** (z.B. einen Peripheral Component Interconnect-Express- (PCIe) -Accelerated Graphics Port, oder HyperTransport-Link) gekoppelt; in einer Ausführungsform ist das Parallel-Verarbeitungs-Subsystem **112** ein Grafik-Subsystem, welches Pixel an ein Anzeigegerät **110** (z.B. ein konventioneller Kathodenstrahlröhre- oder Flüssigkristallanzeige-basierter Monitor) liefert. Eine Systemplatte **114** ist auch mit der I/O-Brücke **107** verbunden. Ein Switch **116** stellt Verbindungen zwischen I/O-Brücke **107** und anderen Komponenten bereit, wie etwa ein Netzwerkadapter **118** und verschiedenen Hinzufügungskarten (Add-in-Cards) **120** und **121**. Andere Komponenten (nicht explizit gezeigt) umfassend Universal Serial Bus USB- oder andere Port-Verbindungen, Kompakt-Disk (CD)-Laufwerke, digitale Video-Disk (DVD)-Laufwerke, Filmaufnahmegegeräte, und dergleichen, können auch mit der I/O-Brücke **107** verbunden sein. Die verschiedenen Kommunikationspfade, welche in **Fig. 1** gezeigt sind, einschließlich die speziell genannten Kommunikations-

pfade **106** und **113**, können unter Benutzung irgendwelcher geeigneten Protokolle implementiert sein, wie etwa PCI-Express, AGP (Accelerated Graphics Port), HyperTransport, oder irgendeines oder irgendwelcher Bus- oder Punkt-zu-Punkt-Kommunikations-Protokoll(e), und Verbindungen zwischen verschiedenen Geräten können verschiedene Protokolle benutzen, wie in der Technik bekannt ist.

**[0019]** In einer Ausführungsform inkorporiert das Parallel-Verarbeitungs-Subsystem **112** Schaltung, welche für Grafik- und Video-Verarbeitung optimiert ist, einschließlich zum Beispiel Videoausgabe-Schaltung, und konstituiert eine Grafik-Verarbeitungseinheit (GPU). In einer anderen Ausführungsform umfasst das Parallel-Verarbeitungs-Subsystem **112** Schaltung, welche für Allgemeinzwirk-Verarbeitung optimiert ist, während die darunter liegende Computer-Architektur, welche im größeren Detail hierin beschrieben ist, beibehalten ist. In noch einer anderen Ausführungsform kann das Parallel-Verarbeitungs-Subsystem **102** mit einem oder mit mehreren anderen Systemelementen in ein einzelnes Subsystem integriert sein, wie etwa die Speicherbrücke **105**, CPU **102** und I/O-Brücke **107** verbindend, um ein System auf dem Chip (system on chip) (SoC) zu bilden.

**[0020]** Es wird geschätzt werden, dass das hierin gezeigte System illustrativ ist und dass Variationen und Modifikationen möglich sind. Die Verbindungstopologie, einschließlich der Anzahl und der Anordnung von Brücken, der Anzahl von CPUs **102**, und der Anzahl von Parallel-Verarbeitungs-Subsystemen **112** kann wie gewünscht modifiziert werden. Zum Beispiel ist in einigen Ausführungsformen Systemspeicher **104** mit CPU **102** direkt gekoppelt anstatt durch eine Brücke, und andere Geräte kommunizieren mit Systemspeicher **104** über Speicherbrücke **105** und CPU **102**. In anderen alternativen Topologien ist das Parallel-Verarbeitungs-Subsystem **112** mit I/O-Brücke **107** oder direkt mit CPU **102** verbunden anstatt mit der Speicherbrücke **105**. In noch anderen Ausführungsformen können die I/O-Brücke **107** und Speicherbrücke **105** in einen einzelnen Chip integriert sein, anstatt als ein oder mehr Geräte zu existieren. Große Ausführungsformen können zwei oder mehr CPUs **102** und zwei oder mehr Parallel-Verarbeitungs-Subsysteme **112** umfassen. Die besonderen Komponenten, welche hierin gezeigt sind, sind optional; z.B. könnte irgendeine Anzahl von Hinzufügungskarten oder peripheren Geräten unterstützt sein. In einigen Ausführungsformen ist der Switch **116** eliminiert und der Netzwerkadapter **116** und Hinzufügungskarten **120**, **121** verbinden direkt mit der I/O-Brücke **107**.

**[0021]** Fig. 2 detailliert die grundsätzliche Geometrie von Stereo-Visions-Kamera **200** gemäß einer Ausführungsform der vorliegenden Erfindung. Die Stereo-Visions-Kamera kann ein oder mehrere Elementen

von Computer-System **100** umfassen. Zum Beispiel kann die Stereo-Visions-Kamera eine CPU **102** und einen System-Speicher **104**, welche durch einen System-Bus verbunden sind, umfassen. Die Stereo-Visions-Kamera **200** kann auch ein Parallel-Verarbeitungs-Subsystem **112** zum Verarbeiten von Bild-Daten umfassen. In einer Ausführungsform umfasst die Stereo-Visions-Kamera die CPU **102** und die GPU **112** auf demselben integrierten Schaltkreis, um einen SoC zu bilden. Die Stereo-Visions-Kamera **200** kann auch Anwendungen oder Firmware umfassen, welche in dem System-Speicher **104** gespeichert sind, welche, wenn mittels CPU **102** oder GPU **112** ausgeführt, die Stereo-Visions-Kamera **200** in die Lage versetzen, Stereo-Bilder zu verarbeiten und Tiefe-Karten zu erzeugen, wie vollständiger unten beschrieben ist. Ferner können die Elemente von Stereo-Visions-Kamera innerhalb des Computer-Systems **100** umfasst sein, etwa mittels eines Umfassens von zwei Bild-Sensoren in einem Tablett-Computer, Mobiltelefon oder einem anderen handgehaltenen Gerät.

**[0022]** Wie in Fig. 2 gezeigt ist, kann eine Stereo-Visions-Kamera **200** auch zwei oder mehr versetzte Bild-Sensoren haben, welche die Stereo-Visions-Kamera **200** in die Lage versetzen, einen Satz von Stereo-Bildern zu erfassen. In einer Ausführungsform können die Digital-Bilder in dem System-Speicher **104** oder in irgendeinem anderen nicht-volatilen Speicher gespeichert sein, welcher entfernbar sein kann oder nicht entfernbar sein kann. Die Bilder können in Speicher über ein nicht komprimiertes Format gespeichert sein, wie etwa eine 32bpp BMP (bitmap)-Format oder ein komprimiertes Format, wie etwa JPEG (Joint Photographic Experts Group)-Format. Die Stereo-Visions-Kamera **200** kann dann die Bilder, welche von dem Speicher **104** gelesen sind, verarbeiten. In einer anderen Ausführungsform können die rohen Bild-Sensor-Daten über eine stereoskopische Bild-Verarbeitungs-Pipeline direkt, nachdem sie von den Bildsensoren gelesen wurden, verarbeitet werden. Die verarbeiteten Daten können dann auf System-Speicher **104** in einem der verfügbaren Formate heraus geschrieben werden.

**[0023]** Fig. 2 illustriert ferner ein Erfassen (capturing) eines Stereo-mikroskopischen Bild-Paares unter Benutzung zweier Bild-Sensoren. Ein linker Bild-Sensor ( $C_L$ ) **201** erfasst ein linkes Bild und ein rechter Bild-Sensor ( $C_R$ ) **202** erfasst ein rechtes Bild im Wesentlichen gleichzeitig. Jeder Bild-Sensor **201**, **202** kann mit einer Linse gepaart sein, welche Licht auf die Oberfläche des Bild-Sensors fokussiert. Wie z.B. gezeigt ist, ist eine linke Linse **203** mit dem linken Bild-Sensor **201** gepaart und eine rechte Linse **204** ist mit dem rechten Bild-Sensor **202** gepaart. Obwohl es in Fig. 2 nicht explizit gezeigt ist, kann jeder Bild-Sensor in eine Bild-Sensor-Anordnung integriert sein, welche eine Linse, einen Bild-Sensor, einen Verschluss-Mechanismus und einen oder meh-

rere Filter umfasst, wie etwa einen Polarisierungs-Filter oder Farb-Filter, welche entfernbar sein können, oder nicht entfernbar zu sein brauchen bzw. nicht entfernbar sein müssen. In einer Ausführungsform können die Bild-Sensoren **201**, **202** CMOS-(Komplementär-Metalloxid-Halbleiter)-Bild-Sensoren sein. In einer anderen Ausführungsform können die Bild-Sensoren **201**, **202** CCD (Ladungsgekoppeltes-Gerät)-Bild-Sensoren sein.

**[0024]** In einer typischen Konfiguration sind die Bild-Sensoren **201**, **202** in einer ersten Dimension um einen Grundlinie-(B)-Versetzungs- bzw. - Verschiebung-Abstandes versetzt bzw. verschoben (offset). Die Grundlinie-Verschiebung ist es, was erlaubt, dass die erfassten Bilder analysiert werden können, um die Tiefe zu bestimmen, welche mit Objekten in einem entsprechenden Stereo-mikroskopischen Bild-Paar assoziiert ist. Wie z.B. in **Fig. 2** gezeigt ist, ist ein Objekt **210** bei einer Position innerhalb des überlappenden Sicht-Feldes für den linken Bild-Sensor **201** und den rechten Bild-Sensor **202** gezeigt. Das Objekt ist ein Abstand (Z) von der Oberfläche der Linsen **203**, **204** der Stereo-Visions-Kamera **200**. Die Linsen **203**, **204** sind mit einer Fokus-Länge (f) assoziiert, was der Abstand von den Linsen **203**, **204** zu einer Oberfläche des Bild-Sensors ist (d.h. die Brennebene), bei welcher alles kollimierte Licht, welches durch die Linse passiert, auf einen einzelnen Punkt auf dem Bild-Sensor konvergiert. Die Fokus-Länge ist auf die Geometrie und das Material der Linsen **203**, **204** bezogen bzw. davon abhängig. Wie für die Fachleute in der Technik wohl bekannt ist, sind die Fokus-Länge und die Form des Sichtfeldes (FOV) invers proportional. Mit anderen Worten, wenn die Fokus-Länge des optischen Systems vermindert wird, vergrößert sich das Sichtfeld des optischen Systems. Die Linsen **203**, **204** werden bei einer Position vor dem Bild-Sensor lokalisiert, welche gleich ist der Fokus-Länge (f), derart, dass Licht, welches durch die Linse passiert, auf dem Bild-Sensor fokussiert wird.

**[0025]** Wenn ein Benutzer ein Bild erfasst bzw. aufnimmt, wird Licht, welches von der Oberfläche von Objekt **210** reflektiert ist, auf die Oberfläche der Bild-Sensoren **201**, **202** projiziert. Die vordere linke Ecke ( $P_1$ ) des Objektes **210** ist auf den linken Bild-Sensor **201** bei einer ersten Pixel-Position  $X_1$  projiziert bzw. projiziert und die vordere rechte Ecke ( $P_2$ ) von Objekt **210** ist auf den linken Bild-Sensor **201** bei einer zweiten Pixel-Position  $X_2$  projiziert bzw. projiziert. Ähnlich wird die vordere linke Ecke ( $P_1$ ) des Objektes **210** auf den rechten Bild-Sensor **202** bei einer dritten Pixel-Position  $X_3$  projiziert und die vordere rechte Ecke ( $P_2$ ) von Objekt **210** wird auf den rechten Bild-Sensor **202** bei einer vierten Pixel-Position  $X_4$  projiziert. Es wird geschätzt werden, dass abhängig von der Konfiguration der Linse (oder Linsen), das Bild, welches mittels eines Abtastens der Bild-Sensoren **201**, **202** erfasst ist, sowohl in der horizontalen (wie

in **Fig. 2** gezeigt) als auch der vertikalen Richtung invertiert sein kann. Die Stereo-Visions-Kamera **200** kann konfiguriert sein, die Inversion dadurch zu korrigieren, dass Pixel in der entgegen gesetzten Ordnung abgetastet bzw. gesammelt werden.

**[0026]** Wie auch in **Fig. 2** gezeigt ist, sind die projizierte Position des Objekts **210** in dem linken und dem rechten Bild nicht identisch aufgrund der relativen Position von Objekt **210** bezüglich der Lokalisierung der Bild-Sensoren **201**, **202**. Dieser Effekt wird Parallaxe genannt und ist von Stereo-Visions-Systemen ausgenutzt, um Benutzern zu ermöglichen, einen 3D-Effekt zu erfahren. Mit anderen Worten passiert Licht von Objekten, welche weit entfernt von der Kamera lokalisiert sind (d.h. im Hintergrund) durch die linke Linse **203** und die rechte Linse **204** bei ungefähr demselben Winkel und wird auf dieselbe Pixel-Stelle in dem linken Bild-Sensor **201** und dem rechten Bild-Sensor **202** projiziert. Im Gegensatz dazu passiert Licht von Objekten, welche näher an der Kamera lokalisiert sind, durch die linke Linse **203** und die rechte Linse **204** bei verschiedenen Winkeln und wird daher auf verschiedene Pixel-Stellen auf den Bild-Sensoren **201**, **202** projiziert. Die Disparität in den projizierten Pixel-Positionen wird immer in derselben Dimension (z.B. horizontal) wie die Grundlinie-Versetzung zwischen den Bild-Sensoren **201**, **202** sein. Diese Disparität (D) zwischen den Pixel-Positionen eines projizierten Punktes in dem linken Bild und demselben projizierten Punkt in dem rechten Bild ermöglicht einer Verarbeitungs-Einheit die erfassten Bilder zu analysieren, um eine Tiefe abzuschätzen, welche mit den verschiedenen Punkten (d.h. Pixeln) innerhalb der Szene assoziiert ist. Ein Berechnen der Disparität (D) für die projizierten Pixelstellen von der vorderen linken Ecke ( $P_1$ ) und der vorderen rechten Ecke ( $P_2$ ) von Objekt **210** ist mittels eines Lösen von Gleichungen 1 bzw. 2 durchgeführt, welche unten gezeigt sind.

$$D(P_1) = x_1 - x_3 \quad (\text{Gl. 1})$$

$$D(P_2) = x_2 - x_4 \quad (\text{Gl. 2})$$

**[0027]** Wie gezeigt ist, sind  $x_1$  und  $x_2$  Dimensionen, welche von der linken Kante des linken Stereo-Bildes, welches über den linken Bild-Sensor **201** erfasst ist, gemessen sind, und  $x_3$  und  $x_4$  sind Dimensionen, welche von der linken Kante des rechten Stereo-Bildes gemessen sind, welches über den rechten Bild-Sensor **202** erfasst ist. Es wird geschätzt werden, dass die linke Kante der erfassten Stereo-Bilder nicht mit den am weitesten links angeordneten Pixeln des entsprechenden Bild-Sensors entsprechen muss. Kalibrierung der Stereo-Visions-Kamera **200** kann durchgeführt werden, um das linke Bild und das rechte Bild derart gleich auszurichten (rectify), dass das am weitesten links liegende Pixel des linken Stereo-Bil-

des mit dem am weitesten links liegenden Pixel des rechten Stereo-Bildes in der horizontalen Richtung für Licht entspricht, welches von einem Objekt ausgeht, welches in dem Hintergrund lokalisiert ist.

**[0028]** Gegeben die berechneten Disparitäten ( $D$ ) der vorderen linken Ecke ( $P_1$ ) und der vorderen rechten Ecke ( $P_2$ ) des Bildes **210** wird ein Berechnen der Tiefen ( $Z$ ) für die vordere linke Ecke ( $P_1$ ) und die vordere rechte Ecke ( $P_2$ ) von Objekt **210** mittels eines Lösen von Gleichungen 3 bzw. 4 durchgeführt, wie unten gezeigt ist.

$$Z(P_1) = \frac{f \cdot B}{D(P_1)} = \frac{f \cdot B}{x_2 - x_3} \quad (\text{Gl. 3})$$

$$Z(P_2) = \frac{f \cdot B}{D(P_2)} = \frac{f \cdot B}{x_2 - x_4} \quad (\text{Gl. 4})$$

**[0029]** Die Fokus-Länge ( $f$ ) und die Grundlinie ( $B$ ) sind bekannte Konstanten, welche basierend auf der Konfiguration der Bild-Sensoren **201**, **202** und den Linsen **203**, **204**, welche in der Stereo-Visions-Kamera **200** implementiert sind, definiert sind. Daher ist, wie Gleichungen 3 und 4 klar machen, die Tiefe invers proportional zu der Menge bzw. Größe von Disparität zwischen entsprechenden Pixeln in dem linken und dem rechten Bild. Folglich wird ein Berechnen der Tiefen für jedes Pixel eines linken oder eines rechten Bildes dadurch durchgeführt, dass jedes Pixel in einem Bild mit einem entsprechenden Pixel in dem anderen Bild zur Übereinstimmung gebracht wird bzw. angepasst wird, dass die Disparitäten zwischen der entsprechenden übereinstimmenden Pixelstellen berechnet werden, und dass die Tiefe jedes Pixels basierend auf der berechneten Disparität bestimmt wird. In Praxis begegnet man jedoch vielen Problemen, wenn versucht wird, die Pixel-Stellen in jedem Bild zur Übereinstimmung zu bringen bzw. anzupassen. Zum Beispiel benutzen herkömmliche Pixel-Anpassungs-Algorithmen typischerweise eine Fenster-Funktion, um ein kleines Fenster von Pixeln in einem Bild mit einem ähnlichen Fenster von Pixeln in dem entsprechenden Bild anzupassen bzw. zur Übereinstimmung zu bringen. Nahe bei Tiefen-Diskontinuitäts-Grenzen (d.h. wo das Fenster Pixel umfassen kann, welche mit mehreren Oberflächen assoziiert sind, welche bei verschiedenen Tiefen lokalisiert sind) können die Fenster um die tatsächlichen entsprechenden Pixel herum verschieden sein aufgrund einer Verdeckung, welche durch verschiedene räumliche Stellen der Bild-Sensoren und Objekte in der Szene verursacht ist. Ferner kann ein zur-Übereinstimmung-Bringen bzw. Anpassen von Fenstern auf gleichförmigen Oberflächen schwierig sein, wo mehrere potentielle Fenster innerhalb der Suche-Ausdehnungen ähnliche Charakteristiken teilen können. Typischerweise stellen

herkömmliche Pixel-Anpassungs-Algorithmen Niedrig-Qualitäts-Tiefe-Karten bereit, welche dazu neigen, eine schwache Auflösung bei Objekt-Kanten zu haben.

#### Blitz-/Nicht-Blitz-Stereo-System

**[0030]** Fig. 3 illustriert eine Stereo-Visions-Kamera **300** zum Erfassen von Blitz/Nicht-Blitz-Stereo-Bild-Paaren gemäß einer Ausführungsform der vorliegenden Erfindung. Wie in Fig. 3 gezeigt ist, ist die Stereo-Visions-Kamera **300** ähnlich zu der Stereo-Visions-Kamera **200**. Stereo-Visions-Kamera **300** umfasst jedoch ein Blitz-Gerät **330**, welches nahe an den Bild-Sensoren **201**, **202** lokalisiert ist. Das Blitz-Gerät **330** kann ein herkömmliches Blitz-Gerät sein, wie etwa eine elektronische Blitz-Röhre, welche mit Xenon-Gas gefüllt ist und mittels einer Hoch-Spannungs-Entladung angeregt wird, um einen kurzen Blitz oder einen Puls von Licht zu emittieren. Es wird geschätzt werden, dass das Blitz-Gerät **330** irgendein herkömmliches Blitz-Gerät sein kann, welches in der Lage ist, einen kurzen Puls (burst) von Licht zu emittieren, um eine Szene zu beleuchten. Eine präzise Positionierung des Blitz-Gerätes **330** relativ zu der Stelle der Bild-Sensoren **201**, **202** (und die entsprechende erforderliche Kalibrierung) ist nicht wichtig, um gewisse Aspekte der vorliegenden Erfindung zu implementieren. Herkömmliche fotometrische Tiefe-Berechnungs-Techniken erfordern einen hohen Grad einer Kalibrierung, um die Tiefe von Objekten innerhalb der Szene zu bestimmen. Zum Beispiel kann ein Abtasten bzw. Sampeln oder ein Steuern der Umgebungs-Licht-Bedingungen und der Intensität und der Positionierung des Blitz-Gerätes erforderlich sein. Die hierin beschriebene Blitz-/Nicht-Blitz-Technik erfordert nicht denselben Grad einer Kalibrierung, um robuste Tiefe-Karten zu erzeugen.

**[0031]** Um ein robustes Stereo-Bild unter Benutzung einer Stereo-Visions-Kamera **300** zu erzeugen, sollten zwei Paare von Stereo-Bildern erfasst werden, ein erstes Stereo-Bild-Paar (hierin als  $G_l$  und  $G_r$  genannt) und ein zweites Stereo-Bild-Paar (hierin als  $F_l$  und  $F_r$  benannt). Das erste Stereo-Bild-Paar ( $G_l$ ,  $G_r$ ) stützt sich nicht auf das Blitz-Gerät **330**, um die Szene zu beleuchten, sondern stützt sich stattdessen auf das Umgebungs-Licht, um die Szene zu beleuchten. Das zweite Stereo-Bild-Paar ( $F_l$ ,  $F_r$ ) stützt sich auf das Blitz-Gerät **330** sowie auf das Bewegungs-Licht, um die Szene zu beleuchten. In alternativen Ausführungsformen können mehr als zwei Bild-Sensoren implementiert sein, um N-Ansicht-3D-Bilder zu erfassen (z.B. zur Benutzung mit Auto-stereoskopischen Anzeigen). Aus Illustrations-Zwecken bezieht sich die vorliegende Offenbarung auf 2-Ansicht-stereoskopische Bilder, aber die hierin beschriebenen Techniken können auf alternative Ausführungsformen unter Benutzung von N-Ansicht-Implementierungen ausgedehnt werden.

**[0032]** Es wird geschätzt werden, dass die Stereo-Visions-Kamera **300** kalibriert werden kann, um die digitalen Bilder, welche mit dem linken Bild-Sensor **201** und dem rechten Bild-Sensor **202** erfasst werden, derart auszugleichen, dass irgendein Oberflächen-Punkt, welcher für beide Kameras sichtbar ist, auf dieselbe Scan-Linie (d.h. in der vertikalen Dimension) in den erfassten Bildern projiziert. Mit anderen Worten können die erfassten Bilder einem Unterabschnitt von Pixeln der Bild-Sensoren **201**, **202** derart entsprechen, dass ein Rand von Pixeln an den Kanten der Bild-Sensoren nicht zu den Pixeln der erfassten Bilder beiträgt. Ferner kann die Anzahl von Pixeln in den Rändern des linken Bild-Sensors **201** verschieden sein von der Anzahl von Pixeln in den entsprechenden Rändern des rechten Bild-Sensors **202**. Zum Beispiel kann der obere Rand des linken Bild-Sensors **201** größer sein als der obere Rand des rechten Bild-Sensors **202**, um eine Fehl-Ausrichtung in der vertikalen Richtung der Bild-Sensoren **201**, **202** zu korrigieren, wie sie in der Stereo-Visions-Kamera **300** installiert sind. Ferner können die linken und rechten Ränder des linken Bild-Sensors **201** und des rechten Bild-Sensors **202** derart eingestellt werden, dass ein Punkt, welcher bei einem Abstand (Z) von unendlich von der Stereo-Visions-Kamera **300** lokalisiert ist (d.h. ein genügend großer Abstand, dass die Strahlen, welche von dem Objekt reflektiert sind, sowohl bei der linken Linse **203** als auch bei der rechten Linse **204** im Wesentlichen parallel ankommen) bei der horizontalen Grenze des erfassten Bildes ausgerichtet ist. Kalibrierung der Stereo-Visions-Kamera **300** in einer solchen Weise stellt sicher, dass Bilder, welche in dem Hintergrund der Szene lokalisiert sind, eine Disparität von Null haben.

**[0033]** Um die Tiefe (Z) für einen Punkt P 320 zu auswerten, welcher auf einer Oberfläche eines Objekts in der Szene lokalisiert ist, werden zwei Sätze von Stereo-Bild-Paaren analysiert. Unter der Annahme, dass die Oberfläche des Objekts, auf welcher Punkt P lokalisiert ist, eine Lambert'sche Oberfläche approximiert (d.h., wo Licht, welches von der Oberfläche reflektiert ist, derart gestreut ist, dass die ersichtliche bzw. scheinbare Helligkeit der Oberfläche für einen Beobachter dieselbe ist unabhängig von dem Sichtwinkel des Beobachters), ist die gemessene Intensität von Licht, welches von Punkt P zu jedem der Bild-Sensoren **201**, **202** während der Erfassung des Nicht-Blitz-Stereo-Bild-Paares ( $G_l$ ,  $G_r$ ) reflektiert ist, gegeben durch  $G(P)$  in Gleichung 5.

$$G(P) = \eta \cdot I_a \cdot R_s \quad (\text{Gl. 5})$$

**[0034]** Wie in Gleichung 5 gezeigt ist, betrifft die Intensität von Licht  $G(P)$  bei Punkt P in dem Nicht-Blitz-Stereo-Bild-Paar ( $G_l$ ,  $G_r$ ) die Umgebungs-Beleuchtung, Oberflächen-Form, und Reflektivität mittels der Beziehung, welche oben ausgeführt ist, wobei  $\eta$  eine

Proportionalitäts-Konstante zwischen einer gemessenen Irradianz und einer Szenen-Radianz ist,  $I_a$  ist die Intensität von Umgebungs-Beleuchtung bei Punkt P und  $R_s$  ist die Oberflächen-Reflektivität bei Punkt P. Im Gegensatz dazu ist die gemessene Intensität von Licht, welches von Punkt P auf jeden der Bild-Sensoren während der Erfassung des Blitz-Stereo-Bild-Paares ( $F_l$ ,  $F_r$ ) reflektiert ist, durch  $F(P)$  in Gleichung 6 gegeben.

$$F(P) = (\eta \cdot I_a \cdot R_s) + \left( \eta \cdot I_f \cdot \frac{\langle \hat{n} \cdot \hat{v} \rangle}{r^2} \cdot R_5 \right) \quad (\text{Gl. 6})$$

**[0035]** Wie in Gleichung 6 gezeigt ist, betrifft bzw. bezieht sich die Intensität von Licht  $F(P)$  bei Punkt P in dem Blitz-Stereo-Bild-Paar ( $F_l$ ,  $F_r$ ) auf Umgebungs-Beleuchtung, Blitz-Beleuchtung, Oberflächen-Form und Reflektivität mittels der Beziehung, welche oben ausgeführt ist, wobei  $\eta$  eine Proportionalitäts-Konstante zwischen der gemessenen Irradianz und der Szene-Radianz ist, wobei  $I_a$  die Intensität von Umgebungs-Licht-Beleuchtung bei Punkt P ist,  $R_s$  die Oberflächen-Reflektivität bei Punkt P ist,  $I_f$  die Intensität des Blitzes ist,  $r$  der Abstand von dem Punkt P zu dem Blitz ist und  $\langle \hat{n} \cdot \hat{v} \rangle$  ist das innere Produkt zwischen dem Oberflächen-Normalen-Vektor  $\hat{n}$  321 und dem Richtungs-Vektor zu dem Blitz  $\hat{v}$  322. Für ein entsprechendes Paar von Blitz-/Nicht-Blitz-Bildern für entweder den linken Bild-Sensor **201** oder den rechten Bild-Sensor **202** können wir  $G(P)$  und  $F(P)$  dadurch messen, dass die erfassten Intensitäts-Werte von jeder Farb-Komponente für Pixel in den entsprechenden Stereo-Bildern analysiert werden.

**[0036]** Die Genauigkeit der Messung nimmt an, dass die Szene statisch ist, was dadurch approximiert werden kann, dass die Blitz-/Nicht-Blitz-Bilder im Wesentlichen simultan erfasst werden. Zum Beispiel kann Stereo-Visions-Kamera **300** in der Lage sein, die Bild-Sensoren **201**, **202** einem Umgebungs-Licht ohne einen Blitz für 1/100 einer Sekunde auszusetzen. Die Stereo-Visions-Kamera **300** schließt dann den Verschluss und tastet die Bild-Sensoren ab, um das Nicht-Blitz-Stereo-Bild-Paar ( $G_l$ ,  $G_r$ ) zu erfassen. Einige Zehntel einer Sekunde später setzt die Stereo-Visions-Kamera **300** die Bild-Sensoren **201**, **202** Umgebungs-Licht mit einem Blitz für 1/100 einer Sekunde aus, schließt den Verschluss, und tastet die Bild-Sensoren ab, um das Blitz-Stereo-Bild-Paar ( $F_l$ ,  $F_r$ ) zu erfassen. Da beide Sätze von Stereo-Bild-Paaren innerhalb weniger Zehntel einer Sekunde erfasst werden, entsprechen ähnliche Pixel in dem linken Bild-Sensor **201** oder dem rechten Bild-Sensor **202** ähnlichen Punkten auf Objekten in sowohl dem Nicht-Blitz-Stereo-Bild-Paar ( $G_l$ ,  $G_r$ ) und dem Blitz-Stereo-Bild-Paar ( $F_l$ ,  $F_r$ ), unter der Annahme, dass die Szene keine sich schnell bewegenden Objekte umfasst. In einer Ausführungsform kann eine kleine Bewe-

gung zwischen der Erfassung von entsprechenden Stereo-Bild-Paaren dadurch korrigiert werden, dass eine Transformation berechnet wird (typischerweise eine Homografie), welche das Blitz-Stereo-Bild-Paar ( $F_l, F_r$ ) mit dem Nicht-Blitz-Stereo-Bild-Paar ( $G_l, G_r$ ) ausrichtet.

**[0037]** Durch Nehmen des Verhältnisses der abgetasteten Intensitäts-Werte für die Komponenten jedes Pixels in den entsprechenden Blitz/Nicht-Blitz-Bildern für sowohl den linken Bild-Sensor **201** als auch den rechten Bild-Sensor **202** wird ein Paar von Verhältnis-Karten ( $R_l, R_r$ ) erzeugt, welche den Unterschied zwischen der Helligkeit von Pixeln in den Stereo-Bild-Paaren vergleichen, welche mit und ohne Blitz-Beleuchtung erfasst sind. In einer Ausführungsform ist eine Verhältnis-Karte für ein Blitz-/Nicht-Blitz-Stereo-Bild als der log des Verhältnisses zwischen der Intensität von Licht ( $F, P$ ) bei irgendeinem Punkt  $P$  in dem Blitz-Stereo-Bild-Paar ( $F_l, F_r$ ) und der Intensität von Licht  $G(P)$  bei dem Punkt  $P$  in dem Nicht-Blitz-Stereo-Bild-Paar ( $G_l, G_r$ ) definiert, wie unten in Gleichung 7 gezeigt ist.

$$R(P) = \log\left(\frac{F(P)}{G(P)}\right) = \log\left(1 + \frac{I_f}{I_a} + \frac{\langle \hat{n} \cdot \hat{v} \rangle}{r^2}\right) \quad (\text{Gl. 7})$$

**[0038]** Wie in Gleichung 7 gezeigt ist, ist die Verhältnis-Karte für irgendeinen Punkt  $P$  unabhängig von einer Oberflächen-Reflektivität  $A_s$  und variiert basierend auf dem Oberflächen-Normalen-Vektor  $n$  321, Richtungs-Vektor zu dem Blitz  $v$  322 und dem Objekt-Abstand  $r$ . Diese Unabhängigkeit gilt selbst dann, wenn die Belichtungs-Zeit ( $t_f, t_r$ ) der Blitz- und Nicht-Blitz-Bilder verschieden ist und selbst dann, wenn Gamma-Korrektur angewendet worden ist, wie mittels Gleichung 8 illustriert ist.

$$R(P) = \log\left(\frac{[t_f \cdot F(P)]^\gamma}{[t_r \cdot G(P)]^\gamma}\right) = \gamma \cdot \left\{ \log\left(\frac{t_f}{t_r}\right) + \log\left(\frac{F(P)}{G(P)}\right) \right\} \quad (\text{Gl. 8})$$

**[0039]** In einer Ausführungsform ist, um eine Berechnung zu vermeiden, welche zu einer Division durch Null führt, die Verhältnis-Karte durch Gleichung 9 definiert, wobei  $\varepsilon$  eine kleine Zahl ist.

$$R(P) = \log(F(P) + \varepsilon) - \log(G(P) + \varepsilon) \quad (\text{Gl. 9})$$

**[0040]** Die Verhältnis-Karte ( $R_l, R_r$ ) ist im Wesentlichen unabhängig von Szene-Albedo und ist stattdessen bezogen auf bzw. abhängig von dem Abstand von dem Blitz-Gerät **330** und der lokale-Oberflächen-Orientierung. Obwohl die Berechnung nicht

vollständig für Nicht-Lambert'sche Oberflächen genau ist, rühren benachbarte Pixel mit ähnlichen Verhältnis-Werten wahrscheinlich von derselben Oberfläche her. Zum Beispiel wird ein Punkt  $O$ , welcher benachbart bzw. nächst einem Punkt  $P$  auf der Oberfläche ist, welche in **Fig. 3** gezeigt ist, ungefähr ähnliche Verhältnis-Werte haben, weil der Oberflächen-Normalen-Vektor  $n$  für Punkt  $O$  ähnlich ist zu Oberflächen-Normalen-Vektor  $n$  für Punkt  $P$ , weil der Richtungs-Vektor  $v$  für Punkt  $O$  ähnlich ist dem Richtungs-Vektor  $v$  für Punkt  $P$  und weil der Abstand  $r$  von Punkt  $O$  zu dem Blitz ähnlich dem Abstand  $r$  von Punkt  $P$  zu dem Blitz ist. Somit sollten die Verhältnis-Werte für nächst zueinander gelegene Pixel derselben Oberfläche ähnlich sein. Wenn die Verhältnis-Werte für nächst gelegene Pixel nicht ähnlich sind, dann kann dies anzeigen, dass die Pixel benachbart oder nächst einer Tiefe-Grenze lokalisiert sind (d.h. die Pixel sind mit verschiedenen Oberflächen von verschiedenen Objekten assoziiert, welche bei verschiedenen Tiefen lokalisiert sind) oder dass die Pixel nächst bzw. benachbart einer Kante des Objektes lokalisiert sind, wo sich die Oberflächen-Normalen-Orientierung abrupt ändert.

**[0041]** **Fig. 4** illustriert ein Beispiel eines Blitz-/Nicht-Blitz-Stereo-Bild-Paares und der entsprechenden Verhältnis-Karten ( $R_l, R_r$ ) gemäß einer Ausführungsform der vorliegenden Erfindung. Ein erstes Stereo-Bild-Paar  $G_l$  **401** und  $G_r$  **402** ist ohne Beleuchtung von dem Blitz-Gerät **330** erfasst. Eine kurze Zeit später wird ein zweites Stereo-Bild-Paar  $F_l$  **411** und  $F_r$  **412** mit Beleuchtung von dem Blitz-Gerät **330** erfasst. Wie in den Bildern **401, 402, 411** und **412** gezeigt ist, ist ein erstes Objekt **451** in dem Vordergrund der Szene lokalisiert und daher ist das Licht, welches auf Pixel der Bild-Sensoren **201, 202** projiziert ist, welches mit dem ersten Objekt **451** korrespondiert, mit einer Nicht-Null-Disparität zwischen den Pixel-Positionen assoziiert, welche dem linken und dem rechten Stereo-Bild entsprechen. Ein zweites Objekt **452** ist dem Hintergrund der Szene lokalisiert und daher ist Licht, welches auf Pixel der Bild-Sensoren **201, 202** projiziert ist, welches dem zweiten Objekt **452** entspricht, mit einer Null-Disparität zwischen den Pixel-Positionen assoziiert, welche dem linken und dem rechten Bild entsprechen. Ein drittes Objekt **453** ist bei einer Tiefe ( $Z$ ) lokalisiert, welche weniger ist als die Tiefe von Objekt **452** und größer ist als die Tiefe von Objekt **451**. Daher ist das Licht, welches auf Pixel der Bild-Sensoren **201, 202** projiziert wird, welches dem dritten Objekt **453** entspricht, auch mit einer Nicht-Null-Disparität zwischen Pixel-Positionen assoziiert, welche dem linken und rechten Bild entspricht. Die Disparität, welche mit den Pixel-Positionen assoziiert ist, welche dem dritten Objekt **453** entsprechen, ist jedoch kleiner als die Disparität, welche mit den Pixel-Positionen assoziiert ist, welche dem ersten Objekt **451** entsprechen und daher ist die relative Positionierung des ersten Objekts **451** und des dritten Objekts

**453** in dem linken und dem rechten Bild nicht dieselbe.

**[0042]** Wie in **Fig. 4** gezeigt ist, ist die Intensität von Licht, welches auf die Pixel projiziert ist, welche mit Objekten in dem Vordergrund der Szene assoziiert sind, höher in dem zweiten Stereo-Bild-Paar  $F_l$  **411** und  $F_r$  **412** als die Intensität von Licht, welche auf die Pixel projiziert ist, welche mit Objekten in dem Vordergrund der Szene in dem ersten Stereo-Bild-Paar  $G_l$  **401** und  $G_r$  **402** assoziiert sind. Zusätzlich ist die Intensität von Licht, welches auf die Pixel projiziert ist, welche mit Objekten in dem Hintergrund der Szene in dem zweiten Stereo-Bild-Paar  $F_l$  **411** und  $F_r$  **412** assoziiert sind, im Wesentlichen dieselbe wie die Intensität von Licht, welches auf die Pixel projiziert ist, welche mit Objekten in dem Hintergrund der Szene in dem ersten Stereo-Bild-Paar  $G_l$  **401** und  $G_r$  **402** assoziiert sind. Die mittels des Blitz-Geräts **330** bereitgestellte Beleuchtung wird höhere-Intensität-Licht von Objekten näher bei dem Blitz-Gerät **330** reflektieren, wie mittels Gleichung 6 wiedergegeben ist (d.h. die Intensität von Licht, welches auf Beleuchtung von dem Blitz-Gerät bezogen ist, ist invers proportional zu dem Quadrat des Abstandes eines Objektes von dem Blitz-Gerät **330**).

**[0043]** Wie in **Fig. 4** gezeigt ist, zeigen die Verhältnis-Karten ( $R_l$ ,  $R_r$ ), welche dem linken Bild-Sensor **203** und dem rechten Bild-Sensor **202** entsprechen, welche von Blitz-/Nicht-Blitz-Stereo-Bild-Paaren erzeugt sind, welche hierin als  $R_l$  **423** bzw.  $R_r$  **422** bezeichnet werden, größere Unterschiede (d.h. Verhältnis-Werte) zwischen Licht-Intensitäts-Werten für Objekte, welche bei seichteren bzw. geringeren Tiefen lokalisiert sind (d.h. in dem Vordergrund der Szene) als für Objekte bei tieferen Tiefe (d.h. in dem Hintergrund der Szene). Zum Beispiel reflektieren die Verhältnis-Karten  $R_l$  **421** und  $R_r$  **422**, dass Objekt **451** näher an der Stereo-Visions-Kamera **300** ist als Objekt **453**, weil der Unterschied in der Licht-Intensität für Pixel, welche mit Objekt **451** assoziiert sind, größer ist (als eine hellere Farbe wie etwa weiß oder grau angezeigt) als der Unterschied in Licht-Intensität für Pixel, welche mit Objekt **453** assoziiert sind (angezeigt als dunklere Farbe, um kleinere Änderungen in Licht-Intensitäts-Werten zu reflektieren). Ähnlich reflektieren die Verhältnis-Karten  $R_l$  **421** und  $R_r$  **422**, dass das Objekt **452** bei einer Tiefe lokalisiert ist, welche ausreichend groß ist, derart dass die Änderung in der Licht-Intensität, welche von Objekt **452** reflektiert ist, aufgrund der Beleuchtung des Blitz-Geräts **330** vernachlässigbar ist.

**[0044]** Die Verhältnis-Karten können zum Erzeugen eines Disparitäts-Raum-Bildes (disparity base image) (DSI) für jedes Pixel in dem zweiten Stereo-Bild-Paar  $F_l$  **411** und  $F_r$  **412** benutzt werden. Ein Disparitäts-Raum-Bild ist ein Vektor von Kosten-Werten, welcher mit potentiellen Disparitäts-Werten für ein eingege-

benes Pixel  $x$  assoziiert sind. Mit anderen Worten berechnet das Disparitäts-Raum-Bild einen Kosten-Wert, welcher mit einem (Ab)Schätzen des Pixels in einem Bild assoziiert ist, was bei jeder diskreten Tiefen-Stelle zu lokalisieren ist, wobei der Kosten-Wert die Verhältnis-Karten ( $R_l$ ,  $R_r$ ) berücksichtigt. Der Kosten-Wert wird basierend auf der Ähnlichkeit zwischen den Pixeln berechnet, welche innerhalb eines Fensters umfasst sind, welches bei einem Pixel  $x$  in dem linken Blitz-Stereo-Bild  $F_l$  **411** zentriert ist, und den Pixeln, welche innerhalb eines entsprechenden Such-Fensters umfasst sind, welches bei einem disparaten Pixel in dem rechten Blitz-Stereo-Bild  $F_r$  **412** zentriert ist. In einer Ausführungsform ist das Disparitäts-Raum-Bild basierend auf Gleichung 10 berechnet, wie unten ausgeführt ist.

$$C(x, d) = \sum_{|\Delta| < r} \underbrace{N_{\sigma\Delta}(\Delta)}_{1^*} \cdot \underbrace{N_{\sigma R}(dR)}_{2^*} \cdot |dF|^2, \Delta = \langle u, v \rangle \quad (\text{Gl. 10})$$

$$dR = R_l(x + \Delta) - R_l(x) \quad (\text{Gl. 11})$$

$$dF = F_l(x + \Delta) - F_r(x - D + \Delta) \quad (\text{Gl. 12})$$

**[0045]** In Gleichung 10 ist  $\Delta$  ein Versatz bzw. eine Verschiebung innerhalb der Such-Ausdehnungen ( $r$ ) des Pixel-Anpassungs-Fensters,  $r$  ist der maximale Radius der Verschiebung (z.B. für ein 7 Pixel x 7 Pixel-Such-Fenster, ist  $r$  gleich 3),  $dR$  ist der Unterschied in den Verhältnis-Werten für ein Versatz-Pixel ( $x + \Delta$ ) innerhalb des Fensters in dem linken Stereo-Bild  $F_l$  **411** und das gegebene Pixel ( $x$ ) in dem linken Stereo-Bild  $F_l$  **411** und  $dF$  ist der Unterschied in der Intensität von Licht zwischen dem Versatz-Pixel ( $x + \Delta$ ) in dem linken Stereo-Bild  $F_l$  **411** und einem entsprechenden Versatz-Pixel ( $x - D + \Delta$ ) in dem rechten Stereo-Bild  $F_r$  **412** basierend auf dem Such-Fenster. Der erste Koeffizient ( $1^*$ ) ist ein Wert, welcher von einer Gauss-Funktion mit einer Standard-Abweichung ( $\sigma_\Delta$ ) abgeleitet ist, welche von der Größe des Fensters und dem Versatz bzw. Verschiebung ( $\Delta$ ) innerhalb des Fensters abhängt. Ein Beispiel einer Gauss-Funktion für den ersten Koeffizienten ( $1^*$ ) ist in Gleichung 13 unten gegeben, wobei  $u$  und  $v$  Pixel-Einheits-Abstände von der Mitte des Fensters zu dem Pixel sind, welches mit dem Versatz ( $\Delta$ ) assoziiert ist. Es wird geschätzt werden, dass Gleichung 13 für eine gegebene Fenster-Größe und Standard-Abweichung ( $\sigma_\Delta$ ) konstant ist und daher kann eine Matrix zuvor berechnet werden, um eine Mehrzahl von ersten Koeffizienten ( $1^*$ ) zu erzeugen, um auf jedes Pixel in dem Fenster angewendet zu werden.

$$N_{\sigma\Delta}(\Delta) = \frac{1}{2\pi \cdot \sigma_{\Delta}^2} e^{-\frac{u^2+v^2}{\sigma_{\Delta}^2}}, \Delta = \langle u, v \rangle \quad (\text{Gl. 13})$$

**[0046]** Ähnlich ist der zweite Koeffizient (2\*) ein Wert, welcher von einer Gauss-Funktion abgeleitet ist, welche von dem lokalen Bereich der Verhältnis-Karte abhängt, welcher benachbart bzw. nächst zu dem Pixel x ist. Ein Beispiel einer Gauss-Funktion für den zweiten Koeffizienten (2\*) ist unten in Gleichung 14 gegeben, wobei die Standard-Abweichung ( $\sigma_R$ ) eine Funktion ist, welche von dem lokalen Bereich der Verhältnis-Karte um das Pixel x herum abhängt.

$$N_{\sigma R}(dR) = \frac{1}{2\pi \cdot \sigma_R^2} e^{-\frac{dR^2}{\sigma_R^2}} \quad (\text{Gl. 14})$$

**[0047]** Mit anderen Worten bestimmen die Größe des Anpassungs-Fensters und der Bereich der Werte in einem lokalen Bereich der erzeugten Verhältnis-Karte, wie Pixel gewichtet werden, wenn der Kosten-Wert zum Wählen eines Disparitäts-Wertes von D für ein gegebenes Pixel x berechnet wird.

**[0048]** Fig. 5A und Fig. 5B illustrieren einen Pixel-Anpassungs-Algorithmus (pixel-matching algorithm) zum Berechnen eines Disparitäts-Raum-Bildes (DSI), welches mit einem Pixel x 501 assoziiert ist, gemäß einer Ausführungsform der vorliegenden Erfindung. Für Illustrations-Zwecke sind die Blitz-Stereo-Bilder  $F_l$  411 und  $F_r$  412 256-Pixel in Breite in der horizontalen oder x-Dimension und 192 Pixel in der Höhe in der vertikalen oder y-Dimension. Wie in Fig. 5A gezeigt ist, hat ein Pixel x 501 einen Stellen-Index von  $\langle 98 \text{ px}, 52 \text{ px} \rangle$  relativ zu dem Ursprung des Bildes  $F_l$  411, welches in der oberen linken Ecke des Bildes lokalisiert ist. Eine Detail-Ansicht eines 16 Pixel x 16 Pixel-Fensters 500 ist aufgeblasen, um eine bessere Ansicht der dem Pixel 501 benachbarten Pixel in dem linken Blitz-Stereo-Bild  $F_l$  411 bereitzustellen. In einer Ausführungsform ist ein Berechnen des Disparitäts-Raum-Bildes für Pixel x 501 unter Benutzung eines Pixel-Anpassungs-Algorithmus implementiert, wie oben beschrieben ist. Ein Fenster 505 in dem linken Blitz-Stereo-Bild  $F_l$  411 ist in Fig. 5A gezeigt, welches 7 Pixel x 7 Pixel in Größe ist, welches auf Pixel x 501 zentriert ist, wie in einer Ausführungsform des Pixel-Anpassungs-Algorithmus benutzt ist.

**[0049]** Die Größe des Fensters kann basierend auf der verfügbaren Verarbeitungs-Kapazität von Stereo-Visions-Kamera 300 gewählt werden. Größere Fenster-Größen erfordern mehr Berechnungen für jeden Kosten-Wert, welcher mit einem Pixel x 501 und einer gegebenen Disparität D assoziiert ist (d.h.  $C(x, D)$ , wie in Gleichung 10 gezeigt ist). Kleinere Fens-

ter-Größen erfordern weniger Berechnungen für jeden Kosten-Wert. Zum Beispiel umfasst, gegeben dass 7x7-Fenster 505, welches in Fig. 5A illustriert ist, Berechnen eines Kosten-Wertes die Summation von 49 verschiedenen Termen, wobei jeder Term eine Nachschau-Operation zum Bestimmen eines ersten Koeffizienten (1\*) erfordert, welcher von der Größe des Anpassungs-Fensters abhängt, eine Berechnung, welche ein Bestimmen von dR betrifft (was zwei Nachschau-Operationen innerhalb der Verhältnis-Karte  $R_l$  421 erfordert), eine Berechnung zum Bestimmen von dF (was zwei Nachschau-Operationen innerhalb des entsprechenden Blitz-Stereo-Bild-Paares  $F_l$  411 und  $F_r$  412 erfordert), eine Berechnung oder eine Nachschau-Operation zum Bestimmen eines zweiten Koeffizienten, welcher den Bereich der Verhältnis-Karte betrifft, und eine Multiplikations-Operation, um den Term von dem ersten Koeffizienten, dem zweiten Koeffizienten und dem Quadrat von dF zu bestimmen. Obwohl die Kosten-Wert-Berechnung aufwendig sein kann bezüglich der Anzahl von Berechnungen, wird es geschätzt werden, dass eine zu starke Verminderung der Fenster-Größe zu weniger genauen Disparitäts-Raum-Bildern führen kann.

**[0050]** Wie in Fig. 5B gezeigt ist, zeigt eine Box 515 die maximalen Such-Ausdehnungen zum Übereinstimmung bringen bzw. Anpassen der Pixel von Fenster 505 in dem linken Blitz-Stereo-Bild  $F_l$  411 mit den Pixeln in einer Mehrzahl von möglichen Such-Fenstern (z.B. Fenster 510) in dem rechten Blitz-Stereo-Bild  $F_r$  412. Jedes mögliche Such-Fenster ist von der entsprechenden Stelle für Pixel x 501 um eine Disparität D versetzt bzw. verschoben. Es wird geschätzt werden, dass für irgendein Pixel x 501 in dem linken Blitz-Stereo-Bild  $F_l$  411 das entsprechende dispartate Pixel (x-D) in dem rechten Blitz-Stereo-Bild  $F_r$  412 bei derselben Stelle wie Pixel x 501 lokalisiert sein sollte oder einen gewissen Abstand D links von dem Pixel x 501. In einer Ausführungsform können die Such-Ausdehnungen auf eine gewisse Maximum-Menge von Disparität  $D_{\max}$  begrenzt werden, wie etwa 64 Pixel, welche auf eine minimale Tiefe  $Z_{\min}$  für irgendein Pixel in dem Stereo-Bild-Paar  $F_l$  411 und  $F_r$  412 bezogen ist. In einer anderen Ausführungsform können die Such-Ausdehnungen unbegrenzt sein und der Pixel-Anpassungs-Algorithmus kann den Kosten-Wert für jedes Pixel links von der Pixel-Stelle berechnen, welche dem Pixel x 501 entspricht, bis zu und einschließlich der Pixel bei der Kante des rechten Blitz-Stereo-Bildes  $F_r$  412. In noch einer anderen Ausführungsform können die Such-Ausdehnungen gemäß dem Verhältnis-Wert ( $R_l$ ) für Pixel x 501 gesetzt werden. Zum Beispiel können kleine Verhältnis-Werte anzeigen, dass die Oberfläche für Pixel x 501 bei einer genügend großen Tiefe lokalisiert ist und dass die Disparität D für das entsprechende übereinstimmende Pixel erwartet ist, klein zu sein.

**[0051]** Sobald das Disparitäts-Raum-Bild für Pixel  $x$  **501** berechnet ist, wird eine Schätzung für eine Disparität  $D$  für Pixel  $x$  **501** bestimmt. In einer Ausführungsform implementiert der Pixel-Anpassungs-Algorithmus eine Gewinner-nimmt-Alles (WTA)-Optimierung, welche die abgeschätzte Disparität für Pixel  $x$  **501** dadurch bestimmt, dass der minimale Kosten-Wert in dem Disparitäts-Form-Bild gefunden ist, welches mit Pixel  $x$  **501** assoziiert ist, wie durch Gleichung 15 illustriert ist. Die ausgewählte Disparität  $D$  für Pixel  $x$  **501** ist der Wert  $D$ , welcher die Kosten-Funktion (d.h. Gleichung 10) innerhalb der Such-Ausdehnungen des Fensters **515** minimiert. In alternativen Ausführungsformen können höhere-Qualitäts-Global-Verfahren zum Abschätzen eines Disparitäts-Wertes  $D$  für Pixel implementiert werden, wie etwa ein Graph-Abschneide-Zugang (Graph-cut approach) oder andere Algorithmen, welche für die Fachleute in der Technik bekannt sind. Ein Berechnen der Disparität  $D$  für jedes Pixel des Blitz-Stereo-Bildes erzeugt eine Disparitäts-Karte ( $D_l$ ), welche benutzt werden kann, um eine Tiefe-Karte  $Z_l$  für jedes Pixel des Blitz-Stereo-Bildes zu erzeugen, gemäß Gleichungen 3 oder 4, welche oben ausgeführt sind.

$$D(x) = \min C(x, D) \quad (\text{Gl. 15})$$

**[0052]** In einer Ausführungsform kann, um eine Tiefe  $Z$  für verdeckte Oberflächen abzuschätzen, eine Links-Rechts-Konsistenz-(LRC)-Technik auch implementiert werden. In der LRC-Technik wird eine Disparitäts-Karte ( $D_l$ ) für jedes Pixel des linken Blitz-Stereo-Bildes  $F_l$  **411** wie oben beschrieben erzeugt. Eine ähnliche Operation wird für Pixel des rechten Blitz-Stereo-Bildes  $F_r$  **412** durchgeführt, um eine entsprechende Disparitäts-Karte ( $D_r$ ) zu erzeugen, außer dass die Terme für  $dR$  und  $dF$  von Gleichung 10 unter Benutzung von Gleichungen 16 und 17 anstatt der Gleichungen 11 und 12 berechnet werden, wenn die Disparitäts-Karte  $D_r$  für das rechte Blitz-Stereo-Bild  $F_r$  **412** erzeugt wird. Ähnlich werden sich Such-Ausdehnungen nach rechts in dem linken Blitz-Stereo-Bild  $F_l$  **411** anstatt nach links wie in dem rechten Blitz-Stereo-Bild  $F_r$  **412** ausdehnen bzw. erstrecken.

$$dR = R_r(x + \Delta) - R_r(x) \quad (\text{Gl. 16})$$

$$dF = F_r(x + \Delta) - F_r(x + D + \Delta) \quad (\text{Gl. 17})$$

**[0053]** In der LRC-Technik sollte die abgeschätzte Disparität für ein Pixel in dem linken Blitz-Stereo-Bild  $F_l$  **411** das Umgekehrte bzw. das Gegensätzliche der abgeschätzten Disparität für das entsprechende Pixel in dem rechten Blitz-Stereo-Bild  $F_r$  **412** sein, wie durch Gleichung 18 illustriert ist. Wenn  $D_l$  und  $D_r$  exakt bzw. präzise für ein Pixel in einem Stereo-Bild und das entsprechende Pixel in dem ande-

ren Stereo-Bild abgeschätzt sind, dann sollte sich ein Disparitäts-Fehler  $E(x)$  Null nähern. In einer Ausführungsform, wenn der Disparitäts-Fehler  $E(x)$  größer als eine Schwellwert-Grenze ist, dann kann die abgeschätzte Disparität  $D(x)$  für ein Pixel  $x$  **501** als unzuverlässig (oder verdeckt) gekennzeichnet werden. Für irgend solche Pixel kann die abgeschätzte Disparität als ein Mittelwert der abgeschätzten Disparität von dem linken Blitz-Stereo-Bild  $F_l$  **411** und die abgeschätzte Disparität von dem rechten Blitz-Stereo-Bild  $F_r$  **412** berechnet werden. In anderen Ausführungsformen wird ein verdecktes Pixel eine Disparität von einem benachbarten Pixel mit einem ähnlichen Verhältnis-Wert kopieren.

$$E(x) = |D_l(x) + D_r(x + D_l(x))| \quad (\text{Gl. 18})$$

**[0054]** Herkömmliche Stereo-Pixel-Anpassungs-Algorithmen können damit Schwierigkeiten haben, eine exakte Disparität ( $D(x)$ ) für Pixel abzuschätzen, welche nächst bzw. benachbart zu Tiefen-Diskontinuitäten sind. Wie Fenster **505** in Fig. 5A zeigt, sind einige der Pixel in Fenster **505** mit einer Oberfläche des dritten Objekts **453** assoziiert, welches hinter dem ersten Objekt **451** in Bezug auf den linken Bild-Sensor **201** lokalisiert ist. Ein Such-Fenster **510**, welches auf einem entsprechenden disparaten Pixel in dem rechten Bild-Stereo-Bild  $F_r$  **512** zentriert ist, braucht jedoch nicht irgendwelche Pixel umfassen, welche mit dem dritten Objekt **453** assoziiert sind, aufgrund der relativen Position der Objekte bezüglich des rechten Bild-Sensors **302**. Herkömmliche Pixel-Anpassungs-Algorithmen können daran scheitern, zwischen den Pixeln von verschiedenen Objekten zu unterscheiden, welche bei verschiedenen Tiefen lokalisiert sind und daher kann der herkömmliche Algorithmus bestimmen, dass Fenster **505** nicht ähnlich zu Fenster **510** in dem entsprechenden Stereo-Bild ist und die abgeschätzte Disparität  $D$  basierend auf einem verschiedenen Such-Fenster innerhalb der Such-Ausdehnungen, kann inkorrekt sein. Mittels eines Benutzens der Verhältnis-Karten, um Pixel zu gewichten, welche mit Oberflächen bei verschiedenen Tiefen assoziiert sind, erzeugt der hierin beschriebene Blitz-/Nicht-Blitz-Pixel-Anpassungs-Algorithmus eine genauere Disparitäts-Abschätzung für jedes Pixel eines Stereo-Bild-Paars. Ferner erfordert der Blitz-/Nicht-Blitz-Pixel-Anpassungs-Algorithmus keine präzise Kalibrierung des Blitz-Gerätes **330**, weil die Verhältnis-Karte benutzt ist, Pixel für Anpassungs-Prozesse zu filtern und nicht für ein direktes Berechnen einer Disparitäts-Schätzung oder eines Tiefe-Wertes.

**[0055]** In einer Ausführungsform können Stereo-Bild-Paare als Teil eines stereoskopischen Videos erfasst werden. In solch einer Ausführungsform kann die Bewegung von Objekten in der Szene dadurch abgeschätzt werden, dass aufeinander folgende Bil-

der (frames) eines stereoskopischen Videos analysiert werden, um abgeschätzte bzw. geschätzte Bewegungs-Vektoren zu erzeugen, welche mit den verschiedenen Objekten (d.h. Setzen von betreffenden Pixeln) in einer Szene assoziiert sind. Der oben beschriebene Pixel-Anpassungs-Algorithmus kann dadurch verbessert werden, dass die geschätzten Bewegungs-Vektoren angewendet werden, um kleine Versetzungen bzw. Verschiebungen von Objekten zu korrigieren, welche zwischen dem Punkt auftreten, an dem Nicht-Blitz-Stereo-Bild-Paar erfasst ist, und dem Punkt, an dem das Blitz-Stereo-Bild-Paar erfasst ist. Zum Beispiel können die Verhältnis-Werte basierend auf Pixeln in dem Nicht-Blitz-Stereo-Bild (d.h. **401**, **402**) berechnet werden, welche von der gegebenen Pixel-Position in dem Blitz-Stereo-Bild (d.h. **411**, **412**) um den geschätzten Bewegungs-Vektor versetzt sind.

Filtern, um Tiefen-Quantisierung zu vermindern

**[0056]** Die Disparitäts-Karte  $D(x)$ , welche unter Benutzung des Blitz-/Nicht-Blitz-Pixel-Anpassungs-Algorithmus, während er genauer ist als Disparitäts-Karten, welcher unter Benutzung von konventionellen Pixel-Anpassungs-Algorithmen erzeugt sind, erzeugt ist, kann noch eine schwache Präzision für gewisse Pixel basierend auf lokalen Unregelmäßigkeiten haben. Sowohl Disparitäts-Werte als auch Verhältnis-Werte sind jedoch typischerweise lokal linear für irgendeine ebene Oberfläche. In anderen Worten werden benachbarte bzw. nächst angeordnete Punkte auf derselben Oberfläche typischerweise bei ungefähr derselben Tiefe lokalisiert sein und sollten daher ähnliche Disparitäts- und Verhältnis-Werte haben. Als ein Ergebnis wird ein Filtern der Disparitäts-Karten, welche mit dem Blitz-/Nicht-Blitz-Pixel-Anpassungs-Algorithmus erzeugt sind, eine robustere Tiefe-Karte des Blitz-Stereo-Bild-Paares ( $F_l$ ,  $F_r$ ) erzeugen.

**[0057]** Fig. 6A und Fig. 6B illustrieren einen iterativen Filterungs-Algorithmus zum Erzeugen robusterer Tiefe-Karten gemäß einer Ausführungsform der vorliegenden Erfindung. Fig. 6A zeigt eine Disparitäts-Karte **600-0** für eine Szene, welche eine Hand einer Person umfasst, welche gemäß dem Blitz-/Nicht-Blitz-Pixel-Anpassungs-Algorithmus berechnet ist, welcher oben ausgeführt ist. Wie Fig. 6A jedoch ersichtlich macht, umfasst die Disparitäts-Karte **600-0** lokale Inkonsistenzen (z.B. **601**), wo der Blitz-/Nicht-Blitz-Pixel-Anpassungs-Algorithmus anscheinend inkorrekte Disparitäts-Werte für bestimmte Pixel hat. In einer Ausführungsform kann die Stereo-Visions-Kamera **300** die ursprüngliche Disparitäts-Karte **600-0** basierend, zumindest im Teil, auf einer Verhältnis-Karte ( $R_l$ ,  $R_r$ ) filtern, um eine konsistentere und genauere Disparitäts-Karte bereitzustellen, welche mit dem Blitz-Stereo-Bild-Paar ( $F_l$ ,  $F_r$ ) assoziiert ist. Gleichung 19, welche unten gezeigt

ist, ist implementiert, um Disparitäts-Werte für ein Pixel  $x$  zu filtern, wobei  $\Omega$  die Summe aller Gewichte  $W(x, \Delta)$  ist und wobei  $\tau$  auf 2 gesetzt ist (d.h. ein 5 Pixel  $x$  5 Pixel-Fenster), gemäß einer Beispiel-Ausführungsform.

$$D^{i+1}(x) = \frac{1}{\Omega} \sum_{|\Delta| < \tau} W(x + \Delta) \cdot D^i(x + \Delta) \quad (\text{Gl. 19})$$

**[0058]** Mit anderen Worten ist jedes Pixel der gefilterten Disparitäts-Karte von einer gewichtigen Summe von benachbarten Pixel in der ursprünglichen Disparitäts-Karte **600-0** berechnet. Die Gewichte für jede Pixel-Position relativ zu einem Pixel  $x$  sind durch Gleichungen 20, 21 und 22 gegeben, welche unten gezeigt sind.

$$W(x, \Delta) = N_{\sigma_R}(\delta F(x)) \cdot N_{\sigma_D}(\delta D^i(x)) \cdot e^{-C(x+\Delta)} \quad (\text{Gl. 20})$$

$$\delta F(x) = F(x + \Delta) - F(x) \quad (\text{Gl. 21})$$

$$\delta D^i(x) = D^i(x + \Delta) - D^i(x) \quad (\text{Gl. 22})$$

**[0059]** Der Exponenten-Term ( $e^{-C(x+\Delta)}$ ) ist das Konfidenz-Niveau des Anpassungs-Algorithmus und ist abhängig von dem Kosten-Wert, welcher durch Gleichung 15 gegeben ist, für ein bestimmtes Pixel  $x$  und geschätzte Disparität  $D$ . Mit anderen Worten führt ein niedriger Kosten-Wert (d.h. gute Anpassungs-Konfidenz) für ein gegebenes Pixel  $x$  zu einem größeren Gewicht als ein Pixel mit einem hohen Kosten-Wert (d.h. schwache Anpassungs-Konfidenz). Die grundsätzliche Idee der Gewichtungsberechnung ist, dass die Tiefe-Information von Hohe-Konfidenz-Pixeln zu Niedrig-Konfidenz-Pixeln fließen sollte. Wieder sind die  $N_{\sigma_R}$ - und  $N_{\sigma_D}$ -Koeffizienten durch Gauss-Funktionen ähnlich zu Gleichungen 13 und 14 erzeugt, wie oben ausgeführt ist. Anwenden von Gleichung 19 auf die Disparitäts-Karte **600-0** erzeugt eine neue gefilterte Disparitäts-Karte. Die Filter-Funktion von Gleichung 19 kann wiederholt über eine Anzahl  $N$  von Iterationen angewendet werden, um gefilterte Resultate zu erzeugen, wie etwa Disparitäts-Karte **600-30** (30 mal iteriert), wie in Fig. 6B gezeigt ist. Der iterative Filterungs-Algorithmus entfernt effektiv einzelne Pixel-Diskontinuitäten, welche nicht mit benachbarten Pixeln übereinstimmen.

**[0060]** Fig. 7 ist ein Flussdiagramm von Verfahrensschritten zum Erzeugen einer Disparitäts-Karte, welche mit einem Stereo-Bild assoziiert ist, gemäß einer Ausführungsform der vorliegenden Erfindung. Obwohl die Verfahrensschritte im Zusammenhang mit den Systemen von Fig. 1-6 beschrieben sind, wer-

den gewöhnliche Fachleute in der Technik verstehen, dass irgendein System, welches konfiguriert ist, die Verfahrensschritte durchzuführen, innerhalb des Geltungsbereichs der vorliegenden Erfindung ist.

**[0061]** Wie gezeigt ist, beginnt Verfahren **700** bei Schritt **710**, wobei Stereo-Visions-Kamera **300** ein Stereo-Bild-Paar  $G_l$  **401**,  $G_r$  **402** ohne die Benutzung einer Blitz-Beleuchtung von einem Blitz-Gerät **330** erfasst. Bald danach erfasst bei Schritt **712** die Stereo-Visions-Kamera **300** ein Stereo-Bild-Paar  $F_l$  **411**,  $F_r$  **412** unter Benutzung von Blitz-Beleuchtung von dem Blitz-Gerät **330**. In einigen Ausführungsformen können mehr als zwei Bild-Sensoren implementiert sein, um einen n-Tuple-Satz von Bildern mit und ohne Blitz-Beleuchtung zu erfassen.

**[0062]** Bei Schritt **714** erzeugt die Stereo-Visions-Kamera **300** einen Satz von Verhältnis-Karten, welche mit den Blitz-Stereo-Bildern assoziiert sind (z.B.  $F_l$  **411**,  $F_r$  **412**), welche mit Blitz-Beleuchtung erfasst sind. Für jedes Bild in dem Satz von Blitz-Stereo-Bildern ist eine Verhältnis-Karte als der Logarithmus des Verhältnisses von Licht-Intensitäts-Werten für Pixel in dem Blitz-Stereo-Bild zu den Licht-Intensitäts-Werten für Pixel in dem entsprechenden Nicht-Blitz-Stereo-Bild definiert. In einer Ausführungsform ist die Verhältnis-Karte gemäß Gleichung 9 definiert, wie oben ausgeführt ist. Die Verhältnis-Karte stellt eine grobe Indikation von Tiefen für jedes Pixel in den Blitz-Stereo-Bildern bereit, wobei Pixel bei ähnlichen Tiefen und Orientierungen auf ähnlichen Oberflächen ähnliche Verhältnis-Werte haben.

**[0063]** Bei Schritt **716** erzeugt die Stereo-Visions-Kamera **300** einen Satz von Disparitäts-Karten, welche mit den Blitz-Stereo-Bildern assoziiert sind, und die entsprechenden Verhältnis-Karten dadurch, dass ein Blitz-/Nicht-Blitz-Pixel-Anpassungs-Algorithmus implementiert ist. Für jedes Bild in dem Satz von Blitz-Stereo-Bildern (z.B.  $F_l$  **411**,  $F_r$  **412**) ist eine Disparitäts-Karte als der Satz von geschätzten Disparitäts-Werten ( $D(x)$ ) definiert, welche mit den Pixeln in dem Blitz-Stereo-Bild assoziiert sind. In einer Ausführungsform entspricht jeder Disparitäts-Wert für ein Pixel mit den minimalen Kosten für einen Bereich von möglichen Disparitäts-Werten für das gegebene Pixel. Die minimalen Kosten können mittels eines Lösens von Gleichung 15 abgeleitet werden, wie oben ausgeführt ist.

**[0064]** Bei Schritt **718** erzeugt die Stereo-Visions-Kamera **300** eine Tiefe-Karte basierend auf zumindest einer der Disparitäts-Karten. In einer Ausführungsform kann die Tiefe-Karte mittels eines Lösens von Gleichungen 3 oder 4, welche oben ausgeführt sind, berechnet werden, wobei eine bekannte Fokus-Länge und Grundlinie-Versetzung bzw. -Verschiebung für die Stereo-Visions-Kamera **300** gegeben ist. Dann terminiert das Verfahren **700**.

**[0065]** Fig. 8 ist ein Flussdiagramm von Verfahrensschritten zum Erzeugen einer gefilterten Disparitäts-Karte, welche mit einem Blitz-Stereo-Bild assoziiert ist, gemäß einer Ausführungsform der vorliegenden Erfindung. Obwohl die Verfahrensschritte im Zusammenhang mit den Systemen von Fig. 1-6 beschrieben sind, werden gewöhnliche Fachleute in der Technik verstehen, dass irgendein System, welches konfiguriert ist, die Verfahrensschritte auszuführen, in irgendeiner Ordnung, innerhalb des Geltungsbereichs der vorliegenden Erfindung ist.

**[0066]** Wie gezeigt ist, beginnt das Verfahren **800** bei Schritt **810**, wobei eine Disparitäts-Karte für ein Blitz-Stereo-Bild erzeugt ist. Zum Beispiel kann die Disparitäts-Karte mittels des Verfahrens **700** erzeugt sein. Bei Schritt **812** filtert die Stereo-Visions-Kamera **300** die Disparitäts-Karte, um eine gefilterte Disparitäts-Karte zu erzeugen. In einer Ausführungsform implementiert die Stereo-Visions-Kamera **300** die Filter-Funktion gemäß Gleichung 19, welche oben ausgeführt ist, um die ursprüngliche Disparitäts-Karte zu filtern. Bei Schritt **814** bestimmt die Stereo-Visions-Kamera **300**, ob die Anzahl von vollendeten Iterationen  $i$  größer ist oder gleich zu der Schwellwert-N-Anzahl von Iterationen. Wenn  $i$  kleiner als  $N$  ist, dann kehrt das Verfahren **800** zu Schritt **812** zurück, wo die Stereo-Visions-Kamera **300** eine zusätzliche Iteration dadurch durchführt, dass die gefilterte Disparitäts-Karte, welche während der vorherigen Iteration erzeugt ist, gefiltert ist und fügt 1 zu  $i$  hinzu. Wenn  $i$  jedoch größer ist als  $N$  oder gleich  $N$  ist, dann terminiert Verfahren **800**.

**[0067]** Zusammenfassend wird eine robuste Tiefe-Karte von einem Satz von Stereo-Bildern erzeugt, welche mit und ohne Blitz-Beleuchtung erfasst sind. Die Tiefe-Karte ist robuster als Tiefe-Karten, welche unter Benutzung von herkömmlichen Techniken erzeugt sind, weil ein Pixel-Anpassungs-Algorithmus implementiert ist, welcher Pixel in einem Anpassungs-Fenster gemäß des Verhältnisses von Licht-Intensität gewichtet, welche unter Benutzung von verschiedener Blitz-Beleuchtung erfasst ist. Die Verhältnis-Karte stellt eine grobe Schätzung von Tiefe relativ zu benachbarten Pixeln bereit, was dem Pixel-Anpassungs-Algorithmus erlaubt, Pixel herunter zu werten, welche bei verschiedenen Tiefen als das zentrale Pixel in dem Anpassungs-Fenster lokalisiert erscheinen. Ein Grund, warum die Technik robust bei Tiefen-Diskontinuitäten ist, ist, dass das Verhältnis-Bild eine Anpassungs-Beschränkung bereitstellt, ohne direkt benachbarte Pixel innerhalb des Anpassungs-Fensters zu involvieren. Traditionelle Techniken stützen sich nur auf Intensitäts-Variationen innerhalb des Anpassungs-Fensters, welche nicht so robust sind, wenn einige Pixel innerhalb des Anpassungs-Fensters mit verschiedenen Hintergrund-Objekten assoziiert sind. Ferner kann die Verhältnis-Karte benutzt werden, um die erzeugte Disparitäts-Karte zu filtern,

um eine glatte Schätzung (smooth estimate) von Tiefe von Objekten innerhalb eines Stereo-Bildes zu erzeugen.

**[0068]** Ein Vorteil des offenbarten Systems ist, dass die Stereo-Visions-Kamera keinen hohen Grad an Kalibrierung erfordert, um robuste Tiefe-Karten zu erzeugen. Die hierin beschriebene Technik kann in vielen kostengünstigen handgehaltenen Geräten implementiert werden. Ein einfaches Vergleichen des Verhältnisses von erfasster Licht-Intensität unter Benutzung von verschiedenen Graden von Beleuchtung stellt Tiefen-Information bereit, welche benutzt werden kann, um effektiv zwischen Pixeln zu differenzieren, welche bei verschiedenen Tiefen in den Pixel-Anpassungs-Algorithmen lokalisiert sind.

**[0069]** Eine Ausführungsform der Erfindung kann als ein Programm-Produkt zur Benutzung mit einer Computer-System implementiert sein. Das Programm oder die Programme des Programm-Produkts definieren Funktionen der Ausführungsformen (einschließlich der hierin beschriebenen Verfahren) und können auf einer Verschiedenheit von Computer-lesbaren Speichermedien beinhaltet sein. Illustrative Computer-lesbare Speichermedien umfassen, sind jedoch nicht darauf beschränkt: (i) nicht-schreibbare Speichermedien, z.B. Nur-Lese-Speicher-Geräte innerhalb eines Computers (wie CD-ROM-Platten, welche mittels eines CD-ROM-Laufwerks lesbar sind, Flash-Speicher, ROM-Chips oder irgendein anderer Typ von Festkörper-nicht-volatilen Halbleiter-Speicher), auf welchen Informationen permanent gespeichert ist; und (ii) schreibbare Speichermedien (z.B. Floppy-Disks innerhalb eines Disketten-Laufwerks oder eines Festplatten-Laufwerks oder irgendein anderer Typ von Festkörper-Halbleiter-Speicher mit willkürlichem Zugriff), auf welchen veränderbare Informationen gespeichert ist.

### Patentansprüche

1. Computer-implementiertes Verfahren zum Erzeugen von Tiefenkarten aus einem Satz von Stereo-Bildern, welche mittels zwei oder mehr Bild-Sensoren erfasst sind, umfassend ein erstes Bild, welches mittels eines ersten Bild-Sensors unter Benutzung von Umgebungs-Licht-Beleuchtung erfasst ist, ein zweites Bild, welches mittels des ersten Bild-Sensors unter Benutzung von Blitz-Beleuchtung erfasst ist, ein drittes Bild, welches mittels eines zweiten Bild-Sensors im Wesentlichen simultan zu der Erfassung des ersten Bildes mittels des ersten Bild-Sensors erfasst wird, und ein viertes Bild, welches mittels des zweiten Bild-Sensors im Wesentlichen simultan zu der Erfassung des zweiten Bildes mittels des ersten Bild-Sensors erfasst ist, wobei das Verfahren aufweist:  
Erzeugen einer Verhältniskarte, welche dem ersten Bild und dem zweiten Bild in dem Satz von Stereobildern zugeordnet ist, wobei die Verhältniskarte für je-

des Pixel in dem zweiten Bild dem Pixel ein Verhältnis eines ersten Intensitätswertes für das Pixel basierend auf dem zweiten Bild zu einem zweiten Intensitätswert für ein entsprechendes Pixel basierend auf dem ersten Bild zuordnet;

Erzeugen einer Disparitätskarte, die dem zweiten Bild zugeordnet ist, basierend auf einem Vergleich des zweiten Bildes und des vierten Bildes in dem Satz von Stereobildern, wobei die Disparitätskarte für jedes Pixel in dem zweiten Bild das Pixel einem Disparitätswert gemäß einem Pixel-Matching-Algorithmus zuordnet, wobei der Pixel-Matching-Algorithmus Pixel in dem zweiten Bild mit entsprechenden Pixeln in dem vierten Bild vergleicht und eine Berechnung umfasst, die auf dem Verhältnis basiert, das dem Pixel zugeordnet ist, wobei die Pixel, die bei verschiedenen Tiefen angeordnet sind, im Matching-Algorithmus in einem Matching-Fenster mit der Verhältniskarte gewichtet werden;

Erzeugen einer Tiefenkarte, die dem zweiten Bild zugeordnet ist, wobei die Tiefenkarte für jedes Pixel im zweiten Bild das Pixel einem Tiefenwert zuordnet, der basierend auf dem Pixel zugeordneten Disparitätswert, einer Fokusslänge und einem Basislinienversatzabstand berechnet wird.

2. Verfahren gemäß Anspruch 1, wobei, für jedes Pixel in dem zweiten Bild, Erzeugen der Disparitätskarte aufweist:

Berechnen zumindest eines Kostenwert zur Bestimmung des Disparitätswertes, wobei der Kostenwert teilweise auf einer Differenz zwischen Werten eines dem Pixel zugeordneten Verhältnisses und einem Verhältnis basiert, das einem versetzten Pixel in dem zweiten Bild zugeordnet ist, das von dem Pixel versetzt ist.

3. Verfahren gemäß Anspruch 1, wobei, für jedes Pixel in dem zweiten Bild, Erzeugen der Disparitätskarte aufweist:

Erzeugen eines Disparitäts-Raum-Bildes, welches eine Mehrzahl von Kostenwerten umfasst, welche für eine Mehrzahl von potentiellen Disparitätswerten berechnet sind; und

Auswählen eines bestimmten Disparitätswertes von der Mehrzahl von potentiellen Disparitätswerten basierend auf der Mehrzahl von Kostenwerten.

4. Verfahren gemäß Anspruch 3, wobei Auswählen des bestimmten Disparitätswertes Auswählen, von der Mehrzahl von potentiellen Disparitätswerten, eines potentiellen Disparitätswertes aufweist, welcher mit einem minimalen Kostenwert in dem Disparitäts-Raum-Bild assoziiert ist.

5. Verfahren gemäß Anspruch 1, ferner aufweisend:

Filtern der Disparitätskarte, um eine gefilterte Disparitätskarte zu erzeugen, welche mit dem zweiten Bild assoziiert ist,

wobei, für jedes Pixel in dem zweiten Bild, die gefilterte Disparitätskarte einen gefilterten Disparitätswert mit dem Pixel assoziiert, wobei der gefilterte Disparitätswert dadurch erzeugt ist, dass eine gewichtete Summe eines Satzes von Disparitätswerten in der Disparitätskarte berechnet wird, welcher mit einem Fenster von Pixeln assoziiert sind, welche nahe bzw. nächst bzw. benachbart zu dem Pixel sind, und wobei Gewichtungskoeffizienten für jeden der Disparitätswerte in dem Satz von Disparitätswerten auf einen Konfidenz-Level des Disparitätswertes bezogen sind, welcher mit diesem Gewichtungskoeffizient assoziiert ist, welcher zumindest zum Teil basiert ist auf dem Verhältnis, welches mit dem Pixel assoziiert ist.

6. Verfahren gemäß Anspruch 5, ferner aufweisend, für eine Anzahl N von Iterationen, Wiederholen des Schrittes eines Filterns, um eine Anzahl N von sukzessiven gefilterten Disparitätskarten zu erzeugen, wobei jede sukzessive gefilterte Disparitätskarte, welche während einer momentanen Iteration erzeugt ist, dadurch erzeugt ist, dass eine sukzessive gefilterte Disparitätskarte gefiltert wird, welche während einer vorherigen Iteration erzeugt ist.

7. Verfahren gemäß Anspruch 1, ferner aufweisend:  
Erzeugen einer Verhältniskarte, welche mit dem vierten Bild assoziiert ist;  
Erzeugen einer Disparitätskarte, welche mit dem vierten Bild assoziiert ist; und  
Einstellen der Tiefenkarte, welche mit dem zweiten Bild assoziiert ist, basierend auf der Disparitätskarte, welche mit dem zweiten Bild assoziiert ist, und der Disparitätskarte, welche mit dem vierten Bild assoziiert ist.

8. Verfahren gemäß Anspruch 7, ferner aufweisend:  
Berechnen eines Disparitäts-Fehlers für ein bestimmtes Pixel in dem zweiten Bild basierend auf einem Unterschied zwischen dem Disparitätswert, welcher mit dem bestimmten Pixel in der Disparitätskarte assoziiert ist, welche mit dem zweiten Bild assoziiert ist, und den Disparitätswert, welcher mit einem entsprechenden Pixel in der Disparitätskarte assoziiert ist, welche mit dem vierten Bild assoziiert ist;  
Bestimmen, dass der Disparitäts-Fehler für das bestimmte Pixel oberhalb eines Schwellwertes ist; und  
Kennzeichnen des bestimmten Pixels als verdeckt.

9. System zum Erzeugen von Tiefenkarten aus einem Satz von Stereo-Bildern, wobei das System aufweist:  
einen Speicher, welcher konfiguriert ist, den Satz von Stereo-Bildern zu speichern, welche mittels zwei oder mehr Bild-Sensoren erfasst sind, wobei der Satz von Stereo-Bildern ein erstes Bild, welches mittels eines ersten Bild-Sensors unter Benutzung von Um-

gebungs-Licht-Beleuchtung erfasst wird, ein zweites Bild, welches mittels des ersten Bild-Sensors unter Benutzung von Blitz-Beleuchtung erfasst wird, ein drittes Bild, welches mittels eines zweiten Bild-Sensors im Wesentlichen simultan erfasst wird während der Erfassung des ersten Bildes mittels des ersten Bild-Sensors, und ein viertes Bild umfasst, welches mittels des zweiten Bild-Sensors im Wesentlichen simultan erfasst wird während der Erfassung des zweiten Bildes mittels des ersten Bild-Sensors; und  
einen Prozessor, welcher mit dem Speicher gekoppelt ist und konfiguriert ist, zum:

Erzeugen einer Verhältniskarte, welche dem ersten Bild und dem zweiten Bild in dem Satz von Stereobildern zugeordnet ist, wobei die Verhältniskarte für jedes Pixel in dem zweiten Bild dem Pixel ein Verhältnis eines ersten Intensitätswertes für das Pixel basierend auf dem zweiten Bild zu einem zweiten Intensitätswert für ein entsprechendes Pixel basierend auf dem ersten Bild zuordnet;

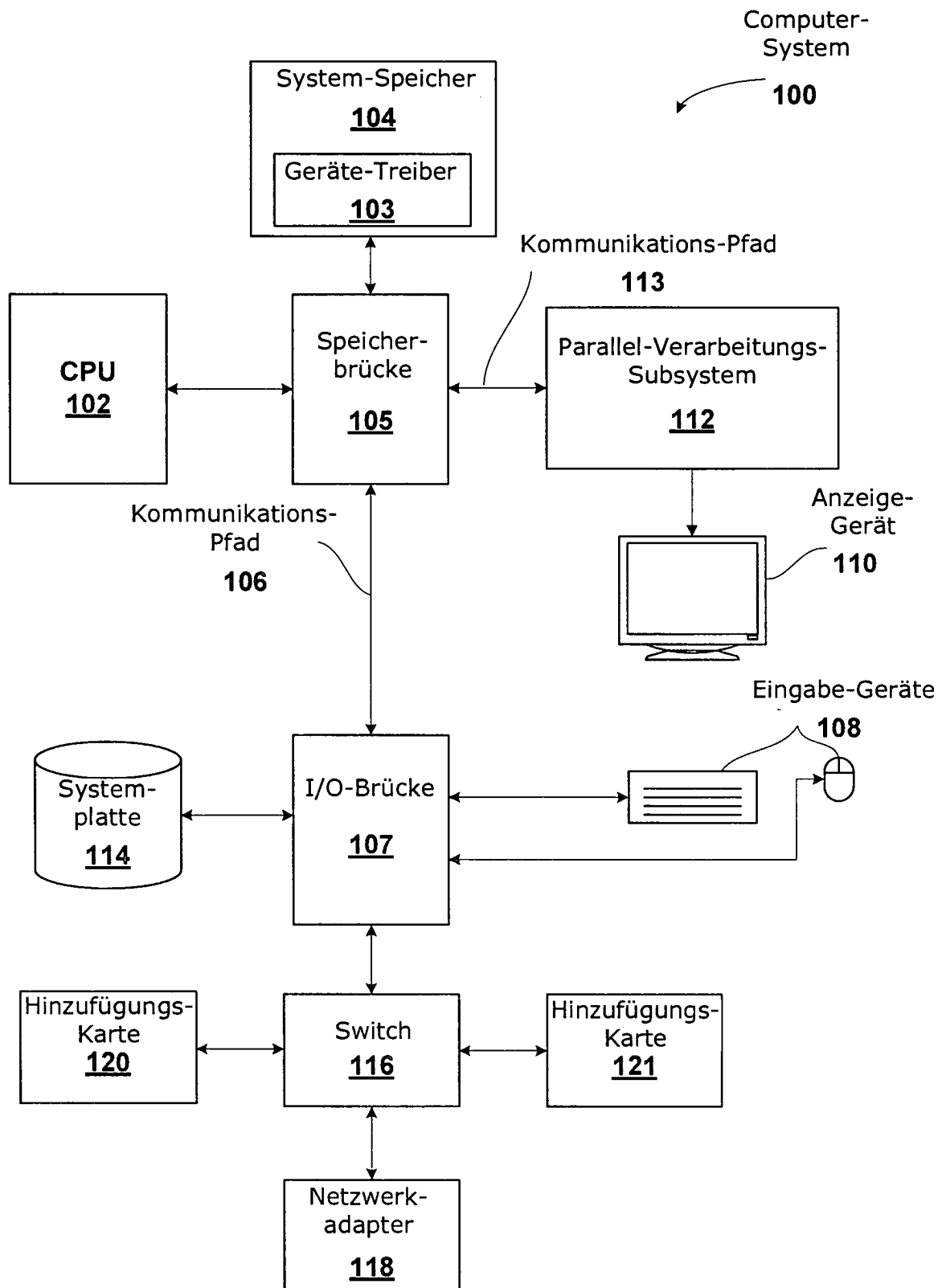
Erzeugen einer Disparitätskarte, die dem zweiten Bild zugeordnet ist, basierend auf einem Vergleich des zweiten Bildes und des vierten Bildes in dem Satz von Stereobildern, wobei die Disparitätskarte für jedes Pixel in dem zweiten Bild das Pixel einem Disparitätswert gemäß einem Pixel-Matching-Algorithmus zuordnet, wobei der Pixel-Matching-Algorithmus Pixel in dem zweiten Bild mit entsprechenden Pixeln in dem vierten Bild vergleicht und eine Berechnung umfasst, die auf dem Verhältnis basiert, das dem Pixel zugeordnet ist, wobei die Pixel, die bei verschiedenen Tiefen angeordnet sind, im Matching-Algorithmus in einem Matching-Fenster mit der Verhältniskarte gewichtet werden;

Erzeugen einer Tiefenkarte, die dem zweiten Bild zugeordnet ist, wobei die Tiefenkarte für jedes Pixel im zweiten Bild das Pixel einem Tiefenwert zuordnet, der basierend auf dem dem Pixel zugeordneten Disparitätswert, einer Fokusslänge und einem Basislinienversatzabstand berechnet wird.

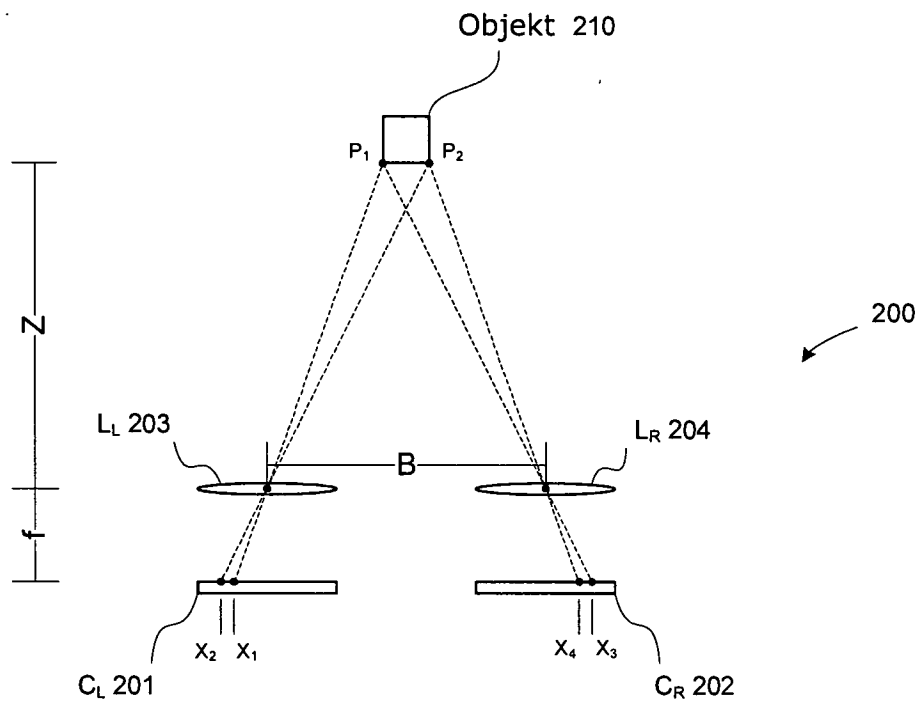
10. System gemäß Anspruch 9, ferner aufweisend:  
eine Stereo-Visionskamera, welche mit dem Speicher und dem Prozessor gekoppelt ist, wobei die Stereo-Visionskamera den ersten Bild-Sensor, den zweiten Bild-Sensor, eine erste Linse, welche mit der Fokuss-Länge assoziiert ist und vor dem ersten Bild-Sensor montiert ist, eine zweite Linse, welche mit der Fokuss-Länge assoziiert ist und vor dem zweiten Bild-Sensor montiert ist, und ein Blitz-Gerät umfasst, wobei der Prozessor ferner konfiguriert ist, um:  
den ersten Bild-Sensor und den zweiten Bild-Sensor abzutasten, um den Satz von Stereo-Bildern zu erzeugen, und  
zu veranlassen, dass der Satz von Stereo-Bildern in dem Speicher gespeichert wird.

Es folgen 7 Seiten Zeichnungen

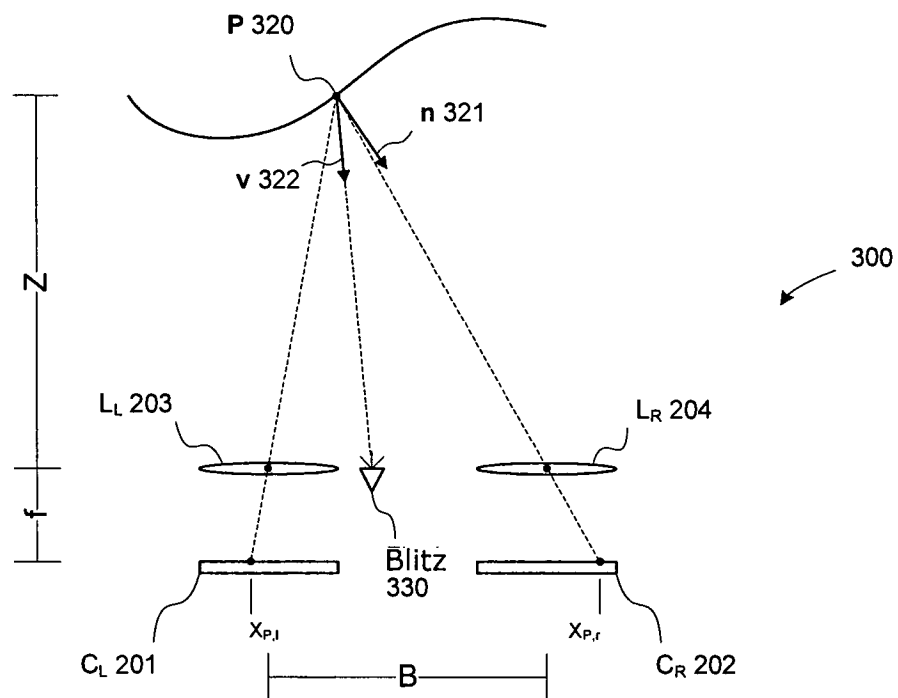
## Anhängende Zeichnungen



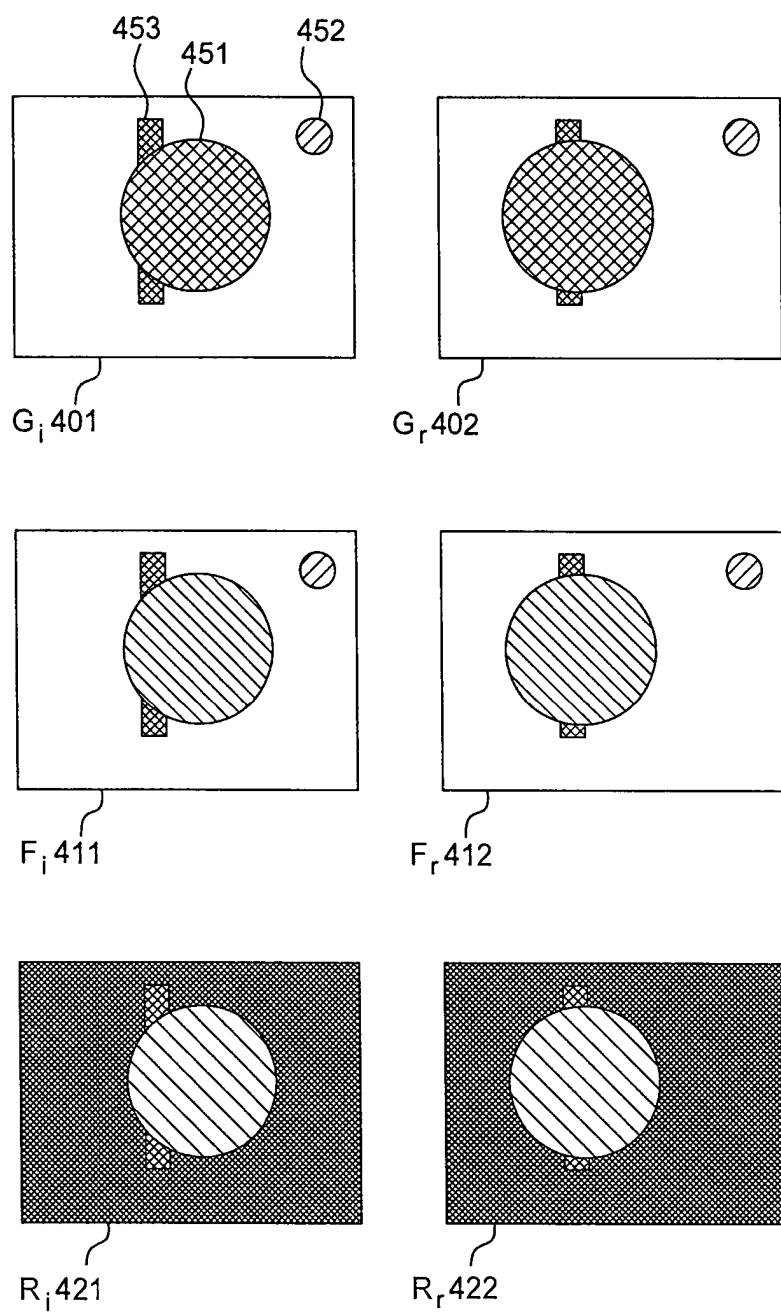
Figur 1



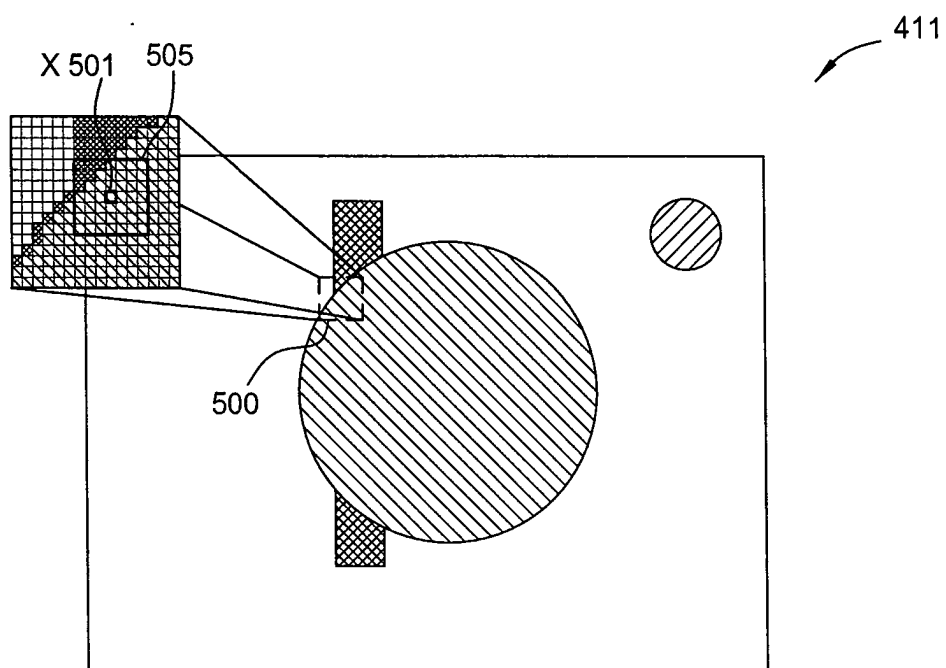
Figur 2



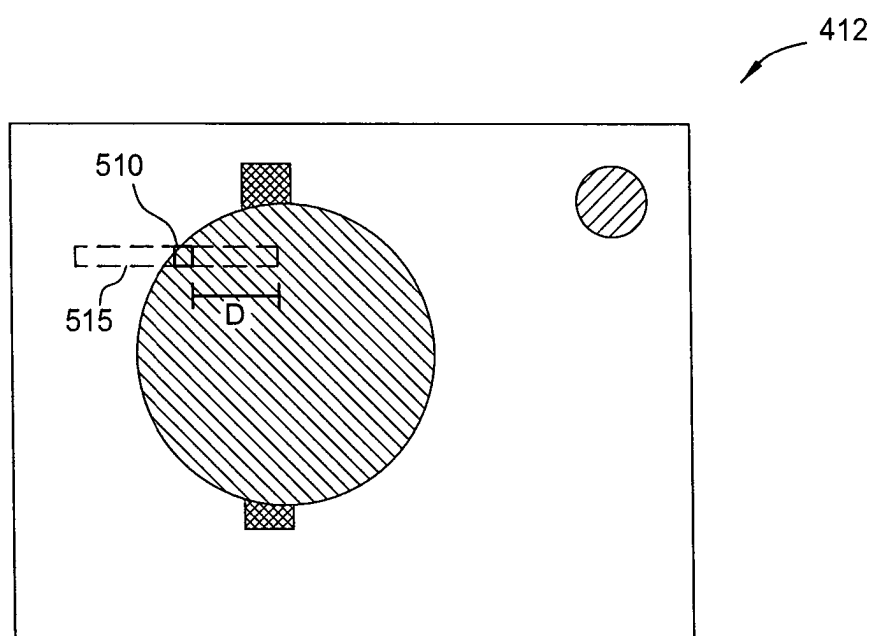
Figur 3



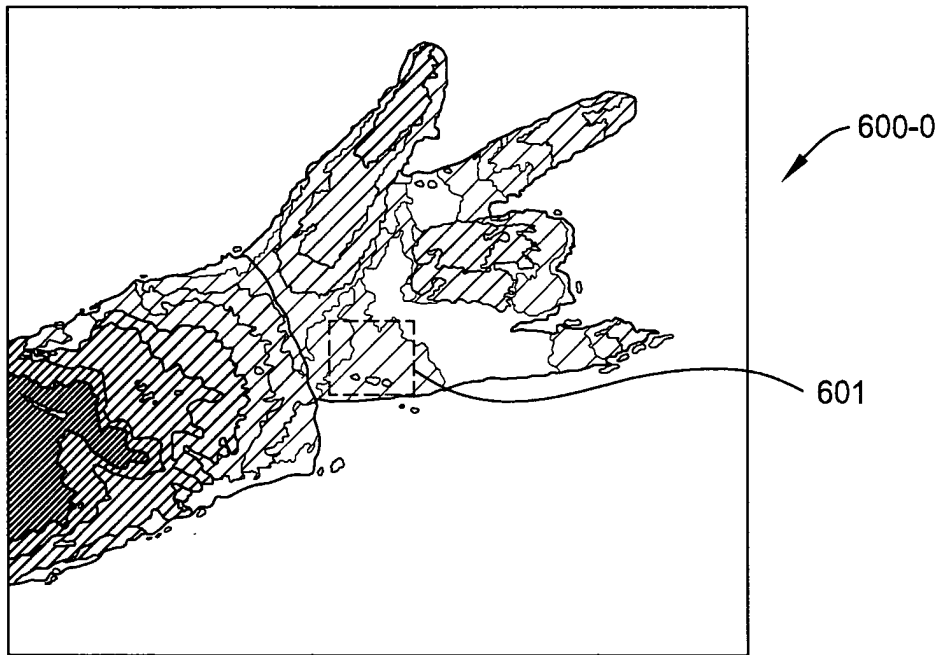
Figur 4



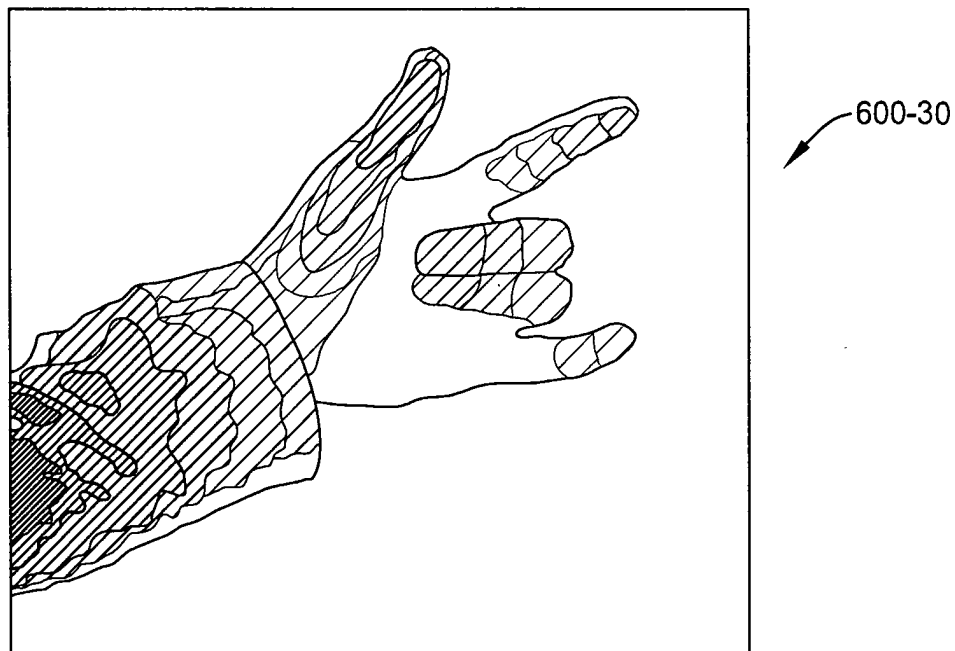
Figur 5A



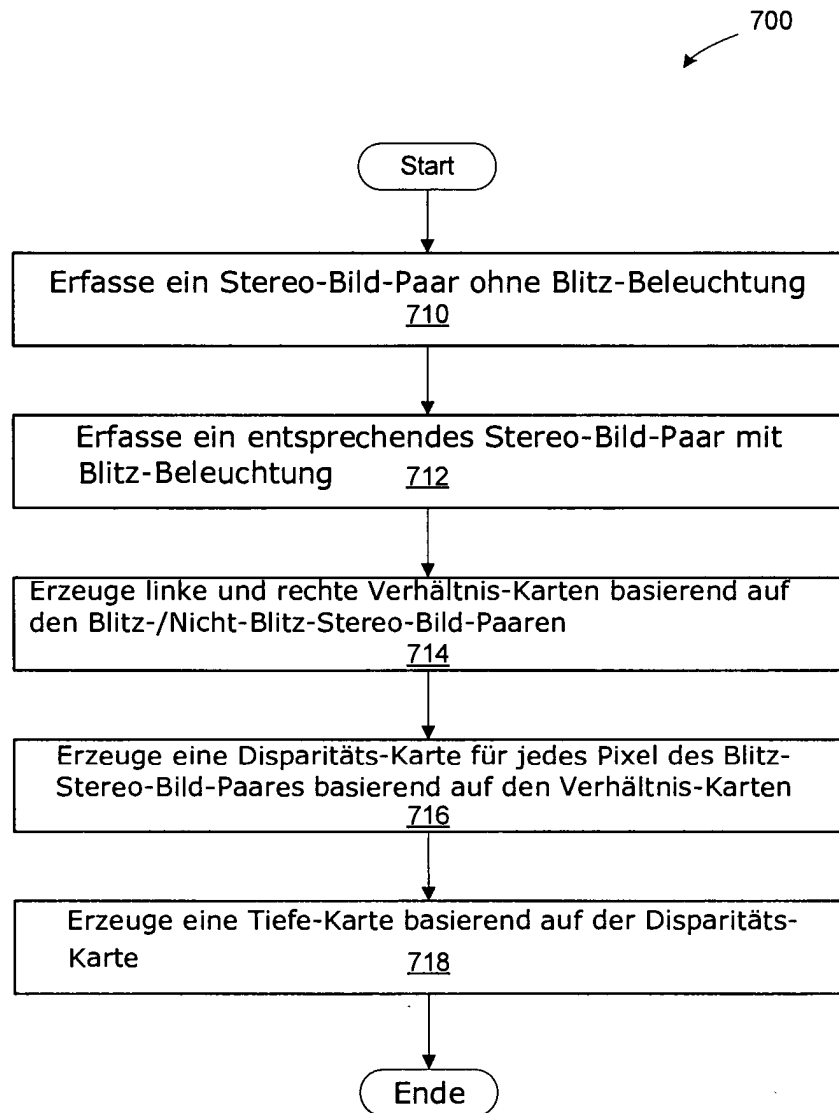
Figur 5B



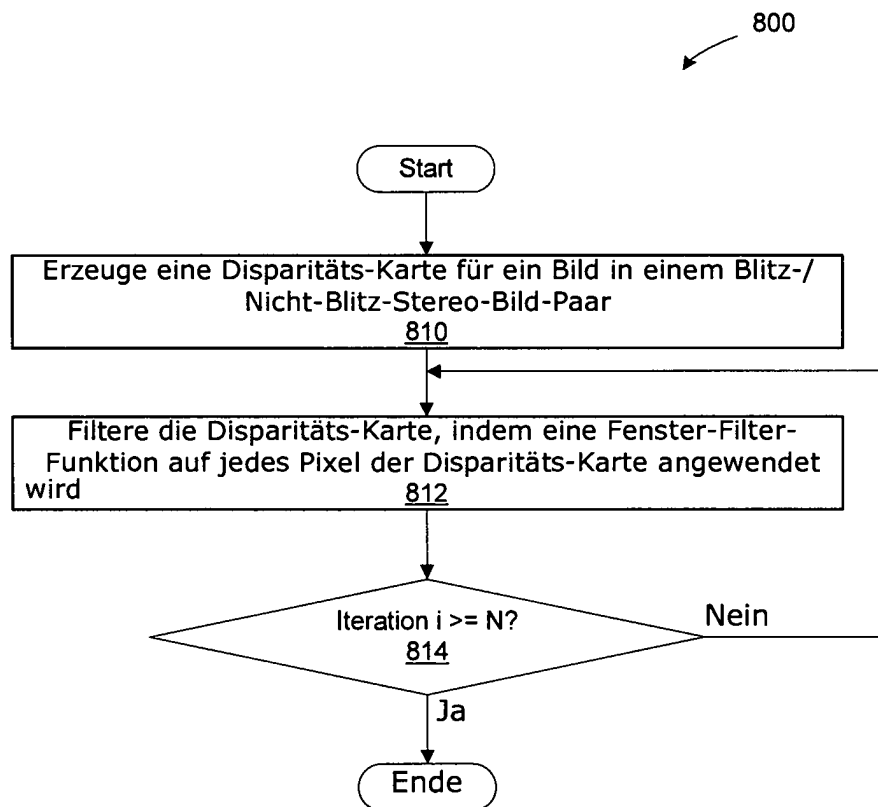
Figur 6A



Figur 6B



**Figur 7**



**Figur 8**