



(19)  
Bundesrepublik Deutschland  
Deutsches Patent- und Markenamt

(10) DE 60 2005 006 347 T2 2009.06.18

(12)

## Übersetzung der europäischen Patentschrift

(97) EP 1 612 738 B1

(21) Deutsches Aktenzeichen: 60 2005 006 347.1

(96) Europäisches Aktenzeichen: 05 105 310.6

(96) Europäischer Anmeldetag: 16.06.2005

(97) Erstveröffentlichung durch das EPA: 04.01.2006

(97) Veröffentlichungstag  
der Patenterteilung beim EPA: 30.04.2008

(47) Veröffentlichungstag im Patentblatt: 18.06.2009

(51) Int Cl.<sup>8</sup>: G06T 15/00 (2006.01)

G06T 15/20 (2006.01)

(30) Unionspriorität:  
879235 28.06.2004 US

(73) Patentinhaber:  
Microsoft Corp., Redmond, Wash., US

(74) Vertreter:  
Grünecker, Kinkeldey, Stockmair &  
Schwanhäusser, 80802 München

(84) Benannte Vertragsstaaten:  
AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB,  
GR, HU, IE, IS, IT, LI, LT, LU, MC, NL, PL, PT, RO,  
SE, SI, SK, TR

(72) Erfinder:  
Zitnick III, Charles, Redmond, WA 98052, US;  
Uyttendaele, Matthew, Redmond, WA 98052, US;  
Szeliski, Richard, Redmond, WA 98052, US;  
Winder, Simon, Redmond, WA 98052, US; Kang,  
Sing Bing, Redmond, WA 98052, US

(54) Bezeichnung: System und Verfahren zur Erzeugung einer zweilagigen 3D-Darstellung einer Szene

Anmerkung: Innerhalb von neun Monaten nach der Bekanntmachung des Hinweises auf die Erteilung des europäischen Patents kann jedermann beim Europäischen Patentamt gegen das erteilte europäische Patent Einspruch einlegen. Der Einspruch ist schriftlich einzureichen und zu begründen. Er gilt erst als eingelebt, wenn die Einspruchsgebühr entrichtet worden ist (Art. 99 (1) Europäisches Patentübereinkommen).

Die Übersetzung ist gemäß Artikel II § 3 Abs. 1 IntPatÜG 1991 vom Patentinhaber eingereicht worden. Sie wurde vom Deutschen Patent- und Markenamt inhaltlich nicht geprüft.

**Beschreibung**

Hintergrund

Technisches Gebiet

**[0001]** Die Erfindung bezieht sich auf geschichtete Darstellungen von digitalen oder digitalisierten Bildern, und insbesondere auf ein System und Verfahren zum Erzeugen einer zweischichtigen 3D-Darstellung einer Szene.

Stand der Technik

**[0002]** Seit mehreren Jahren sehen Zuschauer von Fernsehwerbung und Spielfilmen nun den "Standbild"-Effekt, der eingesetzt wird, um die Illusion von angehaltener Zeit und einem sich verändernden Kamerastandpunkt zu erzeugen. Die ersten Werbespots wurden unter Verwendung eines filmbasierter Systems gedreht, das schnell zwischen verschiedenen Standbildkameras hin- und hersprang, die entlang einer Schiene aufgestellt waren, um die Illusion zu erzeugen, man bewege sich durch einen eingefrorenen Zeitabschnitt.

**[0003]** Als der Effekt zum ersten Mal herauskam, war er neu und sah spektakulär aus und wurde bald in zahlreichen Produktionen nachgeahmt, wobei am bekanntesten wahrscheinlich die "Bullet Time"-Effekte aus dem Film "Matrix" sind. Unglücklicherweise ist dieser Effekt eine einmalige, vorausgeplante Angelegenheit. Die Standpunkt-Bahnkurve wird vorab geplant, und viele Arbeitsstunden werden damit verbracht, die erwünschten interpolierten Ansichten zu produzieren. Neuere Systeme beruhen auf Videokamera-Anordnungen, verlassen sich jedoch nach wie vor auf den Einsatz zahlreicher Kameras, um die Software-Bildinterpolation zu vermeiden.

**[0004]** Die bestehenden Systeme erlauben es einem Benutzer also nicht, interaktiv zu einem erwünschten Standpunkt zu wechseln, während gleichzeitig eine auf dynamischen Bildern beruhende Szene betrachtet wird. Der größte Teil der Arbeit an Image Based Rendering (IBR) in der Vergangenheit betrifft die Erzeugung statischer Szenen, wobei die zwei bekanntesten Techniken das Light-Field-Renderring [11] und der Lumigraph [7] sind. Deren Erfolg bei der hochwertigen Bilderzeugung stammt von der Verwendung einer großen Anzahl aufgezeichneter Bilder und hat umfangreiche Arbeiten auf dem Gebiet inspiriert. Eine spannende potenzielle Erweiterung dieser bahnbrechenden Arbeit betrifft die interaktive Steuerung des Standpunkts beim Ansehen eines Videos. Die Fähigkeit des Benutzers, den Standpunkt eines Videos interaktiv zu steuern verbessert das Betrachtungserlebnis wesentlich, da sie zahlreiche Anwendungen ermöglicht, wie die Sofortwiedergabe neuer Standpunkte, eine Änderung des Standpunkts in

Dramen, und jederzeit das Erzeugen des visuellen "Standbild"-Effekts.

**[0005]** Die Erweiterung von IBR auf dynamische Szenen ist allerdings nicht trivial aufgrund der Schwierigkeit (und der Kosten), so viele Kameras zu synchronisieren sowie die Bilder zu erfassen und zu speichern. Nicht nur gilt es Hürden zu überwinden beim Aufzeichnen, Darstellen und Rendern dynamischer Szenen von mehreren Standpunkten aus, sondern dies auch interaktiv tun zu können, stellt eine beträchtliche weitere Komplikation dar. Bis heute sind die Versuche, dieses Ziel zu verwirklichen, unbefriedigend geblieben.

**[0006]** Im Hinblick auf die Aspekte eines Videosystems mit interaktivem Standpunkt beim videobasierten Rendering war einer der ersten Versuche beim Aufzeichnen und Rendern dynamischer Szenen das Virtualized Reality System [10] von Kanade et al., bei dem 51 Kameras um eine 5-Meter geodätische Kuppel angeordnet waren. Die Auflösung jeder Kamera beträgt  $512 \times 512$  und die Aufzeichnungsrate 30 Einzelbilder pro Sekunde (fps). In jedem Zeitrahmen wird eine globale Oberflächendarstellung extrahiert, wobei eine Form von Voxelfarbgebung [14] verwendet wird, basierend auf der Szenenflussgleichung [17]. Unerwarteterweise sieht das Ergebnis unrealistisch aus, aufgrund von niedriger Auflösung, Abgleichfehlern und ungenauem Verarbeiten der Objektgrenzen.

**[0007]** Carranza et al. [3] nutzten sieben synchronisierte Kameras, die ringsum in einem Raum verteilt sind, mit Blick in dessen Mitte, um menschliche Bewegung 3D aufzuzeichnen. Jede Kamera hat eine CIF Auflösung ( $320 \times 240$ ) und zeichnet mit 15 fps auf. Ein dreidimensionales menschliches Modell wird als Vorabeingabe verwendet, um bei jedem Zeitrahmen eine 3D-Form zu berechnen.

**[0008]** Yang et al. [18] entwarfen ein  $8 \times 8$  Kameragitter (jede  $320 \times 240$ ), um eine dynamische Szene aufzuzeichnen. Anstatt die Daten zu speichern und zu rendern, werden nur die Strahlen übertragen, die notwendig sind, um die erwünschte virtuelle Ansicht zusammenzusetzen. In diesem System sind die Kameras nicht Genlock-synchronisiert, sondern stützen sich auf interne Taktgeber über sechs PCs. Die Kameraaufzeichnungsrate beträgt 15 fps und die Vorführrate beträgt 18 fps.

**[0009]** Diesem Ansatz ist mit den vorangehenden Systemen gemein, dass zahlreiche Bilder für ein realistisches Rendern erforderlich sind, zum Teil weil die Szenengeometrie entweder unbekannt oder nur in einer groben Annäherung bekannt ist. Wenn die Geometrie genau bekannt ist, kann der Bedarf an Bildern wesentlich verringert werden [7]. Ein zweckmäßiger Weg zum Extrahieren der Szenengeometrie ist der

Einsatz von Stereoaufnahmen, und für statische Szenen sind zahlreiche Stereoalgorithmen vorgeschlagen [13] worden. Es gab jedoch bisher wenige Versuche, Stereotechniken bei dynamischen Szenen einzusetzen. Als Teil der Arbeit Virtualized Reality [10] schlugen Vedula et al. [17] einen Algorithmus zum Extrahieren von 3D-Bewegung vor (d. h. eine Korrespondenz zwischen einer Szenenform über die Zeit), unter Verwendung eines 2D optischen Flusses und einer 3D-Szenenform. In ihrem Ansatz nutzen sie ein Auswahlschema ähnlich der Voxelfarbgebung [14], wobei das verwendete Maß ist, wie gut eine hypothetische Voxelposition in die 3D-Flussgleichung passt.

**[0010]** Zhang und Kambhamettu [19] integrierten ebenfalls 3D-Szenenfluss und -struktur in ihr System. Deren 3D-affines Bewegungsmodell wird lokal eingesetzt, mit räumlicher Regularisierung, und Diskontinuitäten werden unter Verwendung von Farbsegmentierung erhalten. Tao et al. [16] gehen davon aus, dass die Szene stückweise planar ist. Außerdem nehmen sie eine konstante Geschwindigkeit für jeden ebenen Fleck an, um die dynamische Tiefenkartenschätzung zu beschränken.

**[0011]** In einem ehrgeizigeren Versuch entdecken Carceroni und Kutulakos [2] die stückweise kontinuierliche Geometrie und Reflexion (Phong-Modell) unter weicher Bewegung mit bekannten Beleuchtungspositionen wieder. Sie diskretisieren den Raum in Oberflächenelemente (Surface Elements, "Surfels") und führen eine Suche über Standort-, Ausrichtungs- und Reflexionsparameter durch, um die Übereinstimmung mit den beobachteten Bildern zu maximieren.

**[0012]** In einer Variante zu der herkömmlichen lokalen Window-Matching-Technik, nutzen Zhang et al. [20] zusammengehörende Fenster, die Raum und Zeit überspannen. Der Vorteil dieses Verfahrens ist, dass weniger Abhängigkeit von einer Beleuchtungskonstanz über die Zeit besteht.

**[0013]** Auch aktive Entfernungsmessungstechniken wurden auf bewegte Szenen angewandt. Hall-Holt und Rusinkiewicz [8] nutzen grenzkodierte Streifenmuster, die über die Zeit variieren. Außerdem gibt es ein handelsübliches System auf dem Markt, das ZCam™ heißt und von 3DV Systems aus Israel hergestellt wird, und bei dem es sich um einen entfernungsmegenden Videokamerazusatz handelt, der in Verbindung mit einer Videokamera für Sendezwecke verwendet wird. Allerdings ist es ein teures System, das nur eine Einzelstandpunkt Tiefe bietet, so dass es weniger gut für ein Mehrfachstandpunktvideo geeignet ist.

**[0014]** Wie auch immer, trotz aller Fortschritte im Stereo- und bildbasierten Rendering ist es immer noch sehr schwierig, hochwertige, hochauflösende Ansichten dynamischer Szenen wiederzugeben. Ein

Ansatz, wie in der Schrift Light Field Rendering [11] vorgeschlagen, ist es, einfach Strahlen zu resampeln, basierend nur auf den relativen Positionen der Eingabe und der virtuellen Kameras. Wie in dem Lumigraph [7] und nachfolgenden Arbeiten demonstriert, kann jedoch die Verwendung eines 3D-Impostors oder -Proxs für die Szenengeometrie die Qualität der interpolierten Ansichten wesentlich verbessern. Ein anderer Ansatz ist es, durch Texture Mapping ein einzelnes 3D-Modell zu erzeugen [10], doch ergibt dies in der Regel minderwertige Ergebnisse, verglichen mit der Verwendung multipler Referenzansichten. Wiederum ein weiterer Ansatz verwendet eine geometrie-unterstützte, bildbasierte Render-Methode, die einen 3D-Proxy erfordert. Eine Möglichkeit ist es, ein einzelnes, globales Polyedermodell zu verwenden, wie in den Schriften zu Lumigraph und Unstructured Lumigraph [1]. Eine andere Möglichkeit ist es, Per-Pixel-Tiefe zu verwenden, wie in Layered Depth Images [15], Offset-Tiefenkarten in Facade [5], oder Sprites mit Tiefe [15]. Im Allgemeinen führt die Verwendung verschiedener lokaler geometrischer Proxies für jede Referenzansicht [12, 6, 9] zu hochwertigeren Ergebnissen.

**[0015]** Allerdings ergeben sich auch bei multiplen Tiefenkarten, wie sie z. B. in "Mesh-based depth map compression and transmission for real-time view-based rendering" (XP 10607952) von Chai, B-B. et al. verwendet werden, noch das Rendern von Artefakten beim Erzeugen neuer Ansichten, d. h. Aliasing (Treppchen) aufgrund der plötzlichen Natur des Übergangs von dem Vordergrund zum Hintergrund und verunreinigte Farben aufgrund vermischter Pixel, die beim Zusammensetzen über neuen Hintergründen oder Objekten sichtbar werden.

**[0016]** Diesem Problem widmet sich die vorliegende Erfindung mittels einer einzigartigen zweischichtigen 3D-Darstellung von Eingabebildern. Es sei angemerkt, dass diese zweischichtige 3D-Darstellung nicht nur dazu verwendet werden kann, die vorwähnten Aliasing-Probleme im Zusammenhang mit dem Rendern neuer Ansichten in einem Videosystem mit interaktivem Standpunkt zu lösen, sondern auch genauso vorteilhaft in anderen Zusammenhängen eingesetzt werden kann. Im Allgemeinen kann jedes digitale oder digitalisierte Bild unter Verwendung dieser zweischichtigen 3D-Darstellung dargestellt werden.

**[0017]** Es sei angemerkt, dass sich die Beschreibung in den vorhergehenden Paragraphen sowie in dem Rest dieser Patentschrift auf verschiedene einzelne Publikationen bezieht, die durch ein numerisches Bezugszeichen in Klammern identifiziert werden. Eine solche Bezugnahme kann zum Beispiel identifiziert werden, indem "Quelle [1]" oder einfach "[1]" angeführt wird. Zahlreiche Bezugnahmen werden durch Klammern identifiziert, die mehr als ein Be-

zugszeichen enthalten, wie zum Beispiel [2, 3]. Eine Literaturangabe einschließlich der Veröffentlichungen, die mit jedem Bezugszeichen korrespondieren, findet sich am Ende des Abschnitts "detaillierte Beschreibung".

### Zusammenfassung

**[0018]** Die vorliegende Erfindung bezieht sich auf ein System und Verfahren zum Erzeugen einer zweischichtigen Darstellung eines digitalen oder digitalisierten Bildes. Im Allgemeinen umfassen die zwei Schichten eine Hauptschicht, die Pixel, die Hintergrundfarben und Hintergrund-Disparitäten aufweisen, die mit entsprechend angeordneten Pixeln von Bereichen von Tiefen-Diskontinuität in dem Bild verbunden sind, sowie Pixel umfasst, die Farben und Disparitäten aufweisen, die mit den entsprechend angeordneten Pixeln des Bildes verbunden sind, die nicht in diesen Bereichen von Tiefen-Diskontinuität gefunden werden. Die andere Schicht ist eine Grenzschicht, die aus Pixeln besteht, die Vordergrundfarben, Vordergrund-Disparitäten und Alpha-Werte aufweisen, die mit den entsprechend angeordneten Pixeln der Bereiche von Tiefen-Diskontinuität verbunden sind. Die Bereiche von Tiefen-Diskontinuität entsprechen Bereichen vorgeschriebener Größe, die Tiefen-Diskontinuitäten umgeben, die in dem Bild gefunden werden.

**[0019]** Diese zweischichtige Darstellung wird erzeugt, indem zunächst die Position der Tiefen-Diskontinuitäten in dem vorliegenden Bild unter Verwendung einer Disparitätskarte des Bildes identifiziert wird. Tiefen-Diskontinuitäten treten an Positionen auf, an welchen eine Differenz zwischen Disparitätswerten benachbarter Pixel mehr als eine vorgeschriebene Anzahl von Stufen beträgt. Die Pixel des Bildes, die in den Bereichen vorgeschriebener Größe gefunden werden, die die Tiefen-Diskontinuitäten umgeben, werden anschließend identifiziert. Im nächsten Schritt werden die Vordergrund- und Hintergrundfarben zusammen mit den Vordergrund-Alpha-Werten für jedes Pixel in diesen Bereichen von Tiefen-Diskontinuität unter Verwendung einer Matting-Methode geschätzt. Zusätzlich werden die Vordergrund- und Hintergrund-Disparitätswerte für jedes Pixel in den Bereichen von Tiefen-Diskontinuität unter Verwendung alpha-gewichteter Mittelwerte naheliegender Disparitäten in den Vordergrund- und Hintergrundabschnitten des Bildes geschätzt. Anschließend wird eine Grenzschicht des Bildes eingerichtet, die ein Pixel an jeder Position umfasst, die einem Pixel der Bereiche von Tiefen-Diskontinuität entspricht. Die Vordergrund-Farbe, Vordergrund-Disparität und der Alpha-Wert, die mit dem entsprechenden Pixel der Bereiche von Tiefen-Diskontinuität verbunden sind, werden dann jedem Grenzschicht-Pixel zugewiesen. Zusätzlich wird eine Hauptschicht des Bildes eingerichtet. Diese Hauptschicht umfasst ein Pixel an

jeder Position, die einem Pixel der Bereiche von Tiefen-Diskontinuität entspricht, sowie ein Pixel an jeder Pixelposition des Bildes, die nicht einem Pixel der Bereiche von Tiefen-Diskontinuität entspricht. Die Hintergrund-Farbe und der Hintergrund-Disparitätswert, die mit dem entsprechenden Pixel der Bereiche von Tiefen-Diskontinuität verbunden sind, werden jedem der entsprechend angeordneten Pixel in der Hauptschicht zugeordnet, wogegen die Farben und Disparitäts-Werte, die mit den Pixeln des Bildes verbunden sind, die nicht in den Bereichen von Tiefen-Diskontinuität gefunden werden, jedem der entsprechend angeordneten Pixel in der Hauptschicht zugeordnet werden. Es sei angemerkt, dass sobald die Vordergrund-Farben, Vordergrund-Disparitäten und Alpha-Werte für jedes Pixel der Bereiche von Tiefen-Diskontinuität eingerichtet worden sind, die Größe dieser Bereiche um eine vorgegebene Größe unter Verwendung herkömmlicher Erweiterungstechniken erweitert werden kann, um das Auftreten von Rissen während des Renderns des Bildes aus diesen Schichten zu verhindern.

**[0020]** Zusätzlich zu den gerade beschriebenen Vorteilen werden weitere Vorteile der vorliegenden Erfindung anhand der detaillierten Beschreibung offensichtlich, die im Folgenden unter Bezug auf die begleitenden Figuren gegeben wird.

### Beschreibung der Zeichnungen

**[0021]** Die spezifischen Merkmale, Aspekte und Vorteile der vorliegenden Erfindung werden besser verständlich unter Bezug auf die folgende Beschreibung, die angehängten Ansprüche und die begleitenden Zeichnungen, wobei:

**[0022]** [Fig. 1](#) ein Diagramm ist, das eine Universal-Recheneinrichtung zeigt, die ein beispielhaftes System zum Implementieren der vorliegenden Erfindung darstellt.

**[0023]** [Fig. 2](#) ein Graph ist, der die Disparitätswerte einer Linie von Pixeln gegen die Pixelposition darstellt, wobei der abrupte Schritt in den Disparitätswerten eine Tiefen-Diskontinuität repräsentiert.

**[0024]** [Fig. 3A](#) und [Fig. 3B](#) ein Ablaufdiagramm zeigen, das einen Prozess zum Erzeugen einer zweischichtigen Darstellung eines digitalen oder digitalisierten Bildes gemäß der vorliegenden Erfindung schematisch darstellt.

**[0025]** [Fig. 4\(a\)–\(e\)](#) Bilder sind, die das Ergebnis des Anwenders des Prozesses zum Erzeugen einer zweischichtigen Bilddarstellung aus [Fig. 3A–B](#) auf ein Bild einer Gruppe von Breakdancern zeigt. [Fig. 4\(a\)](#) zeigt die Hauptschicht-Farbschätzungen und [Fig. 4\(b\)](#) stellt die Hauptschicht-Disparitätsschätzungen dar. [Fig. 4\(c\)](#) zeigt die Grenz-

schicht-Farbschätzungen und [Fig. 4\(d\)](#) stellt die Grenzschicht-Disparitätsschätzungen dar. [Fig. 4\(c\)](#) repräsentiert die Grenzschicht-Alphaschätzungen. Es sei angemerkt, dass die Bilder aus [Fig. 4\(c\)–\(e\)](#) negativ dargestellt sind, so dass die transparenten/leeren Pixel weiß aussehen.

#### Detaillierte Beschreibung der bevorzugten Ausführungen

**[0026]** In der folgenden Beschreibung der bevorzugten Ausführungen der vorliegenden Erfindung wird Bezug auf die begleitenden Zeichnungen genommen, die einen Teil derer darstellen, und in welchen spezifische Ausführungen, in welchen die Erfindung ausgeübt werden kann, als Beispiel gezeigt sind. Es wird davon ausgegangen, dass andere Ausführungen verwendet und strukturelle Änderungen vorgenommen werden können, ohne dabei von dem Geltungsbereich der vorliegenden Erfindung abzuweichen.

#### 1.0 Die Rechnerumgebung

**[0027]** Bevor eine Beschreibung der bevorzugten Ausführungen der vorliegenden Erfindung erfolgt, wird eine kurze, allgemeine Beschreibung einer geeigneten Rechnerumgebung, in der die Erfindung implementiert werden kann, gegeben. [Fig. 1](#) illustriert ein Beispiel einer geeigneten Rechensystemumgebung **100**. Die Rechensystemumgebung **100** ist lediglich ein Beispiel einer geeigneten Rechnerumgebung und soll keine Beschränkung hinsichtlich des Geltungsbereichs bei der Verwendung oder Funktionalität der Erfindung andeuten. Ebenso wenig sollte interpretiert werden, dass die Rechnerumgebung **100** eine Abhängigkeit oder Erfordernis bezüglich einer der Komponenten oder einer Kombination aus den Komponenten hat, die in der beispielhaften Betriebsumgebung **100** dargestellt sind.

**[0028]** Die Erfindung ist mit zahlreichen anderen Universal- oder Sonderzweckrechensystemumgebungen oder -konfigurationen betriebsfähig. Zu Beispielen bekannter Rechensysteme, -umgebungen, und/oder -konfigurationen, die für die Verwendung der Erfindung geeignet sein können, zählen, ohne jedoch darauf beschränkt zu sein, Personalcomputer, Serverrechner, Handgeräte oder Laptops, Mehrprozessorsysteme, Mikroprozessor-basierte Systeme, Set-Top-Boxen, programmierbare Unterhaltungselektronik, Netzwerkcomputer, Minicomputer, Großrechner, verteilte Rechnerumgebungen, die eines der oben genannten Systeme oder Geräte umfassen, und dergleichen.

**[0029]** Die Erfindung kann in dem allgemeinen Zusammenhang computer-ausführbarer Befehle beschrieben werden, wie zum Beispiel Programmmodulen, die von einem Computer ausgeführt werden

können. Im Allgemeinen umfassen Programmmodule Routinen, Programme, Objekte, Komponenten, Datenstrukturen, etc, die spezielle Aufgaben durchführen oder spezielle abstrakte Datentypen implementieren. Die Erfindung kann auch in verteilten Rechnerumgebungen angewendet werden, wo die Aufgaben von entfernten Datenverarbeitungseinrichtungen durchgeführt werden, die über ein Kommunikationsnetzwerk miteinander verbunden sind. In einer verteilten Rechnerumgebung können die Programmmoduln sowohl in lokalen als auch in entfernten Computerspeichermedien einschließlich Speichergeräten angeordnet sein.

**[0030]** Bezug nehmend auf [Fig. 1](#) umfasst ein beispielhaftes System zum Implementieren der Erfindung eine Universal-Recheneinrichtung in der Form eines Computers **110**. Zu den Komponenten des Computers **110** können, ohne darauf beschränkt zu sein, ein Prozessor **120**, ein Systemspeicher **130**, und ein Systembus **121** zählen, der die verschiedenen Systemkomponenten einschließlich des Systemspeichers mit dem Prozessor **120** verbindet. Der Systembus **121** kann einer von mehreren Typen von Busstrukturen sein, darunter ein Speicherbus oder eine Speichersteuerung, ein Peripheriebus, und ein lokaler Bus, wobei eine Vielzahl von Busarchitekturen verwendet werden kann. Zu solchen Architekturen zählen, um ein Beispiel zu geben und nicht einzuschränken, der Industry Standard Architecture (ISA) Bus, der Micro Channel Architecture (MCA) Bus, Enhanced ISA (EISA) Bus, der Video Electronics Standards Association (VESA) Localbus, und der Peripheral Component Interconnect (PCI) Bus, auch bekannt als Mezzanine-Bus.

**[0031]** Der Computer **110** beinhaltet üblicherweise eine Vielzahl computerlesbarer Medien. Computerlesbare Medien können alle verfügbaren Medien sein, auf die der Computer **110** zugreifen kann, darunter sowohl flüchtige als auch nichtflüchtige Medien, wechselbare und nicht-wechselbare Medien. Computerlesbare Medien können, um ein Beispiel zu geben und nicht einzuschränken, Computerspeichermedien und Kommunikationsmedien umfassen. Zu den Computerspeichermedien zählen sowohl flüchtige als auch nichtflüchtige, wechselbare und nicht-wechselbaren Medien, die in einem Verfahren oder einer Technologie zum Speichern von Informationen, wie zum Beispiel computerlesbare Befehle, Datenstrukturen, Programmmoduln oder andere Daten, implementiert sind. Zu den Computerspeichermedien zählen, ohne darauf beschränkt zu sein, RAM, ROM, EEPROM, Flash-Memory oder andere Speichertechnik, CD-ROM, DVDs oder andere optische Plattspeicher, Magnetkassetten, Magnetbänder, Magnetplattspeicher oder andere Magnetspeichereinrichtungen, oder jedes andere Medium, das verwendet werden kann, um die erwünschte Information zu speichern, und auf das der Computer **110** zu-

greifen kann. Kommunikationsmedien umfassen üblicherweise computerlesbare Befehle, Datenstrukturen, Programmmodulen oder andere Daten in einem modulierten Datensignal wie zum Beispiel einer Trägerwelle oder einem anderen Transportmechanismus und beinhalten Informationsausgabemedien. Der Begriff "moduliertes Datensignal" bezeichnet ein Signal, bei dem ein oder mehrere seiner Merkmale auf solche Weise eingestellt oder verändert sind, dass sie Informationen in dem Signal kodieren. Zu Kommunikationsmedien zählen, um ein Beispiel zu geben und nicht einzuschränken, verdrahtete Medien wie zum Beispiel ein festverdrahtetes Netz oder ein direkter Anschluss, und drahtlose Medien, wie zum Beispiel akustische, HF-, Infrarot- und anderweitig drahtlose Medien. Auch Kombinationen aus dem oben genannten sollen in den Bereich der computerlesbaren Medien mit eingeschlossen werden.

**[0032]** Der Systemspeicher **130** umfasst Computerspeichermedien in der Form eines flüchtigen und/oder nichtflüchtigen Speichers wie beispielsweise ein Nur-Lese-Speicher (ROM) **131** und ein Lese-Schreib-Speicher (RAM) **132**. Ein Basic Input-Output System (BIOS) **133**, das die Grundroutinen enthält, die dabei helfen, Informationen zwischen Elementen in dem Computer **110** zu übertragen, wie beispielsweise beim Starten, ist üblicherweise im ROM-Speicher **131** gespeichert. Der RAM-Speicher **132** enthält Daten und/oder Programmmodulen, auf die von dem Prozessor **120** unmittelbar zugegriffen werden kann, und/oder die von dem Prozessor momentan bearbeitet werden. [Fig. 1](#) illustriert, um ein Beispiel zu geben und nicht einzuschränken, das Betriebssystem **134**, die Anwenderprogramme **135**, weitere Programmmodulen **136**, und Programmdaten **137**.

**[0033]** Der Computer **110** kann außerdem weitere wechselbare/nicht-wechselbare, flüchtige/nichtflüchtige Computerspeichermedien beinhalten. Nur um ein Beispiel zu geben, illustriert [Fig. 1](#) ein Festplattenlaufwerk **141**, das nicht-wechselbare, nichtflüchtige magnetische Medien liest oder beschreibt, ein Magnetplattenlaufwerk **151**, das eine wechselbare, nichtflüchtige Magnetplatte **152** liest oder beschreibt, und ein optisches Laufwerk **155**, das wechselbare, nichtflüchtige optische Speicherplatten **156** liest oder beschreibt, wie beispielsweise eine CD-ROM oder andere optische Medien. Zu weiteren wechselbaren/nicht-wechselbaren, flüchtigen/nichtflüchtigen Computerspeichermedien, die in der beispielhaften Betriebsumgebung verwendet werden können, zählen, ohne jedoch darauf beschränkt zu sein, Magnetbandkassetten, Flash-Memory-Karten, DVDs, digitale Videobänder, Solid State RAM, Solid State ROM, und dergleichen. Das Festplattenlaufwerk **141** ist üblicherweise über eine nicht-wechselbare Speicherschnittstelle, wie zum Beispiel die Schnittstelle **140**, mit dem Systembus **121** verbunden, und das Mag-

netplattenlaufwerk **151** und das optische Laufwerk **155** sind üblicherweise über eine wechselbare Speicherschnittstelle, wie zum Beispiel die Schnittstelle **150**, mit dem Systembus **121** verbunden.

**[0034]** Die Laufwerke und deren oben beschriebene und in [Fig. 1](#) illustrierte verbundene Computerspeichermedien bieten Speicher für computerlesbare Befehle, Datenstrukturen, Programmmodulen und andere Daten für den Computer **110**. In [Fig. 1](#) ist das Festplattenlaufwerk **141** zum Beispiel als Speicher für das Betriebssystem **144**, Anwenderprogramme **145**, weitere Programmmodulen **146** und Programmdaten **147** dargestellt. Es sei angemerkt, dass es sich bei diesen Komponenten entweder um die gleichen oder andere als das Betriebssystem **134**, die Anwenderprogramme **136**, die weiteren Programmmodulen **136** und die Programmdaten **137** handeln kann. Das Betriebssystem **144**, die Anwenderprogramme **145**, die weiteren Programmmodulen **146** und die Programmdaten **147** haben hier andere Nummern erhalten, um zu illustrieren, dass es sich zumindest um andere Kopien handelt. Ein Benutzer kann in den Computer **110** Befehle und Informationen mit Hilfe von Eingabeeinrichtungen eingeben, wie zum Beispiel eine Tastatur **162** und ein Zeigegerät **161**, allgemein als eine Maus bezeichnet, Trackball oder Touch Pad. Zu weiteren Eingabeeinrichtungen (nicht dargestellt) können ein Mikrofon, Joystick, Gamepad, Satellitenschüssel, Scanner oder dergleichen zählen. Diese und weitere Eingabeeinrichtungen sind häufig über eine Benutzeereingabeschnittstelle **160** mit dem Prozessor **120** verbunden, die an den Systembus **121** gekoppelt ist, können jedoch auch über andere Schnittstellen- und Busstrukturen verbunden sein, wie zum Beispiel eine parallele Schnittstelle, Joystick-Schnittstelle oder einen universellen seriellen Bus (USB). Ein Monitor **191** oder ein anderer Typ von Anzeigegerät ist ebenfalls mit dem Systembus **121** über eine Schnittstelle, wie beispielsweise eine Video-Schnittstelle **190**, verbunden. Neben dem Monitor können die Computer auch weitere periphere Ausgabegeräte wie zum Beispiel Lautsprecher **197** und einen Drucker **196** umfassen, die durch eine periphere Ausgabeschnittstelle **195** verbunden sein können. Eine Kamera **192** (wie zum Beispiel eine digitale/elektronische Standbild- oder Videokamera, oder ein Film-/Fotoscanner), die dazu eingerichtet ist, eine Bildersequenz **193** aufzuzeichnen kann ebenfalls als eine Eingabeeinrichtung für den Personalcomputer **110** eingeschlossen sein. Weiterhin können, auch wenn nur eine Kamera abgebildet ist, mehrere Kameras als Eingabeeinrichtungen für den Personalcomputer **110** eingeschlossen sein. Die Bilder **193** von der einen Kamera oder mehreren Kameras werden über eine geeignete Kamerabeschaltung **194** in den Computer **110** eingegeben. Diese Schnittstelle **194** ist mit dem Systembus **121** verbunden, so dass es möglich ist, die Bilder zu dem RAM **132** zu leiten und dort zu speichern, oder zu einer der anderen Datenspeichereinrichtungen, die mit

dem Computer **110** verbunden sind. Wie auch immer, es sei jedoch angemerkt, dass Bilddaten auch von einem der zuvor genannten computerlesbaren Medien in den Computer **110** eingegeben werden können, ohne dass die Nutzung der Kamera **192** notwendig ist.

**[0035]** Der Computer **110** kann in einer vernetzten Umgebung unter Verwendung logischer Verbindungen zu einem oder mehreren entfernten Computer arbeiten, wie zum Beispiel dem entfernten Computer **180**. Der entfernte Computer **180** kann ein Personalcomputer sein, ein Server, ein Router, ein Netzwerkcomputer, eine Peer-Einrichtung oder ein anderer gemeinsamer Netzwerkknoten, und umfasst viele oder alle der bezüglich des Computers **110** oben beschriebenen Elemente, auch wenn in [Fig. 1](#) nur ein Speichergerät **181** illustriert worden ist. Die in [Fig. 1](#) abgebildeten logischen Verbindungen umfassen ein lokales Netzwerk (LAN) **171** und ein Weitverkehrsnetz (WAN) **173**, können jedoch auch andere Netzwerke umfassen. Solche vernetzten Umgebungen sind in Büros üblich, unternehmensweiten Computernetzwerken, Intranets und dem Internet.

**[0036]** Wenn der Computer **110** in einer LAN-Netzwerkumgebung verwendet wird, ist er über eine Netzwerkschnittstelle oder einen -Netzwerkadapter **170** mit dem LAN **171** verbunden. Wenn der Computer **110** in einer WAN-Netzwerkumgebung genutzt wird, umfasst er üblicherweise ein Modem **172** oder andere Mittel zum Einrichten von Kommunikationen über das WAN **173**, wie beispielsweise das Internet. Das Modem **172**, bei dem es sich um ein internes oder externes handeln kann, kann über die Benutzereingabeschmittstelle **160** oder einen anderen geeigneten Mechanismus mit dem Systembus **121** verbunden sein. In einer vernetzten Umgebung können Programmmodulen, die bezüglich des Computers **110** abgebildet sind, oder Teile davon in dem entfernten Speichergerät gespeichert sein. In [Fig. 1](#) sind, um ein Beispiel zu geben und nicht einzuschränken, entfernte Anwenderprogramme **185** dargestellt, die dem Speichergerät **181** innewohnen. Es versteht sich, dass die gezeigten Netzwerkverbindungen exemplarisch sind und andere Mittel zum Einrichtung einer Kommunikationsverbindung zwischen den Computer verwendet werden können.

## 2.0 Die zweischichtige Bilddarstellung

**[0037]** Nachdem nun die beispielhafte Betriebsumgebung erörtert worden ist, beschäftigt sich der verbleibende Teil dieses Beschreibungsabschnitts mit einer Beschreibung der Programmmodulen, welche die Erfindung verkörpern. Im Allgemeinen betrifft die vorliegende Erfindung das Erzeugen einer einzigartigen zweischichtigen 3D-Darstellung eines Bildes, die dessen Komprimierung, Übertragung und Speicherung erleichtert. Diese Darstellung ist insbesondere

von Nutzen, wenn das Bild ein Frame eines Videos einer dynamischen Szene ist, und die Framedaten für das Echtzeit-Rendering kodiert sind. Außerdem schließt sie Pixel-Disparität oder Tiefeninformationen ein, und somit das Bereitstellen der 3D-Aspekte der Darstellung. Das Bild oder der Frame ist ein digitales Bild, das entweder von einer Digitalkamera aufgezeichnet worden ist, oder falls nicht, vor dem Durchlaufen weiterer Verarbeitungsprozesse digitalisiert worden ist. Weiterhin sei angemerkt, dass die digitalen Bilddaten insofern bildbasiert sein können, als sie durch das Aufzeichnen eines tatsächlichen Bildes einer Szene unter Verwendung einer Kamera erfasst worden sind, oder synthetisierte Bilddaten sind.

**[0038]** Die zweischichtige Darstellung wird erzeugt, indem zunächst die Tiefen-Diskontinuitäten in einer Disparitätskarte des vorliegenden Bildes oder Frames lokalisiert werden. Diese Tiefen-Diskontinuitäten sind definiert als jeder Sprung, der größer als eine vorgeschriebene Anzahl an Disparitätsstufen ist (z. B. 4 Stufen in getesteten Ausführungen der vorliegenden Erfindung). Die Disparitätskarte des vorliegenden Bildes kann auf jede herkömmliche Weise erfasst werden. In den getesteten Ausführungen der vorliegenden Erfindung jedoch wurde ein neuer Ansatz verwendet, der Gegenstand der Parallelanmeldung mit dem Titel "Color Segmentation-Based Stereo Reconstruction System And Process" von den Erfindern dieser Anmeldung und dem gemeinsamen Erwerber zugewiesen ist. Die Parallelanmeldung wurde am ... zu den Akten genommen und erhielt die Seriennummer ....

**[0039]** Als nächstes werden kleine Bereiche in der Nähe der Tiefen-Diskontinuität identifiziert, die in dem vorliegenden Bilde gefunden werden. Diese kleinen Bereiche sind so definiert, dass sie alle Pixel innerhalb 3 Pixeln von der Position der Tiefen-Diskontinuitäten umfassen. Dies ist in dem Graph aus [Fig. 2](#) illustriert, in dem die Disparitätswerte einer Linie von Pixeln gegen die Pixelposition dargestellt sind. Der abrupte Schritt in den Disparitätswerten stellt eine Tiefen-Diskontinuität dar, wobei angenommen wird, dass sie größer als eine vorgeschriebene Anzahl von Disparitätsstufen ist. Der zuvor erwähnte kleine Bereich, der als der Bereich von Tiefen-Diskontinuität bezeichnet wird, wird rund um die Diskontinuität eingerichtet. In dem Graph aus [Fig. 2](#) dargestellten Profilansicht hat dieser Bereich eine Breite, die die Position der Tiefendisparität in dem Bild überspannt. Wenn die Tiefen-Diskontinuität dem Umriss eines Objekts in dem Bild folgt, wie es üblicherweise der Fall ist, verschmelzen die Bereiche von Tiefen-Diskontinuität miteinander, um einen Streifen zu bilden, der diesem Umriss folgt.

**[0040]** Einige Pixel entlang der Objektgrenzen erhalten Einbringungen von sowohl den Hintergrund- als auch den Vordergrundgebieten. Wenn allerdings

die originalen gemischten Pixelfarben während des Renders verwendet werden, können sichtbare Artefakte entstehen. Daher ist es wichtig, diese zwei Eingaben zu trennen. Entsprechend werden als nächstes die Vordergrund- und Hintergrundfarben jedes Pixels in den Bereichen von Tiefen-Diskontinuität eingerichtet, sowie die Opazität der Pixel (in der Form von Alpha-Werten). Dies wird unter Verwendung einer Matting-Methode ausgeführt. Im Allgemeinen beinhalten Matting-Methoden das Extrahieren der Vordergrundelemente und der Hintergrundelemente eines Bildes durch Schätzen der Farbe und Opazität, die von den Vordergrund- und Hintergrundelementen an jedem Pixel stammen. Die Opazität eines Pixels ist durch einen Alpha-Wert in dem Bereich zwischen 0 und 1 definiert. Tatsächlich definiert der Alpha-Wert den Prozentsatz der Farbe eines Pixels, der dem Vordergrundelement zuzuschreiben ist. Das Ergebnis eines typischen Matting-Vorgangs ist es, eine Vordergrundfarbe, eine Hintergrundfarbe und einen Alpha Wert für jedes berücksichtigte Pixel zu identifizieren. Zwar kann jedes Matting-Verfahren im Zusammenhang mit der vorliegenden Erfindung eingesetzt werden, doch wurde bei den getesteten Ausführungen eine bayesianische Bild-Matting-Methode verwendet, die in Quelle [4] beschrieben ist. Es sei angemerkt, dass in Quelle [4] auch eine Beschreibung zahlreicher anderer bestehender Matting-Methoden enthalten ist.

**[0041]** Wie zuvor beschrieben, beinhaltet eine Primärerwendung der vorliegenden Erfindung das Rendern virtueller Bilder einer Szene aus Standpunkten, die zwischen den Standpunkten liegen, die mit zwei tatsächlichen Bildern der Szene verbunden sind. Zwar liegen die Details dessen, wie dies zur Ausführung gelangt, nicht im Geltungsbereich der vorliegenden Erfindung, doch sei angemerkt, dass der Prozess üblicherweise die Kenntnis der mit jedem Pixel verbundenen Tiefen voraussetzt. Zu diesem Zweck beinhaltet die Bilddarstellung gemäß der vorliegenden Erfindung die Pixeltiefen (oder Disparitätswerte, die problemlos in Tiefenwerte umgewandelt werden können). Auch wenn diese Werte für die meisten Pixel aus der zuvor erwähnten Disparitätskarte zur Verfügung stehen, sei angemerkt, dass es zwei Tiefen gibt, die mit den gemischten Pixeln verbunden sind, die in der Nähe einer Tiefen-Diskontinuität gefunden werden, nämlich die Tiefe, die mit dem Vordergrundelement verbunden ist und die Tiefe, die mit dem Hintergrund verbunden ist. Also werden als nächstes diese Vordergrund- und Hintergrundtiefen-(oder Disparitäts-)Werte für jedes Pixel in dem Bereich von Tiefen-Diskontinuität geschätzt. Im Allgemeinen wird dies unter Verwendung von alpha-gewichteten Mittelwerten naheliegender Disparitäten in den Vordergrund-respektive Hintergrundabschnitten des Bildes ausgeführt. Genauer gesagt, werden die Vordergrund-Disparitäten unter Verwendung eines alpha-gewichteten Mittelwertes innerhalb eines Fen-

ters (z. B.  $7 \times 7$  Pixel) der Disparitätswerte von Pixeln innerhalb des originalen Vordergrundbereichs des Bereichs von Tiefen-Diskontinuität gefunden. In [Fig. 2](#) bezieht sich der originale Vordergrundbereich auf die Pixel links von der Tiefen-Diskontinuität. Die Hintergrund-Disparitäten werden durch einen gewichteten Mittelwert innerhalb eines Fensters der Disparitätswerte in dem originalen Hintergrundbereich des Bereichs von Tiefen-Diskontinuität gefunden. Die Gewichte, die zum Mitteln der Hintergrund-Disparitäten verwendet werden, werden gleich eins minus dem Alpha-Wert gesetzt.

**[0042]** Sobald die Pixelfarben, Disparitäten und Opazitäten des Bereichs von Tiefen-Diskontinuität eingerichtet sind, werden die zwei Schichten der Bilddarstellung gebildet. Genauer gesagt werden die zuvor berechnete Vordergrundfarbe, Vordergrund-Disparität und die mit jedem entsprechenden Pixel, das in den Bereichen von Tiefen-Diskontinuität gefunden wird, verbundenen Alpha-Werte einem entsprechend angeordneten Pixel der Grenzschicht des Bildes zugewiesen. Entsprechend werden die zuvor berechneten Hintergrundfarb- und Hintergrund-Disparitätswerte, die mit jedem entsprechenden Pixel in den Bereichen von Tiefen-Diskontinuität verbunden sind, einem entsprechend angeordneten Pixel der Hauptschicht des Bildes zugewiesen, zusammen mit den Farb- und Disparitätswerten aller Pixel in dem Bild, die sich nicht in den Bereichen von Tiefen-Diskontinuität befinden (wie aus dem Bild und dessen Disparitätskarte entnommen).

**[0043]** Demgemäß umfasst die resultierende zweischichtige Bilddarstellung eine Grenzschicht, die die Vordergrundfarbe, Vordergrund-Disparität und den Alpha-Wert für jedes Pixel in dieser Schicht identifiziert. Zusätzlich umfasst die Darstellung eine Hauptschicht, die die Hintergrundfarbe und Hintergrundtiefe der Pixel in der Schicht identifiziert, die mit den Bereichen von Tiefen-Diskontinuität verbunden sind, und den Farb- und Disparitätswert für jedes andere Pixel in der Schicht. In Anbetracht dessen ist ersichtlich, dass die Hauptschicht Pixelinformationen über jedes Pixel in dem Bild enthält, wobei jedoch diese Daten von einem Bild der Szene zum nächsten relativ ähnlich sind. Dies gilt sowohl zeitlich, wenn die Bilder Video-Frames sind, als auch räumlich, wenn mehrere Bilder derselben Szene übertragen oder gespeichert werden. Somit kann die Hauptschicht unter Verwendung herkömmlicher Bildkomprimierungsverfahren stark komprimiert werden. Außerdem enthält die Grenzschicht relativ wenige Daten, wie auch üblicherweise nur ein kleiner Prozentsatz der Gesamtanzahl der Pixel in dem Bild in der Schicht enthalten ist. Somit gibt es, auch wenn sich diese Daten voraussichtlich von einem Bild einer dynamischen Szene zum nächsten wesentlich verändern, auf die Zeit gerechnet, nicht so viele Daten, die übertragen oder gespeichert werden müssen. Aufgrund dessen können

diese Daten sogar ohne Komprimierung problemlos übertragen oder gespeichert werden. Mit der Verwendung geeigneter Komprimierungssysteme können die Größe der Daten und die Übertragungsdauer derselben noch weiter reduziert werden. Auf diese Weise bietet die vorerwähnte zweischichtige Bilddarstellung eine erwünschte Erleichterung bei der Datenübertragung und -speicherung und macht das Echtzeit-Rendering durchführbar.

**[0044]** Der vorerwähnte Erzeugungsprozess einer zweischichtigen Bilddarstellung wird nun unter Bezug auf das Ablaufdiagramm, das in [Fig. 3A–B](#) gezeigt ist, umrissen. Zunächst werden das digitale oder digitalisierte Bild, das dargestellt werden soll, und seine Disparitätskarte eingegeben (Verfahrensschritt 300). Die Positionen der Tiefen-Diskontinuitäten werden dann in dem Bild unter Verwendung der Disparitätskarte identifiziert (Verfahrensschritt 302). Als nächstes werden die Pixel, die in Bereichen vorgeschriebener Größe gefunden werden, die die Tiefen-Diskontinuitäten umgeben, identifiziert und als in den Bereichen von Tiefen-Diskontinuität des Bildes befindlich gekennzeichnet (Verfahrensschritt 304). Im nächsten Schritt werden die Vordergrund- und Hintergrundfarben jedes Pixels in den Bereichen von Tiefen-Diskontinuität gekennzeichnet, zusammen mit seinem Alpha-Wert, wobei eine Matting-Methode verwendet wird (Verfahrensschritt 306). Außerdem wird, in Verfahrensschritt 308, ein Vordergrund- und Hintergrund-Disparitätswert für jedes der Pixel in den Bereichen von Tiefen-Diskontinuität geschätzt. Dann wird eine Grenzschicht eingerichtet, die Pixel entsprechend den Pixeln der Bereiche von Tiefen-Diskontinuität aufweist (Verfahrensschritt 310). Die Vordergrundfarbe, Vordergrund-Disparität und die mit jedem entsprechenden Pixel, das in den Bereichen von Tiefen-Diskontinuität gefunden wird, verbundenen Alpha-Werte werden dem entsprechenden Pixel der Grenzschicht zugewiesen (Verfahrensschritt 312). Außerdem wird eine Hauptschicht eingerichtet, die Pixel entsprechend aller Pixel des Bildes aufweist (Verfahrensschritt 314). In diesem Fall werden die Hintergrundfarbe und Hintergrund-Disparität, die mit den Pixeln der Bereiche von Tiefen-Diskontinuität verbunden sind, den entsprechenden Pixeln der Hauptschicht zugewiesen und die Farbe und Disparität, die mit den Bildpixeln verbunden ist, die nicht in den Bereichen von Tiefen-Diskontinuität vorhanden sind, werden dem entsprechenden Pixel der Hauptschicht zugewiesen.

**[0045]** Eine Verbesserung, die in den vorerwähnten Prozess zum Erzeugen der zweischichtigen Bilddarstellung gemäß der vorliegenden Erfindung einbezogen werden kann, umfasst das Erweitern der Vordergrundmatte der Bereiche von Tiefen-Diskontinuität vor dem Bilden der Grenz- und Hauptschichten der Bilddarstellung. Dies hat den Vorteil, das Auftreten von Rissen während des Renderns des Bildes aus

den Schichten zu verhindern, wie es aufgrund von Ungenauigkeiten in dem Prozess der Fall sein könnte. Genauer gesagt wird, sobald die Vordergrundfarben, Vordergrund-Disparitäten und Alpha-Werte für jedes Pixel der Bereiche von Tiefen-Diskontinuität eingerichtet worden sind, die Größe dieser Bereiche um eine vorgegebene Größe (z. B. 1 Pixel) unter Verwendung herkömmlicher Erweiterungstechniken erweitert. Den neuen Pixeln, die der Grenzschicht während der Erweiterung hinzugefügt wurden, werden die gleichen Farb- und Disparitätswerte zugewiesen wie dem entsprechenden Pixel in der Hauptschicht, und ein Alpha-Wert von 1.

**[0046]** [Fig. 4\(a\)–\(e\)](#) zeigen die Ergebnisse des Anwenders des Prozesses zum Erzeugen einer zweischichtigen Bilddarstellung gemäß der vorliegenden Erfindung auf ein Bild einer Gruppe von Breakdancern. [Fig. 4\(a\)](#) zeigt die Hauptschicht-Farbschätzungen und [Fig. 4\(b\)](#) stellt die Hauptschicht-Disparitäts-schätzungen dar. Entsprechend zeigt [Fig. 4\(c\)](#) die Grenzschicht-Farbschätzungen und [Fig. 4\(d\)](#) stellt die Grenzschicht-Disparitätsschätzungen dar. Schließlich repräsentiert [Fig. 4\(e\)](#) die Grenzschicht-Alpha-schätzungen. Es sei angemerkt, dass die Bilder aus [Fig. 4\(c\)–\(e\)](#) negativ dargestellt sind, so dass die transparenten/leeren Pixel weiß aussehen. Man beachte, wie nur eine kleine Menge an Informationen übertragen werden muss, um die weichen Objektgrenzen zu bedingen, und wie die Grenzschichtopazitäten und die Farben in beiden Schichten sauber wiederhergestellt werden.

**[0047]** Es sei angemerkt, dass während der vorangehenden Beschreibung die Disparitätswerte der Bildpixel verwendet wurden. Diese Werte könnten jedoch an einem beliebigen Punkt in dem Erzeugungsprozess unter Verwendung herkömmlicher Verfahren in Tiefenwerte umgewandelt werden. In diesem Falle würden die Haupt- und Grenzschicht Tiefenwerte anstelle von Disparitätswerten für die in ihnen enthaltenen Pixel beinhalten.

### 3.0 Quellenangabe

- [1] Buehler, C., Bosse, M., McMillan, L., Gortler, S. J., und Cohen, M. F. 2001. Unstructured Lumigraph rendering. Proceedings of SIGGRAPH 2001 (August), 425–432.
- [2] Carceroni, R. L., und Kutulakos, K. N. 2001. Multi-view scene capture by surfel sampling: From video streams to non-rigid 3D motion, shape and reflectance. In Eighth international Conference an Computer Vision (ICCV 2001), Bd. 11, 60–67.
- [3] Carranza, J., Theobalt, C., Magnor, M. A., und Seidel, H.-P. 2003. Free-viewpoint video of human actors. ACM Transactions on Graphics 22, 3 (Juli), 569–577.
- [4] Chuang, Y.-Y., et al. 2001. A Bayesian approach to digital matting. In Conference on Computer Vision

- and Pattern Recognition (CVPR'2001), Bd. II, 264–271.
- [5] Debevec, P. E., Taylor, C. J., und Malik, J. 1996. Modeling and rendering architecture from photographs: A hybrid geometry- and image-based approach. Computer Graphics (SIGGRAPH'96) (August), 11–20.
- [6] Debevec, P. E., Yu, Y., und Borshukov, G. D. 1998. Efficient view-dependent image-based rendering with projective texture-mapping. Eurographics Rendering Workshop 1998, 105–116.
- [7] Gortler, S. J., Grzeszczuk, R., Szeliski, R., und Cohen, M. F. 1996. The Lumigraph. In Computer Graphics (SIGGRAPH'96) Proceedings, ACM SIGGRAPH, 43–54.
- [8] Hall-Holt, O., und Rusinkiewicz, S. 2001. Stripe boundary codes for realtime structured-light range scanning of moving objects. In Eighth International Conference on Computer Vision (ICCV 2001), Bd. II, 359–366.
- [9] Heigl, B., et al. 1999. Plenoptic modeling and rendering from image sequences taken by hand-held camera. In DAGM'99, 94–101.
- [10] Kanade, T., Rander, P. W., und Narayanan, P. J. 1997. Virtualized reality: constructing virtual worlds from real scenes. IEEE MultiMedia Magazine 1, 1 (Jan-März), 34–47.
- [11] Levoy, M., und Hanrahan, P. 1996. Light field rendering. In Computer Graphics (SIGGRAPH'96) Proceedings, ACM SIGGRAPH, 31–42.
- [12] Pulli, K., et al. 1997. View-based rendering: Visualizing real objects from scanned range and color data. In Proceedings of the 8th Eurographics Workshop on Rendering.
- [13] Scharstein, D., und Szeliski, R. 2002. A taxonomy and evaluation of dense twoframe stereo correspondence algorithms. International Journal of Computer Vision 47, 1 (Mai), 7–42.
- [14] Seitz, S. M., und Dyer, C. M. 1997. Photorealistic scene reconstruction by voxel coloring. In Conference an Computer Vision and Pattern Recognition (CVPR'97), 1067–1073.
- [15] Shade, J., Gortler, S., He, L.-W., und Szeliski, R. 1998. Layered depth images. In Computer Graphics (SIGGRAPH'98) Proceedings, ACM SIGGRAPH, Orlando, 231–242.
- [16] Tao, H., Sawhney, H., und Kumar, R. 2001. A global matching framework for stereo computation. In Eighth International Conference an Computer Vision (ICCV 2001), Bd. 1, 532–539.
- [17] Vedula, S., Baker, S., Seitz, S., und Kanade, T. 2000. Shape and motion carving in 6D. In Conference an Computer Vision and Pattern Recognition (CVPR'2000), Bd. II, 592–598.
- [18] Yang, J. C., Everett, M., Buehler, C., und McMillan, L. 2002. A real-time distributed light field camera. In Eurographics Workshop an Rendering, P. Debevec und S. Gibson, Hrsg., 77–85.
- [19] Zhang, Y., und Kambhamettu, C. 2001. On 3D scene flow and structure estimation. In Conference an Computer Vision and Pattern Recognition (CVPR'2001), Bd. II, 778–785.
- [20] Zhang, L., Curless, B., und Seitz, S. M. 2003. Spacetime stereo: Shape recovery for dynamic scenes. In Conference an Computer Vision and Pattern Recognition, 367–374.
- [21] Zhang, Z. 2000. A flexible new technique for camera calibration. IEEE Transactions an Pattern Analysis and Machine Intelligence 22, 11, 1330–1334.

## Patentansprüche

1. Computerimplementiertes Verfahren zum Erzeugen einer zweischichtigen Darstellung eines digitalen oder digitalisierten Bildes aus dem Bild und einer Disparitätskarte (disparity map) des Bildes, das Verwendung eines Computers zum Durchführen der folgenden Verfahrensvorgänge umfasst:  
 Identifizieren der Position von Tiefen-Diskontinuitäten in dem Bild unter Verwendung seiner Disparitätskarte;  
 Identifizieren der Pixel des Bildes, die in Bereichen vorgeschriebener Größe gefunden werden, die die Tiefen-Diskontinuitäten umgeben, und Kennzeichen dieser Pixel als in Bereichen von Tiefen-Diskontinuität des Bildes befindlich;  
 Schätzen einer Vordergrund- und einer Hintergrundfarbe sowie eines Alpha-Wertes für jedes Pixel in den Bereichen von Tiefen-Diskontinuität;  
 Schätzen eines Vordergrund- und eines Hintergrund-Disparitätswertes für jedes Pixel in den Bereichen von Tiefen-Diskontinuität;  
 Einrichten einer Grenzschicht des Bildes, die ein Pixel an jeder entsprechenden Position umfasst, die einem Pixel der Bereiche von Tiefen-Diskontinuität entspricht, wobei jedem Grenzschichtpixel die Vordergrund-Farbe, Vordergrund-Disparität und der Alpha-Wert zugewiesen werden, die mit dem entsprechenden Pixel der Bereiche von Tiefen-Diskontinuität verbunden sind; und  
 Einrichten einer Hauptschicht des Bildes, die umfasst:  
 ein Pixel an jeder entsprechenden Position, die einem Pixel der Bereiche von Tiefen-Diskontinuität entspricht, wobei jedem der Pixel die Hintergrund-Farbe und der Hintergrund-Disparitätswert zugewiesen werden, die mit dem entsprechenden Pixel der Bereiche von Tiefen-Diskontinuität verbunden sind, und  
 ein Pixel an jeder Position, die nicht einem Pixel der Bereiche von Tiefen-Diskontinuität entspricht, wobei jedem der Pixel die Farbe und der Disparitätswert zugewiesen werden, die mit dem entsprechenden Pixel des Bildes verbunden sind.

2. Verfahren nach Anspruch 1, wobei der Verfahrensvorgang des Identifizierens der Position von Tiefen-Diskontinuitäten in dem Bild einen Vorgang des Identifizierens jeder Position, die eine Differenz zwischen Disparitätswerten benachbarter Pixel von mehr als einer vorgeschriebenen Anzahl von Dispari-

tätsstufen aufweist, als eine Tiefen-Diskontinuität.

3. Verfahren nach Anspruch 2, wobei die vorgeschriebene Anzahl von Disparitätsstufen 4 beträgt.

4. Verfahren nach Anspruch 1, wobei die Bereiche vorgeschriebener Größe, die die Tiefen-Diskontinuitäten umgeben, als ein Bereich definiert sind, der sich 3 Pixel in jeder Richtung von einer identifizierten Position von Tiefen-Diskontinuität aus erstreckt.

5. Verfahren nach Anspruch 1, wobei der Verfahrensschritt des Schätzens einer Vordergrund- und einer Hintergrundfarbe sowie eines Alpha-Wertes für jedes Pixel in den Bereichen von Tiefen-Diskontinuität den Schritt des Verwendens einer Matting-Methode zum Schätzen der Farben und Alpha-Werte umfasst.

6. Verfahren nach Anspruch 1, wobei der Verfahrensschritt des Schätzens eines Vordergrund- und eines Hintergrund-Disparitätswertes für jedes Pixel in den Bereichen von Tiefen-Diskontinuität den Schritt des Verwendens alpha-gewichteter Mittelwerte naheliegender Disparitäten in den Vordergrund- und Hintergrundabschnitten des Bildes zum Schätzen der Vordergrund- und Hintergrund-Disparitätswerte umfasst.

7. Verfahren nach Anspruch 6, wobei der Verfahrensschritt des Verwendens alpha-gewichteter Mittelwerte naheliegender Tiefen in den Vordergrund- und Hintergrundabschnitten des Bildes zum Schätzen der Vordergrund- und Hintergrund-Disparitätswerte die folgenden Schritte umfasst:

Berechnen des Vordergrund-Disparitätswertes für jedes Pixel in jedem Bereich von Tiefen-Diskontinuität jeweils durch Multiplizieren jedes Pixels in einem Fenster vorgeschriebener Größe in dem Vordergrundabschnitt des Bildes an das betreffende Pixel angrenzend mit seinem Alpha-Wert und Mitteln der resultierenden Produkte; und

Berechnen des Hintergrund-Disparitätswertes für jedes Pixel in jedem Bereich von Tiefen-Diskontinuität jeweils durch Multiplizieren jedes Pixels in einem Fenster vorgeschriebener Größe in dem Hintergrundbereich des Bildes an das betreffende Pixel angrenzend mit 1 abzüglich seines Alpha-Wertes und Mitteln der resultierenden Produkte.

8. Verfahren nach Anspruch 1, das des Weiteren den Verfahrensvorgang des Erweiterns der Bereiche, die den Grenzschicht-Pixeln entsprechen, um ein vorgegebenes Maß und des Zuweisens der gleichen Farb- und Disparitätswerte zu jedem hinzugefügten Pixel wie dem entsprechenden Pixel in der Hauptschicht sowie eines Alpha-Wertes von 1 umfasst.

9. Verfahren nach Anspruch 8, wobei das vorgegebene Erweiterungsmaß ein Pixel ist.

10. System zum Erzeugen einer zweischichtigen Darstellung eines digitalen oder digitalisierten Bildes aus dem Bild und einer Pixel-Tiefenkarte des Bildes, wobei es umfasst:

eine Universal-Computervorrichtung; und ein Computerprogramm, das durch die Computervorrichtung ausführbare Programmmoduln umfasst, wobei die Computervorrichtung durch die Programmmoduln des Computerprogramms angewiesen wird: die Position von Tiefen-Diskontinuitäten in dem Bild unter Verwendung seiner Tiefenkarte zu identifizieren,

die Pixel in dem Bild zu identifizieren, die in Bereichen vorgeschriebener Größe gefunden werden, die die Tiefen-Diskontinuitäten umgeben, und Kennzeichnen dieser Pixel als in Bereichen von Tiefen-Diskontinuität des Bildes befindlich, eine Vordergrund- und eine Hintergrundfarbe sowie einen Alpha-Wert für jedes Pixel in den Bereichen von Tiefen-Diskontinuität zu schätzen,

einen Vordergrund- und einen Hintergrund-Tiefenwert für jedes Pixel in den Bereichen von Tiefen-Diskontinuität zu schätzen, und

eine Hauptschicht des Bildes, die Pixel, die die Hintergrundfarben und Hintergrundtiefen aufweisen, die mit entsprechend angeordneten Pixeln der Bereiche von Tiefen-Diskontinuität verbunden sind, sowie Pixel umfasst, die die Farben und Tiefen aufweisen, die mit den entsprechend angeordneten Pixeln des Bildes verbunden sind, die nicht in den Bereichen von Tiefen-Diskontinuität gefunden werden, sowie eine Grenzschicht des Bildes zu erzeugen, die Pixel umfasst, die die Vordergrundfarben, Vordergrundtiefen und Alpha-Werte aufweisen, die mit den entsprechend angeordneten Pixeln der Bereiche von Tiefen-Diskontinuität verbunden sind.

11. System nach Anspruch 10, wobei das Programmmodul zum Identifizieren der Position von Tiefen-Diskontinuitäten in dem Bild ein Teilmodul umfasst, das jede Position, die eine Differenz zwischen Tiefenwerten aneinanderliegender Pixel von mehr als einer vorgegebenen Größe aufweist, als Tiefen-Diskontinuität identifiziert.

12. Computerlesbares Medium, das durch Computer ausführbare Befehle zum Erzeugen einer zweischichtigen Darstellung eines digitalen oder digitalisierten Bildes aus dem Bild und einer Pixel-Disparitätskarte des Bildes aufweist, wobei die durch Computer ausführbaren Befehle umfassen:

Identifizieren der Position von Tiefen-Diskontinuitäten in dem Bild unter Verwendung seiner Disparitätskarte;

Einrichten von Bereichen von Tiefen-Diskontinuität vorgegebener Größe, die die identifizierten Tiefen-Diskontinuitäten umgeben;

Schätzen einer Vordergrund- und einer Hintergrundfarbe sowie eines Alpha-Wertes für jedes Pixel in den Bereichen von Tiefen-Diskontinuität;

Schätzen eines Vordergrund- und eines Hintergrund-Disparitätswertes für jedes Pixel in den Bereichen von Tiefen-Diskontinuität; und

Erzeugen einer Hauptschicht des Bildes, die Pixel, die die Hintergrundfarben und Hintergrund-Disparitäten aufweisen, die mit entsprechend angeordneten Pixeln der Bereiche von Tiefen-Diskontinuität verbunden sind, sowie Pixel umfasst, die die Farben und Disparitäten aufweisen, die mit den entsprechend angeordneten Pixeln des Bildes verbunden sind, die nicht in den Bereichen von Tiefen-Diskontinuität gefunden werden, und einer Grenzschicht des Bildes, die Pixel umfasst, die die Vordergrundfarben, Vordergrund-Disparitäten und Alpha-Werte aufweisen, die mit den entsprechend angeordneten Pixeln der Bereiche von Tiefen-Diskontinuität verbunden sind.

13. Zweischichtige Darstellung eines digitalen oder digitalisierten Bildes, die umfasst:  
eine Hauptschicht, die Pixel, die Hintergrundfarben und Hintergrund-Disparitäten aufweisen, die mit allen entsprechend angeordneten Pixeln von Bereichen von Tiefen-Diskontinuität verbunden sind, sowie Pixel umfasst, die Farben und Disparitäten aufweisen, die mit allen entsprechend angeordneten Pixeln des Bildes verbunden sind, die nicht in den Bereichen von Tiefen-Diskontinuität gefunden werden; und  
eine Grenzschicht, die Pixel umfasst, die Vordergrundfarben, Vordergrund-Disparitäten und Alpha-Werte aufweisen, die mit allen entsprechend angeordneten Pixeln der Bereiche von Tiefen-Diskontinuität verbunden sind, wobei  
die Bereiche von Tiefen-Diskontinuität Bereichen vorgeschriebener Größe entsprechen, die Tiefen-Diskontinuitäten umgeben, die in dem Bild unter Verwendung einer Disparitätskarte des Bildes gefunden werden.

Es folgen 4 Blatt Zeichnungen

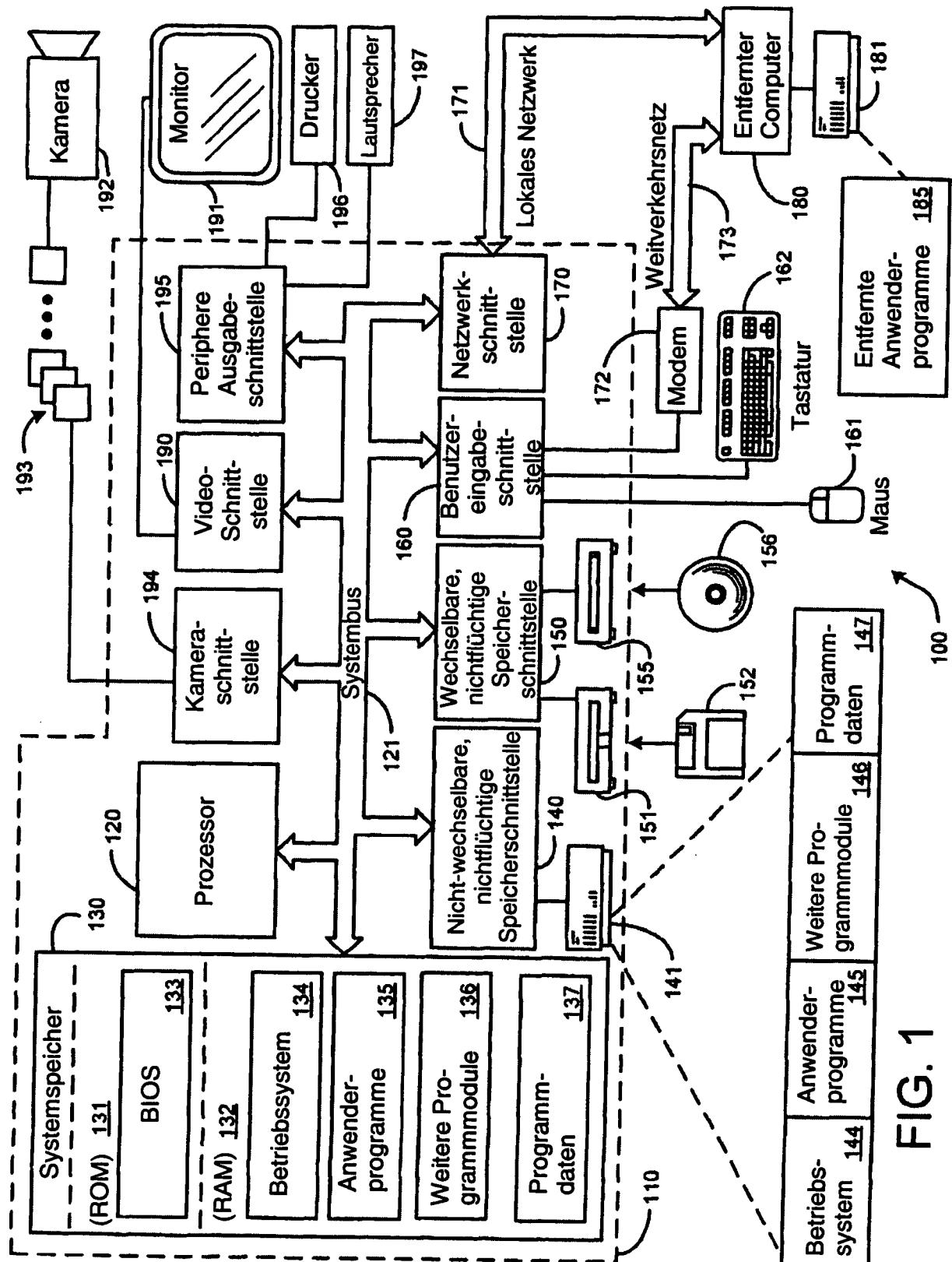


FIG.

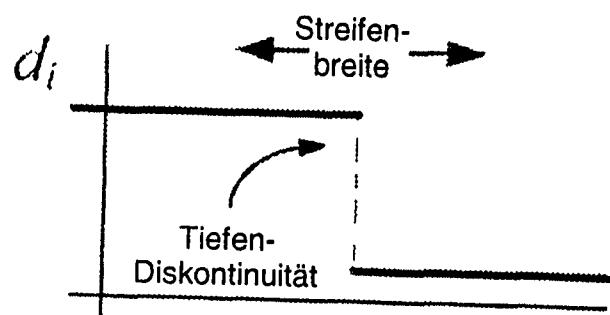


FIG. 2

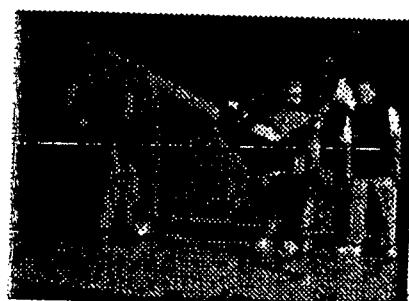


FIG. 4(a)

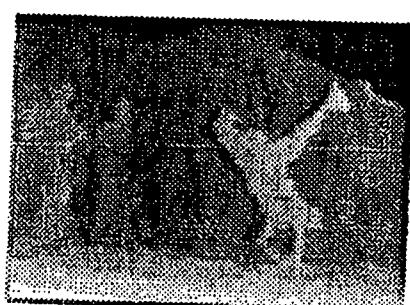


FIG. 4(b)



FIG. 4(c)



FIG. 4(d)



FIG. 4(e)

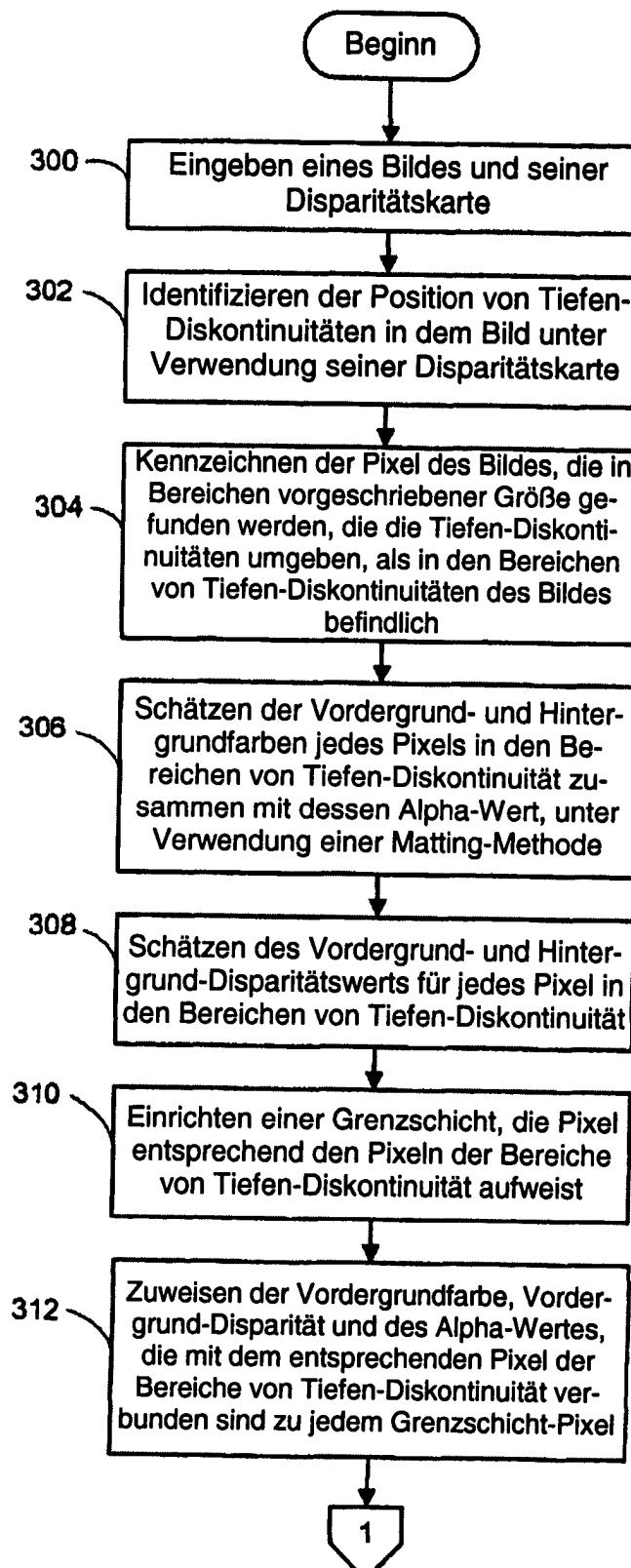
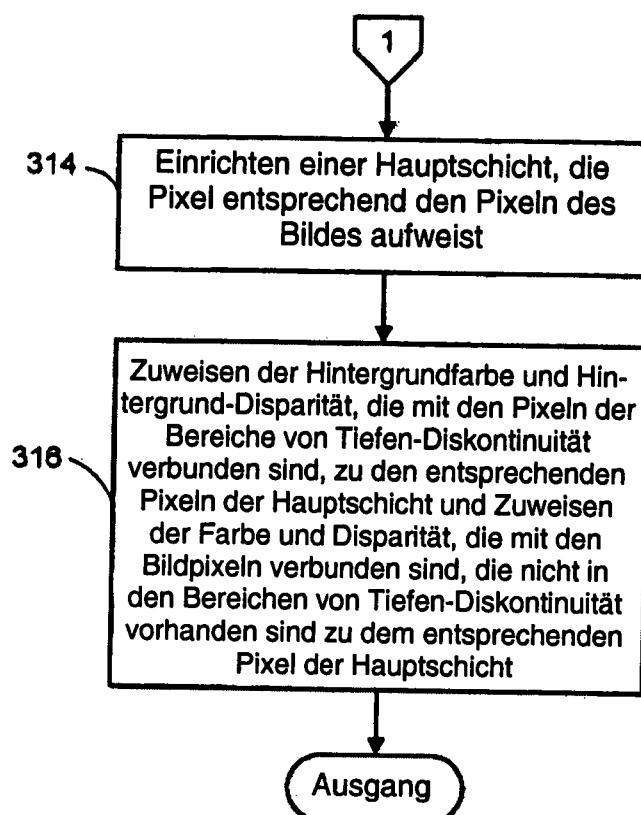


FIG. 3A



**FIG. 3B**