



(19)
Bundesrepublik Deutschland
Deutsches Patent- und Markenamt

(10) **DE 698 31 181 T2** 2006.05.18

(12) **Übersetzung der europäischen Patentschrift**

(97) **EP 1 211 520 B1**

(21) Deutsches Aktenzeichen: **698 31 181.7**

(96) Europäisches Aktenzeichen: **02 004 163.8**

(96) Europäischer Anmeldetag: **29.05.1998**

(97) Erstveröffentlichung durch das EPA: **05.06.2002**

(97) Veröffentlichungstag

der Patenterteilung beim EPA: **10.08.2005**

(47) Veröffentlichungstag im Patentblatt: **18.05.2006**

(51) Int Cl.⁸: **G01S 5/16** (2006.01)

G01S 17/87 (2006.01)

H04N 5/272 (2006.01)

(30) Unionspriorität:

9711316 **30.05.1997** **GB**

9711373 **02.06.1997** **GB**

(73) Patentinhaber:

British Broadcasting Corp., London, GB

(74) Vertreter:

Schwabe, Sandmair, Marx, 81677 München

(84) Benannte Vertragsstaaten:

BE, DE, FR, GB, IT

(72) Erfinder:

**Russell, Richard Thomas, Orsett, Essex RM16
3DF, GB; Thomas, Graham Alexander, Uckfield,
East Sussex TN22 3HA, GB**

(54) Bezeichnung: **Positionsbestimmung**

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 eingelegt, 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

[0001] Die vorliegende Erfindung betrifft eine Positionsbestimmung und betrifft insbesondere, aber nicht ausschließlich, die Ableitung der Position einer Kamera für eine so genannte "Virtual Studio"-Fernsehproduktion. In solchen Anwendungen wird eine virtuelle Szene, beispielsweise ein mit einem Computer generierter Hintergrund, einer realen Szene überlagert, beispielsweise mit Schauspielern. Es ist wichtig, die exakte Position und Orientierung der Fernsehkamera in dem Studio zu kennen, so dass die virtuelle Szene so berechnet werden kann, dass diese dem Gesichtspunkt der realen Kamera entspricht, um eine korrekte Übereinstimmung zwischen realen und virtuellen Elementen in der Szene sicherzustellen.

[0002] Es gibt eine Vielzahl von kommerziell erhältlichen Systemen, die Information über die Position einer Kamera bereitstellen können. Die meisten basieren auf mechanischen Kamerahalterungen, wie beispielsweise robotischen Sockeln, Sockeln, die auf Schienen montiert sind, oder Roboterarmen, die Sensoren aufweisen, um die Position und Orientierung zu messen. Diese Systeme können nicht für von Hand gehaltene Kameras verwendet werden und können voluminös und schwierig in der Verwendung sein.

[0003] Es gibt auch Verfahren, die mit mechanischen Sensoren arbeiten, aber stattdessen einen mit einem Muster versehenen blauen Hintergrund verwenden, der in dem Kamerabild sichtbar ist. Durch Analysieren des Videosignals können diese Verfahren auf die Orientierung und Position der Kamera schließen. Ein Beispiel für ein solches Verfahren ist in unserer früheren Patentanmeldung GB-A-2271241 beschrieben, die einen Kameraschwenk, eine Kameraverkipfung und einen Zoomfaktor unter Verwendung eines beliebig gemusterten Hintergrunds ableitet. Ein weiteres Beispiel, das dieselbe Technik einsetzt, ist in WO-A-95/30312 beschrieben, die einen speziellen Mustertyp verwendet, um eine Bestimmung des Kameraschwenks, der Verkipfung, des Zooms und der Position zu ermöglichen. Diese Verfahren beruhen jedoch auf dem Vorhandensein eines blauen Hintergrunds mit zwei Farbtönen. Dies ist in gewissen Situationen unzuverlässig, beispielsweise dann, wenn es gewünscht ist, Schatten von Objekten aus dem Vordergrundbild zu extrahieren. In gewissen Situationen kann auch nur sehr wenig oder gar kein blauer Hintergrund sichtbar sein, beispielsweise während einer Nahaufnahme eines Schauspielers. Manchmal kann es auch erforderlich sein, ein virtuelles Objekt gegen einen realen Hintergrund zu platzieren, in welchem Fall es überhaupt keinen blauen Hintergrund geben wird.

[0004] WO-A-9711386 offenbart eine Vorrichtung zur Bestimmung einer Position und Orientierung, bei

der eine Kamera auf ein optisch moduliertes Zielobjekt gerichtet wird.

[0005] EP-A-706105 offenbart ein Navigationssystem für einen autonomen mobilen Roboter, bei dem codierte Zeichen an verschiedenen Positionen angeordnet werden. Die Marken werden anhand des Verhältnisses der Radien von Ringen unterschieden.

[0006] Ein Verfahren zur Bestimmung der Position eines aufrecht montierten Displays, das auf dem Gebiet der verbesserten Wirklichkeit (augmented reality) eingesetzt worden ist, wird in dem Artikel von Azuma u. a. mit dem Titel "A Demonstrated Optical Tracker with Scalable Work Area for Head-Mounted Display Systems", beschrieben, veröffentlicht in ACM Computer Graphics: Proceedings of the 1992 Symposium on Interactive 3D Graphics (Cambridge, Massachusetts, April 1992), S. 43–52. Dieses Verfahren verwendet eine Mehrzahl von infraroten LEDs, die in Deckenpaneelen eingebaut sind, die mit vier nach oben schauenden Sensoren betrachtet werden, die auf dem Kopfhörer eines Nutzers montiert sind. Die Sensoren stellen jeweils Information der Koordinaten von hellen Punkten in dem Bild (die LEDs) bereit und aus diesen Koordinaten wird anhand der bekannten Position der LEDs und der Geometrie der Sensoren die Position des Kopfhörers berechnet. Von den Erfindern wurde angedacht, das Verfahren von Azuma auf das Problem der Bestimmung einer Kameraposition anzuwenden. Das Verfahren von Azuma ist jedoch nicht dafür gedacht, um zur Bestimmung der Position einer Kamera in einem großen Studio verwendet zu werden, sondern ist stattdessen ausgelegt, um in einem vergleichsweise kleinen Volumen zu arbeiten, wo die Gesamtzahl von LEDs klein ist. Die Erfinder haben einige potenzielle Probleme bei der Anwendung einer solchen Technik auf das Gebiet der Kamerapositionsbestimmung identifiziert.

[0007] In einem Fernsehstudio könnte die Kamera sich mehrere zehn Meter bewegen, so dass das Verfahren von Azuma u. a. eine sehr große Anzahl von Deckenmarken erfordern würde, von denen nur ein kleiner Teil zu irgendeinem Zeitpunkt sichtbar wäre. Das Verfahren von Azuma beruht auf einer aktiven Kontrolle der LEDs, um die LEDs zu identifizieren. Die Kontrollelektronik und -verdrahtung, die erforderlich sind, um dies in einem großen Studio umzusetzen, wäre komplex und nicht praktikabel. Außerdem enthält die Decke eines Fernsehstudios für gewöhnlich eine Vielzahl von Lampen und anderen Objekten, die die Identifizierung der Marken viel schwieriger machen könnten als in den sorgfältig kontrollierten Bedingungen, für die das Verfahren von Azuma ausgelegt ist, zu funktionieren; die Hellpunktsensoren, die von Azuma verwendet werden, würden falsche Signale erzeugen. Ein weiteres Problem, das die Erfinder identifiziert haben, besteht darin, dass der Satz von Marken, welche die Kamera sehen kann, sich

nicht nur deshalb ändern wird, weil sich die Marken in das Gesichtsfeld der Kamera hinein bewegen und aus diesem heraus bewegen, sondern auch deshalb, weil die Marken durch Objekte, wie beispielsweise Mikrofonarme und Studioleuchten, verdeckt werden. Bei Verwendung des Verfahrens von Azuma ist es jedes Mal dann, wenn eine Marke auftaucht oder verschwindet, wahrscheinlich, dass es eine kleine aber plötzliche Änderung in der berechneten Position und Orientierung der Kamera geben wird, weil jegliche Fehler in der Kamerakalibrierung oder der Messung der Markerpositionen zu Ergebnissen führen werden, die davon abhängen, welche Marken gerade verwendet werden.

[0008] Somit kann das Verfahren von Azuma nicht direkt auf das erfindungsgemäße Problem angewendet werden. Die Erfinder haben neue Techniken entwickelt, die besonders geeignet für die anspruchsvollen Anforderungen an die Bestimmung der Kameraposition in einem Fernsehstudio geeignet sind, welche die Nachteile konventioneller Techniken überwinden oder mindern. Verweise in dieser Patentbeschreibung auf eine Videokamera sollen jegliche Kamera beinhalten, die in der Lage ist, Videobilder zu verfassen; obwohl die Videokamera vorteilhaft eine konventionelle Fernsehstudiokamera ist, ist keine Beschränkung auf einen speziellen Typ von Kamera oder auf einen beabsichtigten Einsatzzweck impliziert.

[0009] Gesichtspunkte der Erfindung werden in den unabhängigen Patentansprüchen dargelegt. Bevorzugte Merkmale werden in den abhängigen Patentansprüchen dargelegt.

[0010] Ein Gesichtspunkt der Erfindung stellt ein Verfahren zur Bestimmung der Position eines Objekts, relativ zu einer Vielzahl von Referenzpositionen bereit, basierend auf der Identifizierung einer entsprechenden Vielzahl von Markierungen an jenen Referenzpositionen, wobei das Verfahren die Anwendung von Korrekturfaktoren auf gespeicherte oder gemessene Positionen der Markierungen relativ zu dem Objekt umfasst, um entsprechende, korrigierte Markierungspositionen zu erzeugen, und um die Objektposition relativ zu den Referenzpositionen, basierend auf den korrigierten Referenzpositionen, zu ermitteln, wobei die Korrekturfaktoren auf der Basis der Differenz zwischen gemessenen Markierungspositionen und erwarteten Markierungspositionen berechnet werden, so dass die korrigierten Markierungspositionen korrigiert werden, um gegenseitig selbstkonsistente Werte zu werden, wobei die Korrekturfaktoren zeitlich Tiefpass gefiltert werden, um die Rate der Änderungen von Korrekturfaktoren zu reduzieren.

[0011] Das Verfahren kann den Schritt einer Anpassung der bestimmten Position umfassen, um plötzliche Änderungen der bestimmten Position auszuglät-

ten, wenn Marken zum Vorschein kommen oder verdeckt werden. Dies kann das Problem von kleinen, aber dennoch plötzlichen und sehr sichtbaren Änderungen in der bestimmten Objektposition auf Grund von kleinen Fehlern in der Positionsbestimmung der Marken mindern, wenn sich verschiedene Marken in das Gesichtsfeld der Kamera hinein bewegen oder aus diesen heraus bewegen.

[0012] Eine Anpassung kann dadurch erzielt werden, dass eine Bewegung der Marken verfolgt wird, wenn sich die Objektposition ändert, und dass dreidimensionale Korrekturvektoren erzeugt werden. Auf diese Weise können stabile genauere Markenpositionen bestimmt werden. Eine Verfolgung der Marken über mehrere Bilder kann jedoch rechenaufwändig sein und dabei versagen, stabile Werte für die Markenpositionen zu erzeugen, wenn Fehler auf irgendeine Nichtlinearität in der Kamera oder in dem Objektiv zurückgehen. Somit umfasst eine bevorzugte einfachere Realisierung der Anpassung den Schritt, dass Korrekturfaktoren mit zwei Freiheitsgraden erzeugt werden (beispielsweise Verschiebungsvektoren parallel zu der Bildebene oder parallel zu der Referenzoberfläche) und dass die Markenposition korrigiert werden, so dass diese dazu neigen, die Markenpositionen selbstkonsistent für ein Bild oder für eine Folge von Bildern zu machen.

[0013] Die Änderungsrate kann beschränkt sein, so dass eine Anpassung der Korrekturfaktoren über eine Periode von einigen Sekunden erfolgt (vorzugsweise von zumindest 2 Sekunden, vorzugsweise von weniger als etwa 15 Sekunden, typischerweise etwa 5–10 Sekunden), so dass eine allmähliche Drift in der bestimmten Position auftritt, die weniger gut wahrgenommen werden kann wie eine plötzliche Verschiebung.

[0014] Ein Maß für den Fehler oder die Genauigkeit der Positionsbestimmung kann auf der Grundlage der Selbstkonsistenz der bestimmten Position, die für jede Marke berechnet wird, bereitgestellt werden. Diese kann auf der Größe der Korrekturfaktoren basieren.

[0015] Obwohl die Marken an einer ebenen Referenzoberfläche anhaften können, sind die Marken vorzugsweise unter variierenden Abständen zu einer Referenzoberfläche positioniert, vorzugsweise der Decke eines Raums, in welchem sich das Objekt bewegt; dies hat sich als eine überraschende Verbesserung der Genauigkeit der Detektion der Bewegung des Objekts herausgestellt. Insbesondere in einem solchen Fall enthält die gespeicherte Information vorzugsweise ein Maß für den Abstand von jeder Marke von der Referenzoberfläche. Dies kann dadurch erzielt werden, dass die dreidimensionalen Koordinaten von jeder Marke relativ zu einem definierten Ursprung gespeichert werden (in kartesischer oder ir-

gendeiner polaren oder krummlinigen Form, welche für den Algorithmus zur Positionsbestimmung geeignet ist).

[0016] Vorzugsweise wird von einer Kamera, die mit dem Objekt in Verbindung gebracht wird, ein Bild von mindestens einer Teilmenge der besagten Vielzahl von Markierungen erhalten.

[0017] Vorzugsweise ist das besagte Objekt eine Videokamera, die ein Bildfeld bzw. Blickfeld aufweist, welches vom Bildfeld bzw. Gesichtsfeld der besagten Kameramittel abweicht.

[0018] Bei der bevorzugten Anwendung ist das Objekt eine Videokamera und umfasst das Kameramittel eine separate, für gewöhnlich nach oben gerichtete Kamera, die auf der Videokamera montiert ist.

[0019] Bevorzugter stellt das Verfahren ferner zumindest eine zusätzliche Markierung an einer vorgegebenen Position bereit, so dass diese für die Videokamera sichtbar ist, wobei die oder jede Markierung Identifizierungsinformationen aufweist, die darin in binärer Form als Folge von konzentrischen Ringen von zwei Farbtönen codiert sind, wobei die Farbtöne so gewählt werden, dass die Markierung durch Einblendung bzw. Chroma-Keying aus dem Videokamerabild ausgetastet werden kann, wobei die Bestimmung der Position der Videokamera auf der Position der zusätzlichen Markierung, falls diese vorhanden ist, in dem Videokamerabild basiert.

[0020] Eine weitere Ausführungsform stellt eine Vorrichtung zur Positionsbestimmung bereit, umfassend eine Kamera zur Montage auf einem Objekt, dessen Position bestimmt werden soll; ein Speichermittel, das ausgelegt ist, um Maßgrößen der Positionen einer Mehrzahl von Marken zu speichern und um Information, die bemusterte codierte Marken identifiziert, zu speichern; ein Bildverarbeitungsmittel, das ausgelegt ist, um ein von der Kamera ausgegebenes Bild zu verarbeiten, um die Positionen der Marken in dem Bild zu identifizieren; Decodierungsmittel, um Information, die in den bemusterten Marken codiert ist, zu decodieren; und ein Positionsbestimmungsmittel, das ausgelegt ist, um die Position des Objekts auf der Grundlage des Ausgangssignals des Bildverarbeitungsmittels, des Decodierungsmittels und auf der Grundlage der in dem Speichermittel gespeicherten Information zu bestimmen.

[0021] Das Speichermittel, das Bildverarbeitungsmittel und das Positionsbestimmungsmittel können in einem einzigen Computer integriert sein. Vorzugsweise ist jedoch zumindest ein Teil der Funktion des Bildverarbeitungsmittels mit Hilfe eines Hardwarebeschleunigers bereitgestellt, der ausgelegt ist, um eine oder mehrere Marken zu identifizieren oder um eine oder mehrere bemusterte Marken zu decodieren

oder um beide Funktionen auszuführen.

[0022] Die Vorrichtung umfasst vorzugsweise außerdem ein Mittel zum Anwenden von Korrekturfaktoren auf gespeicherte Maßgrößen bzw. Messwerte der Markenpositionen, um jeweils korrigierte Markenpositionen zu erzeugen und um die Objektposition auf der Grundlage der korrigierten Markenpositionen zu bestimmen, wobei die Korrekturfaktoren so berechnet werden, dass die korrigierten Markenpositionen dazu neigen, zueinander selbstkonsistente Werte anzunehmen, wobei die Änderungsrate der Korrekturfaktoren auf einen vorbestimmten Wert begrenzt ist.

[0023] In Fällen, wo die Position einer Videokamera (eine andere als das Kameramittel) bestimmt wird, kann die Positionsbestimmung durch Information unterstützt werden, die von dem Bild der Videokamera erhalten wird. Außerdem kann Information, welche den Zoom der Videokamera und Fokuseinstellungen betrifft, eingegeben werden. Eine oder mehrere codierte Marken können in dem Gesichtsfeld der Videokamera platziert werden. Die Anordnung von codierten Marken bei speziellen Positionen in dem Gesichtsfeld ist viel einfacher in der Praxis zu realisieren als die Bereitstellung eines mit einem Muster versehenen Hintergrunds für ein gesamtes Studio, welches Verfahren im Stand der Technik verwendet wurde. Außerdem können die codierten Marken eine genauere Messung als im Falle eines konventionellen Hintergrunds ermöglichen, wo es schwierig ist, für eine genaue Bemusterung über einen großen Bereich zu sorgen.

[0024] Somit kann das Verfahren außerdem die Schritte umfassen, dass zumindest eine zusätzliche Marke bei einer vorgegebenen Position bereitgestellt wird, so dass diese von der Videokamera gesehen werden kann, wobei die oder jede Marke eine Identifizierungsinformation aufweist, die darauf in binärer Form als Folge von konzentrischen Ringen von zwei Farbtönen codiert ist, wobei die Farbtöne ausgewählt werden, so dass die Marke aus dem Bild der Videokamera mit Hilfe einer Einblendung bzw. Chroma-Keying ausgetastet werden kann, wobei die Positionsbestimmung der Videokamera auf der Position der zusätzlichen Marke beruht, falls diese in dem Bild der Videokamera vorhanden ist.

[0025] Für die zusätzliche Positionierung kann unabhängig in einem Verfahren zur Bestimmung oder Korrektur einer Maßgröße oder eines Messwerts der Position einer Videokamera gesorgt werden, welches Verfahren die Schritte umfasst, dass zumindest eine Marke bei einer Position bereitgestellt wird, die von der Kamera gesehen werden kann, wobei die oder jede Marke eine Identifizierungsinformation aufweist, die darauf in binärer Form als Folge von konzentrischen Ringen mit zwei Farbtönen codiert ist, wobei

die Farbtöne so gewählt werden, dass die Marke aus dem Bild der Videokamera mit Hilfe einer Einblendung bzw. Chroma-Keying ausgetastet werden kann, wobei eine Maßgröße bzw. ein Messwert der Position der Videokamera auf der Grundlage der Position der Marke, falls diese in dem Bild der Videokamera vorhanden ist, bestimmt wird. Eine vollständige Positionsbestimmung kann dadurch bewerkstelligt werden, dass mehrere solche Markierungen so verwendet werden, dass vorzugsweise zumindest drei einzelne Marken stets sichtbar sind, oder dadurch, dass eine mechanische Abtastung oder andere optische Verfahren verwendet werden, wo die codierten Marken für einen Prüf- bzw. Referenzpunkt sorgen. Vorzugsweise ist ein Satz von vorzugsweise zumindest drei einzelnen Marken auf einer im Wesentlichen steifen Halterung montiert, um ein Cluster von Marken zu bilden, wobei die Marken vorzugsweise nicht in einer Ebene liegen.

[0026] Eine Ausführungsform wird nun ausführlicher unter Bezugnahme auf die beigefügten Figuren beschrieben werden, worin:

[0027] [Fig. 1](#) ein Beispiel einer Studioanordnung zeigt;

[0028] [Fig. 2](#) ein Beispiel einer Marke zeigt;

[0029] [Fig. 3](#) ein Schemadiagramm eines Algorithmus zeigt, der dazu verwendet werden kann, um Marken in dem Bild zu lokalisieren;

[0030] [Fig. 4](#) die Öffnung der Filter zeigt, die in dem Markendetektionssystem verwendet werden;

[0031] [Fig. 5](#) im Detail das Verfahren zum Zählen der Anzahl von Proben oberhalb eines vorgegebenen Schwellenwerts in einem rechteckigen Bildbereich zeigt;

[0032] [Fig. 6](#) das Verfahren gemäß einem Gesichtspunkt der vorliegenden Erfindung zeigt;

[0033] [Fig. 7](#) eine Marke darstellt, die dazu geeignet ist, um in dem Gesichtsfeld einer Videokamera platziert zu werden.

[0034] Die [Fig. 1](#) bis [Fig. 5](#) und die ihnen zugeordnete Beschreibung umreißen als Hintergrund eine Ausführungsform eines Systems, in dem die vorliegende Erfindung realisiert werden kann. Die [Fig. 6](#) und [Fig. 7](#) und die zugeordnete Beschreibung umreißen eine Ausführungsform der vorliegenden Erfindung.

[0035] Eine Ausführungsform der vorliegenden Erfindung kann in einem virtuellen Studio angewendet werden, wie es in der [Fig. 1](#) gezeigt ist. Eine Anzahl von Marken **1** sind auf der Studiodecke befestigt.

Weitere Marken sind in einem geringen Abstand unterhalb der Decke befestigt, und zwar sind diese an steifen Stangen **2** befestigt. Obwohl das System mit Marken in einer Ebene arbeiten kann, kann man eine größere Genauigkeit mit Marken erzielen, die auf unterschiedlichen Höhen angeordnet sind, wie dies dargestellt ist. Das System kann alternativ mit Hilfe von Marken realisiert werden, die an anderen Positionen als an der Decke angeordnet sind (beispielsweise auf einer Wand oder sogar auf dem Boden), obwohl die Anordnung von diesen auf der Decke wahrscheinlich am praktikabelsten in vielen Situationen ist.

[0036] Die Marken werden von einer nach oben schauenden Kamera **3** gesehen, die auf der Seite der normalen Studiokamera **4** befestigt ist. Die nach oben schauende Kamera, die als zusätzliche Kamera bezeichnet wird, ist vorzugsweise eine monochrome Kamera und wird vorzugsweise mit dem Zeilenfolgeverfahren abgetastet. Die Marken können von einem Licht **12** beleuchtet werden. Das Signal von der Kamera wird von einer Analyseeinrichtung **5** analysiert, um die Kameraposition und -orientierung der zusätzlichen Kamera zu berechnen. Die Analyseeinrichtung berechnet dann die Position und Orientierung **6** der Studiokamera, wobei die relative Orientierung und Position der zusätzlichen Kamera **3** und der Studiokamera **4** ausgeglichen wird.

[0037] Die Analyseeinrichtung kann unter Verwendung entweder von spezieller Hardware oder durch Kombination eines Frame-Grabbers und eines Mehrzweckcomputers realisiert werden. Information bezüglich der Einstellungen von Zoom und Fokus der Studiokamera kann unter Verwendung von konventionellen Sensoren erhalten werden und mit Hilfe einer Verzögerungseinrichtung **8** verzögert werden, um die Verzögerung durch die Verarbeitungseinrichtung **5** zu kompensieren. Die Darstellungseinrichtung **9** berechnet die Ansicht der virtuellen Welt entsprechend der gemessenen Positions-, Orientierungs- und Zoom-Werte. Solche Darstellungseinrichtungen sind kommerziell erhältlich und basieren für gewöhnlich auf Graphikcomputern. Das normale Videosignal von der Kamera wird mit Hilfe der Verzögerungseinrichtung **10** verzögert, um die Verzögerungen bei der Berechnung der Kameraposition und bei der Bilddarstellung auszugleichen, bevor dieses mit dem dargestellten Bild unter Verwendung der herkömmlichen Tasteinrichtung **11** verknüpft wird.

[0038] Ein Beispiel einer Marke, die in einer Ausführungsform der vorliegenden Erfindung verwendet werden kann, ist in der [Fig. 2](#) gezeigt. Die gezeigte Marke weist helle konzentrische Ringe auf einem dunklen Hintergrund auf. Die hellen Teile bestehend vorzugsweise aus einem rückstrahlenden Material, das von einem Licht nahe der Kamera beleuchtet wird, die die Marken sieht, wie dies in der [Fig. 1](#) gezeigt ist. Eine geeignete Form von rückstrahlendem

Material ist der "reflektierende Silberadhäsionsübertragungsfilm", Teil Nr. 8850, hergestellt von 3M. Bei dem in der [Fig. 2](#) gezeigten Beispiel ist die Marke in eine Anzahl von konzentrischen Ringen mit gleicher Breite unterteilt. Der äußere Ring ist stets weiß und dient dazu, den Rand der Marke zu begrenzen. Jeder innere Ring ist entweder dunkel oder hell, um eine "1" oder "0" in dem entsprechenden Bit der Markennummer anzuzeigen. Andere Arten von Streifencode, beispielsweise basierend auf dem Abstand zwischen Übergängen anstelle der Helligkeit von bestimmten Bändern, sind ebenfalls möglich. Das Beispiel in der [Fig. 2](#) hat sechs innere Ringe, was einen 6-Bit-Code ergibt, der Zahlen in dem Bereich von 0–63 repräsentieren kann. Es wird bevorzugt, dass nicht sämtliche möglichen Codenummern verwendet werden, was eine gewisse Redundanz ermöglicht. Beispielsweise könnte ein Bit als Paritätsbit verwendet werden, so dass ein Fehler beim Lesen von einem Bit detektiert werden kann. Es wird auch bevorzugt, dass nicht der Code verwendet wird, dessen Darstellung ein massiver weißer Kreis (**63** in diesem Fall) ist, weil andere Objekte, beispielsweise eine Studioluchte vor einem schwarzen Hintergrund, fälschlicherweise als Marke interpretiert werden könnten.

[0039] Die Positionierung der Marken für eine typische Anwendung wird nun beschrieben. Für Studios mit Bodenflächen von bis zu etwa 50 qm reicht ein 9-Bit-Streifencode aus, um jede erforderliche Marke eindeutig zu identifizieren. Für größere Studios können gewisse Markennummern wieder verwendet werden und das System kann darauf schließen, bei welcher Marke es sich um welche Marke handelt, indem beispielsweise auf die nächste eindeutig bezifferte Marke Bezug genommen wird. Die minimale Größe der Marken muss gewählt werden, um zu gewährleisten, dass der Streifencode mit Hilfe der zusätzlichen Kamera erfolgreich gelesen werden kann. Der erfindungsgemäße Streifencodeentwurf und Lesealgorithmus erfordern vorzugsweise einen minimalen Abstand zwischen den Ringen, der etwa 1,5 Pixel auf dem Kamerasensor entspricht, um ein zuverlässiges Auslesen zu erlauben. Falls sich beispielsweise die Marken etwa 4 m oberhalb der zusätzlichen Kamera befinden, jede in einem 9-Bit-Streifencode enthalten, und falls eine Kamera mit herkömmlicher Auflösung mit einem minimalen Blickfeld von 30 Grad verwendet wird, beträgt der minimale Markendurchmesser etwa 12 cm. Die Marken sollten so positioniert sein, dass die zusätzliche Kamera mindestens 3 und Idealerweise **10** oder mehr gleichzeitig sehen kann. Es muss möglich sein, dass gewisse Marken von Leuchten, Mikrofonhaltern usw. verdeckt werden. Beispielsweise würde in einer Situation mit Marken auf Höhen von 4 und 4,5 m, wenn eine maximale Arbeitshöhe der Kamera 2 m beträgt und das minimale Blickfeld der zusätzlichen Kamera **30** Grad beträgt, eine Positionierung der Marken in einem quadratischen Gitter mit Intervallen von 0,4 m gewähr-

leisten, dass mindestens 12 potenziell sichtbar sind, was ein gutes Leistungsverhalten ermöglichen sollte, selbst wenn die Hälfte von diesen von Hindernissen, beispielsweise von Leuchten, verdeckt werden. Obwohl die Marken so positioniert werden können, dass diese in einer Ebene liegen, wird die Genauigkeit der berechneten Kameraposition deutlich verbessert, falls die Marken unter verschiedenen Abständen zu der Kamera angeordnet werden, beispielsweise könnten gewisse Marken direkt an der Decke eines Studios angebracht werden, wobei andere auf Balken oder Stangen montiert etwas herunterhängen. Die Position von jeder Marke sollte idealerweise bis auf eine Genauigkeit von etwa 1 mm bekannt sein, um die Gesamtanforderungen an die Genauigkeit zu erfüllen. Dies könnte dadurch erzielt werden, dass die Position der Marken unter Verwendung eines Theodolits gemessen wird, obwohl dies zeitaufwändig wäre. Ein alternatives Verfahren ist entwickelt worden. Zunächst werden die Positionen der Marken grob gemessen, mit einer Genauigkeit von einigen Zentimetern. Eine Anzahl von Bildern der Marken werden dann analysiert, um genaue Positionen zu ergeben.

[0040] Die erste Aufgabe, welche die Analysevorrichtung **5** in der [Fig. 1](#) ausführt, besteht darin, mögliche Marken in dem Bild zu identifizieren. Das in der [Fig. 2](#) gezeigte Aussehen der Marke ermöglicht es, dass dies mit Hilfe eines einfachen Algorithmus bewerkstelligt werden kann, der vergleichsweise wenig Rechenzeit erfordert. Ein schematisches Diagramm eines solchen Algorithmus ist in der [Fig. 3](#) gezeigt. Dieser basiert auf der Verwendung von Filtern mit rechteckigen Öffnungen, die unter Verwendung von eindimensionalen, horizontalen und vertikalen Filtern realisiert werden können. Man beachte, dass aus Gründen der Übersichtlichkeit die [Fig. 3](#) nicht irgendwelche zusätzlichen Verzögerungen zeigt, um die Unterschiede in der Verarbeitungsverzögerung der verschiedenen Elemente zu kompensieren. Ein Beispiel der verwendeten Öffnungen ist in der [Fig. 4](#) gezeigt. Das innerste Rechteck A wird gewählt, so dass dieses groß genug ist, um sämtliche Ringe der Marke komplett zu überdecken, und zwar für die größte erwartete Größe einer Marke in dem Bild. Das äußere Rechteck B wird so gewählt, dass es zu der schwarzen äußeren Umgebung der Marke passt, und zwar für die kleinste erwartete Größe einer Marke in dem Bild.

[0041] Nun wird die in der [Fig. 3](#) gezeigte Betriebsweise des Markendetektors im Detail beschrieben. Das Videosignal von der nach oben schauenden Kamera wird in einen Probenzähler **20** eingegeben, der die Anzahl von Proben oberhalb eines Schwellenwertes T1 in einem Rechteck A zählt, das auf dem aktuellen Pixel zentriert ist. Dieser Schwellenwert wird so gewählt, dass er einem Helligkeitswert zwischen den Werten von schwarz und weiß entspricht,

wie diese auf dem Zielobjekt auftreten. Dies kann unter Verwendung der in der [Fig. 5](#) gezeigten Schaltung realisiert werden, die nun beschrieben wird. Ein Komparator **40** vergleicht das Eingangssignal **39** mit einem Schwellenwert T und erzeugt ein Signal **41**, das für Proben, deren Werte größer oder gleich T sind, 1 ist und 0 für Proben, die kleiner sind als T . Dieses Signal wird einem vertikalen Summierer **46** zugeführt, der die Summe seiner Eingangsproben über eine vertikale Öffnung der Höhe " h " berechnet. Dieses Signal wird dann an einen horizontalen Summierer **47** weitergeleitet, der eine vergleichbare Operation für eine horizontale Öffnung einer Breite " w " vornimmt. Summierer wie diese werden in vielen Bildverarbeitungsoperationen verwendet, aber aus Gründen der Vollständigkeit wird deren Betriebsweise beschrieben. Die folgende Beschreibung bezieht sich auf den vertikalen Summierer; der horizontale Summierer arbeitet in ähnlicher Weise. Die Verzögerung **42** verzögert das Signal um die Anzahl von Zeilen, die gleich der Höhe des Rechtecks A ist. Der Subtrahierer **43** bildet die Differenz zwischen dem Eingang und dem Ausgang der Verzögerung. Die Zeilenverzögerung **44** und der Addierer **45** bilden einen Integrierer, der das Ausgangssignal des Subtrahierers **43** integriert. Wenn Proben in den Bereich mit einer Höhe von h Zeilen eintreten, werden diese in dem Integrator integriert, und wenn diese den Bereich verlassen, werden diese von dem integrierten Wert subtrahiert. Das Ausgangssignal des Addierers **45** wird somit gleich der Anzahl von Proben oberhalb des Schwellenwertes T in einer Spalte von Pixeln mit einer Höhe von h Zeilen. In ähnlicher Weise ist das Ausgangssignal des horizontalen Summierers **48** gleich der Anzahl von Proben, deren Werte gleich dem oder größer dem Schwellenwert T über einem rechteckigen Bereich von h Zeilen mal w Pixeln sind.

[0042] Das Ausgangssignal des Probenzählers **20** in der [Fig. 3](#) wird einem Komparator **24** zugeführt. Der Komparator erzeugt ein "wahres" Ausgangssignal, wann immer es mehr als $N1$ Proben in dem Rechteck A über dem Schwellenwert $T1$ gibt. Die Werte von $T1$ und $N1$ werden so gewählt, dass diese Bedingung erfüllt wird, wann immer eine Marke sich innerhalb des Rechtecks A befindet. Eine ähnliche Verarbeitung wird dazu verwendet, um zu überprüfen, ob die Marke von einem schwarzen Rand umgeben ist. Der Probenzähler **21** zählt die Anzahl von Proben oberhalb eines Schwellenwertes $T2$ in dem Rechteck A . $T2$ ist kleiner als $T1$ und wird so gewählt, um gerade oberhalb der nominellen Helligkeit des schwarzen von dem Hintergrund zu sein. Der Probenzähler **22** führt dieselbe Aufgabe für das Rechteck B aus. Die Differenz zwischen den zwei Werten entspricht der Anzahl von Proben oberhalb des Schwellenwertes $T2$ in dem Bereich zwischen den Rechtecken A und B in der [Fig. 4](#). Der Komparator **25** testet, ob diese Zahl kleiner ist als $N2$, welche eine kleinere Zahl als im Vergleich zu der Anzahl von Pi-

xeln in diesem Bereich sein sollte. Sein Ausgangssignal wird "wahr", wann immer diese Bedingung erfüllt ist.

[0043] Das mögliche Vorhandensein einer Marke wird angedeutet, wenn die Ausgangssignale der Komponenten **24** und **25** beide "wahr" sind. Dies wird jedoch wahrscheinlich für viele benachbarte Probenpositionen auftreten, die um die Mitte der aktuellen Marke zentriert sind. Um die exakte Position der Markenmitte zu identifizieren, werden einige weitere Verarbeitungsschritte ausgeführt. Das Bilddarstellungssystem wird zuerst zwischen zwei Werten $T3$ und $T4$ von der Abschneidevorrichtung **23** abgeschnitten. Das heißt, falls das Eingangssignal an die Abschneidevorrichtung zwischen $T3$ und $T4$ liegt, ist das Ausgangssignal gleich dem Eingangssignal. Falls das Eingangssignal kleiner als $T3$ ist, ist das Ausgangssignal $T3$, und falls das Eingangssignal größer ist als $T4$, ist das Ausgangssignal $T4$. Die Werte $T3$ und $T4$ werden gewählt, um zu gewährleisten, dass die schwarze Umgebung der Marken unterhalb von $T3$ liegt und dass der weiße Wert der Ringe oberhalb von $T4$ liegt. Dies verhindert jegliche geringfügigen Änderungen in dem Beleuchtungswert über die Marke, was die Messung der Mittenposition stören würde. Das abgeschnittene Signal wird dann verarbeitet, um die Position der Mitte der Marke abzuschätzen. Die Verarbeitung wird separat für die horizontale und vertikale Position ausgeführt. Um die horizontale Position zu bestimmen, werden die Werte des abgeschnittenen Bildsignals unabhängig in der linken und rechten Hälfte des Rechtecks A aufsummiert. Dies kann mit Hilfe einer Anordnung ausgeführt werden, die ähnlich zu derjenigen ist, die in der [Fig. 5](#) gezeigt ist, ohne den anfänglichen Komparator **40**. Die Summe von Werten in der rechten Hälfte wird von der auf der linken Hälfte subtrahiert. Dies wird mit Hilfe des Prozessschrittes **26** in der [Fig. 3](#) ausgeführt. Dieser Wert wird positiv sein, wenn die Marke überwiegend auf der rechten Hälfte auftritt, und wird andererseits negativ sein, wenn diese überwiegend in der linken Hälfte liegt. Das Signal wird auf sein Vorzeichen hin untersucht, was sich von positiv auf negativ ändert, und zwar für aufeinander folgende Pixel und mit Hilfe des Vorzeichenänderungsdetektors **29**, der anzeigt, dass sich die Position des Rechtecks A von einer Position unmittelbar auf der linken Seite zu einer Position unmittelbar auf der rechten Seite der Mitte der Marke bewegt hat. Dieser Detektor erzeugt ein "wahres" Ausgangssignal, wenn dies auftritt. Die horizontale Position der Markenmitte wird mit Hilfe des Interpolators **31** auf eine Genauigkeit von unter einem Pixel geschätzt, der die Position schätzt, bei der das Ausgangssignal von **26** durch 0 geht. Dies kann beispielsweise dadurch erzielt werden, dass eine lineare Interpolation zwischen den zwei Werten zu beiden Seiten von 0 ausgeführt wird. Eine vergleichbare Verarbeitung (**27**, **30**) wird dazu verwendet, um die vertikale Position der Markenmitte zu lokalisieren.

[0044] Dieses Verfahren zum Auffinden der Mitte einer Marke hat den Vorteil, dass es recheneffizient ist, weil die Mehrzahl der Verarbeitungsschritte nur eine geringe Anzahl von Addierern, Subtrahierern und Verzögerungselementen erfordert. Aufwändigere Verfahren zum Auffinden der Mitte des Bildes eines Kreises können jedoch auch verwendet werden. Viele mögliche Verfahren sind aus der Literatur bekannt, die beispielsweise auf einer Berechnung des Schwerpunkts oder auf einem Nachzeichnen der Randkontur beruhen. Solche Verfahren können auf eine Ausführungsform angewendet werden und können die Möglichkeit bieten, eine größere Genauigkeit zulasten von höheren Rechenanforderungen zu erzielen.

[0045] Das Vorhandensein einer Marke wird angezeigt, wenn die Ausgangssignale von **24**, **25**, **29** und **30** sämtlich "wahr" sind. Wenn dies geschieht, kann die Position der Marke, die mit einer Genauigkeit von unter einem Pixel interpoliert wurde, von dem Interpolator **31** gelesen werden.

[0046] Wenn einmal die Mitte einer Marke aufgefunden worden ist, kann das Bild um diese Position herum verarbeitet werden, um die Markennummer, die durch den Strichcode repräsentiert wird, zu lesen. Der Algorithmus, der dazu verwendet wird, um dies auszuführen, wird von dem verwendeten Streifencode-typ abhängen. Beispielsweise ist bei dem Typ von Streifencode, der in der [Fig. 2](#) gezeigt ist, ein Weg, um voranzukommen, wie folgt. Eine Anzahl von aufeinander folgenden Videoproben wird ausgewählt, beginnend bei dem Pixel am nächsten zu der Markenmitte und bis hin zu dem Rand des Rechtecks A. Diese Pixel werden Proben entlang eines Radius der Marke repräsentieren. Der Außenrand der Marke wird durch Untersuchen der Proben lokalisiert, und zwar beginnend von der äußersten und davon fortschreitend nach innen, wobei nach einer Probe oberhalb eines vorgegebenen Schwellenwerts gesucht wird. Der Schwellenwert würde normalerweise gewählt werden, so dass dieser sich in der Mitte zwischen den Videowerten von weiß und schwarz befindet. Wenn einmal der Rand lokalisiert wurde, können die Positionen entlang des Radius, die den Mitten von jedem Bit des Streifencodes entsprechen, untersucht werden und mit dem Schwellenwert verglichen werden. Dieser Prozess kann Verwendung machen von einer Interpolation, um die Videowerte bei Positionen zwischen existierenden Proben zu schätzen. Für jede Bitposition wird das Bit in dem Code auf 1 gesetzt, falls der Videowert oberhalb des Schwellenwertes liegt, und ansonsten auf 0. Dieser Prozess kann für einige Sätze von Proben, beginnend bei der Mitte, wiederholt werden, beispielsweise in Richtung nach links, nach rechts oberhalb und unterhalb der Mitte. Dies ermöglicht, dass eine Anzahl von unabhängigen Auslesevorgängen des Streifencodes ausgeführt werden können. Die Ausleseergebnisse kön-

nen verglichen werden und die Marke kann entweder akzeptiert oder zurückgewiesen werden. Beispielsweise könnte eine Marke als gültig akzeptiert werden, falls drei oder vier Auslesewerte des Streifencodes übereinstimmen, oder alternativ kann eine einstimmige Vorhersage erforderlich sein. Weitere Modifikationen sind möglich, beispielsweise eine Änderung der Schwellenwerte, die bei dem Vorgang zum Auslesen des Streifencodes verwendet werden, um die bestmögliche Übereinstimmung zwischen Auslesewerten und den verschiedenen Radien zu erhalten.

[0047] Falls der Vorgang zum Auslesen des Streifencodes ein akzeptables Ergebnis erzeugt, wird eine Datenbank, die alle Marken in dem Studio auflistet, untersucht, um zu überprüfen, ob die Marke existiert, und falls dies so ist, um deren Koordinaten in dem Studio zu bestimmen. Falls die Marke nicht in der Tabelle vorliegt, braucht diese nicht berücksichtigt zu werden, was einem fehlerhaften Auslesevorgang entspricht. Alternativ könnten gewisse Fehlerkorrekturtechniken dazu verwendet werden, um zu versuchen, falsch ausgelesene Bits in dem Streifencode zu korrigieren. Es gibt verschiedene wohlbekannte Fehlerkorrekturverfahren, die verwendet werden können, was zulasten der Tatsache geht, dass eine Redundanz in den Streifencode eingebaut werden muss. Dies erfordert außerdem das Hinzufügen von mehr Bits zu dem Code, was größere Marken erfordert. Andere Information könnte dazu verwendet werden, um für eine bekannte Kameraposition, die auf dem vorherigen Bild gemessen wurde, dabei zu helfen, einen nicht erfolgreich gelesenen Streifencode aufzulösen, beispielsweise eine Abschätzung, welche Marke bei dieser Position in dem Bild erwartet werden könnte.

[0048] Aus der Liste der Markenpositionen in dem Bild und aus den entsprechenden Positionen der Marken in dem Studio ist es möglich, die Kameraposition und -orientierung zu berechnen. Dies kann beispielsweise dadurch ausgeführt werden, dass die photogrammetrische Standardtechnik einer "Bündelanpassung" ("bundles adjustment") verwendet wird, die die Kameraposition und -orientierung berechnet, die die Summe der Quadrate der Abstände zwischen den Positionen in dem Bild minimiert, bei denen (unter der Annahme der bekannten Position der Marken in dem Studio und der internen Kameraparameter) erwartet würde, dass die Marken auftauchen und bei denen diese tatsächlich gefunden werden. Geeignete Techniken zum Ausführen der Bündelanpassung sind im "Manual of Photogrammetry, 4. Auflage", American Society of Photogrammetry, Falls Church, VA, 1980, herausgegeben von Slama, C. C., beschrieben. Aufwändigere Verfahren können alternativ verwendet werden, die keine vollständige Kenntnis der internen Kameraparameter oder der Markenpositionen erfordern.

[0049] Eine Ausführung dieser Berechnung unter Verwendung der Markenpositionen, die in jedem Bild exakt gefunden werden, kann jedoch zu einer plötzlichen Änderung in der berechneten Kameraposition führen, wenn verschiedene Marken in das Blickfeld hinein gelangen oder aus diesem sich hinaus bewegen, was zuvor erläutert wurde. Es hat sich herausgestellt, dass es für die berechnete Kameraposition viel akzeptabler ist, wenn diese sich unter gewissen Umständen langsam ändert. Dies ist besonders hilfreich in einer Situation, wo eine Marke vorübergehend detektiert wird, weil der Effekt ihres Verschwindens von einem Bild und des erneuten Auftauchens erheblich verringert werden kann. Die Weise, auf die dies erzielt wird, ist in der [Fig. 6](#) gezeigt, die nachfolgend erklärt wird.

[0050] Die Positionen und Streifencodenummern der in dem Bild gefundenen Marken werden zuerst an einen Prozessschritt **40** weitergeleitet, der die Kameraposition auf der Grundlage dieser Information unter Verwendung von photogrammetrischen Standardtechniken berechnet. Die berechnete Position und Orientierung, **47**, wird abhängen von einer plötzlichen Änderung, falls eine Marke verschwindet oder wieder auftaucht, wie dies zuvor erklärt wurde. Die Position und Orientierung wird in einen Prozessschritt **41** eingegeben, der die Positionen in dem Bild berechnet, bei denen das Verschwinden der Marken erwartet würde, und zwar basierend auf der gemessenen Kameraposition, den internen Parametern der Kamera und den Positionen der Marken in dem Studio. Die Differenz zwischen den erwarteten und tatsächlichen Markenpositionen wird mit Hilfe eines Subtrahierers **42** berechnet. Die Subtraktionsoperationen sind für jede sichtbare Marke notwendig, und zwar eine für die horizontale Koordinate und eine für die vertikale Koordinate. Falls die Kamerakalibrierung und die absoluten Markenpositionen perfekt genau sind, wird das Ausgangssignal dieses Subtrahierers für jede Marke 0 sein; in einer praktischen Situation wird dieses jedoch eine kleine Differenz sein. Es sei angemerkt, dass der Prozessschritt **40** typischerweise darauf abzielt, die Summe der Quadrate dieser Differenzen zu minimieren. In gewissen Anwendungen von **40** werden die Differenzen während des Berechnungsprozesses erzeugt, so dass ein separater Prozessschritt nicht erforderlich ist. Diese Differenzsignale (umfassend horizontale und vertikale Verschiebungsdifferenzen für jede Marke) werden mit Hilfe eines seitlichen Tiefpassfilters **43** gefiltert. Dieses Filter kann beispielsweise ein rekursives Filter erster Ordnung sein. Das Filter kann separat auf die horizontalen und vertikalen Positionsdifferenzen für jede Marke angewendet werden, obwohl auch eine aufwändigere Verarbeitung möglich ist, die beide Komponenten gemeinsam behandelt. Die gefilterten Positionsdifferenzen werden zu den beobachteten Markenpositionen mit Hilfe eines Addierers **44** addiert, um einen Satz von "angepassten" Markenposi-

tionen **50** zu bilden. Die Position und Orientierung werden aus diesen angepassten Positionen mit Hilfe eines Prozessschrittes **45** berechnet, der im Wesentlichen auf dieselbe Weise wie der Prozessschritt **40** arbeitet. Es sei angemerkt, dass es in gewissen Anwendungen der Prozessschritte **40** und **45** nützlich ist, einen Schätzwert der Kameraposition zu haben, beispielsweise um eine rekursive Minimierungsprozedur zu initialisieren. Falls dies der Fall ist, kann es zweckmäßig sein, die von **40** berechnete Kameraposition zu verwenden, um den Prozess **45** zu initialisieren. Die durch **45** berechnete Position kann dazu verwendet werden, um den Prozess **40** auf dem nächsten Bild zu initialisieren.

[0051] In einem Bereitschaftszustand ohne Kamerabewegung und ohne Marken, die auftauchen oder verschwinden, wird das Ausgangssignal des Tiefpassfilters **43** dasselbe sein wie dessen Eingangssignal, so dass die angepassten Markenpositionen **50** gleich den von **41** berechneten erwarteten Positionen sein werden. Der Prozess **45** zur Berechnung der Kameraposition wird somit Werte für die Position und Orientierung berechnen, die identisch zu denjenigen sind, die von **40** berechnet werden, und deshalb denjenigen entsprechen, die direkt von den beobachteten Markenpositionen abgeleitet werden.

[0052] Falls eine Marke verschwindet, kann sich die aus den Markenrohdaten berechnete Kameraposition und -orientierung **47** geringfügig ändern. Dies wird zu einer Änderung der erwarteten Markenpositionen führen, die von **41** berechnet werden. Diese Änderung wird jedoch sofort bei dem Ausgang des Filters **43** auftreten, und zwar wegen der Natur des zeitlichen Tiefpasses. Die angepassten Markenpositionen **50** werden deshalb im Wesentlichen unverändert zu den angepassten Positionen sein, die der Situation entsprechen, bevor die Marke verschwand, obwohl eine Markenposition weniger vorhanden sein wird. (Es sei angemerkt, dass der Subtrahierer **42** und der Addierer **44** nur ein gültiges Ausgangssignal für Marken erzeugt, die bei den beiden Eingangssignalen vorhanden sind.) Die Kameraposition, die von dem Prozessschritt **45** berechnet wird, wird deshalb im Wesentlichen unverändert zu derjenigen sein, die berechnet worden wäre, wenn die Marke nicht verschwunden wäre. Dies ist deshalb der Fall, weil die Markenpositionen **50** selbstkonsistent sind; d. h. jegliche Kombination von Marken wird dieselbe berechnete Kameraposition und -orientierung erzeugen. Während die Zeit verstreicht (und unter der Annahme, dass nicht mehr Marken verschwinden oder auftauchen), wird das Ausgangssignal des Filters **43** allmählich zu seinem Eingangswert hin tendieren. Die angepassten Markenpositionen werden sich somit langsam in Richtung zu den Positionen, die von dem Prozess **41** berechnet wurden, bewegen. Deshalb wird die zum Schluss berechnete Kameraposition **48** langsam zu derjenigen bei **47** tendieren, die aus den

Markenrohdaten berechnet wurde.

[0053] Falls eine neue Marke in dem Bild auftaucht, muss das Filter **43** initialisiert werden, so dass diese Marke gefiltert werden kann. Das heißt, es müssen Werte für die vorherige Anpassungsgröße für die neue Marke angenommen werden, weil diese erforderlich sind, damit das zeitliche Filter arbeiten kann. Die einfachste Option wäre es, anzunehmen, dass eine vorherige Anpassungsgröße 0 ist, aber dies würde wahrscheinlich zu einem Sprung in der berechneten Kameraposition führen, weil die Anpassungsgröße wahrscheinlich exakt 0 ist. Eine bevorzugte Option besteht darin, einen Wert abzuleiten, der gewährleistet wird, dass es keinen plötzlichen Sprung in der berechneten Kameraposition gibt. Eine Weise, auf die dies ausgeführt wird, ist wie folgt. Wenn die Marke zuerst in dem Bild auftaucht, wird ihre Präsenz von den Prozessschritten zur Berechnung der Kameraposition und -orientierung und von dem Filter ignoriert, der nur auf die Marken einwirkt, die sowohl in dem augenblicklichen als auch in dem vorhergehenden Bild sichtbar sind. Die erwartete Position der neuen Marke wird dann auf der Grundlage der berechneten Kameraposition und -orientierung von dem Prozessschritt **46** berechnet. Die Differenz zwischen dieser erwarteten Position und der Position, bei der diese tatsächlich auftaucht, wird von dem Subtrahierer **51** berechnet und dann dazu verwendet, um das Filter **43** zu initialisieren. Dies initialisiert das Filter auf einen Zustand, in welchem dieses sich befunden hätte, wenn die neue Marke stets diese Differenz zwischen der erwarteten und beobachteten Position gehabt hätte. Auf dem nächsten Bild wird die neue Marke von sämtlichen der Prozessstufen verwendet. Ihr Auftauchen führt wahrscheinlich zu einer Änderung der Kameraposition und -orientierung **47**, die aus den Markenrohdaten berechnet wird, aber diese Änderung wird das Ausgangssignal des zeitlichen Filters **43** nicht unmittelbar berühren. Der Prozess **45** wird deshalb ein Ergebnis erzeugen, das sehr ähnlich zu demjenigen ist, das erzeugt worden wäre, wenn die neue Marke nicht aufgetaucht wäre. Allmählich werden jedoch die Korrekturverschiebungen bei dem Ausgang des Filters **43** zu denjenigen hin tendieren, die der von **40** berechneten Kameraposition entsprechen, und die bei **45** berechnete Kameraposition wird deshalb zu der bei **40** berechneten hin tendieren.

[0054] Falls weniger als drei Marken in irgendeinem Bild gesehen werden, wird das System nicht in der Lage sein, eine Position oder einen Winkel zu berechnen, weil die Anzahl von Unbekannten (drei Winkel und drei Positionskoordinaten) größer ist als die Anzahl von beobachteten Werten (zwei Bildkoordinaten für drei Marken). Das System kann auch bei der Berechnung eines gültigen Ergebnisses unter gewissen anderen Umständen versagen, beispielsweise dann, wenn eine Marke falsch identifiziert oder ihre

gemessene Position grob fehlerhaft ist. Wenn dies auftritt, ist es von Vorteil, die Ausgabe der zuletzt gültigen Messung solange fortzusetzen, bis ein gültiges Ergebnis erhalten wird. Außerdem kann es ein Vorteil sein, das Filter **43** auf den Zustand zu initialisieren, den dieses gehabt hätte, falls sämtliche Marken Positionsdifferenzen von 0 haben würden. Dies erlaubt es, dass sich das System rasch von einer Situation erholen kann, in der dieses gewisse ungenaue Messungen ausgeführt hat.

[0055] Eine Variante, bei der Marken in dem Blickfeld der Kamera positioniert sind, wird nun beschrieben werden; diese kann dazu verwendet werden, um die von dem vorgenannten Verfahren erhaltene Positionsinformation zu ergänzen, oder kann dazu verwendet werden, um die Positionsinformation unabhängig zu erhalten. Die Variante verwendet ein optisches System, das auf Clustern bzw. Mengen von Marken beruht, die in der Szene platziert sind. Jeder Cluster enthält vorzugsweise zumindest drei einzelne Markierungen (beispielsweise codierte konzentrische Ringe), vorzugsweise auf unterschiedlichen Höhen oberhalb einer Referenzoberfläche, und bevorzugter enthält dieser zumindest sechs einzelne Markierungen. Die Cluster werden vorzugsweise als starre Einheiten gebildet, auf denen die einzelnen Marken befestigt werden; dies ermöglicht eine Kenntnis der relativen Positionen von jeder Marke innerhalb von jedem Cluster mit einem hohen Grad an Genauigkeit.

[0056] Bei einer bevorzugten Realisierung, wie sie in der [Fig. 7](#) gezeigt ist, besteht jeder Cluster von Marken aus sechs Kreisen in einer Schattierung von blau gegen einen Hintergrund in einer geringfügig anderen Schattierung, so dass diese ausgetastet werden können. Die Verwendung von zwei Blaufarbtönen für eine Messung der Kameraposition ist bereits wirkungsvoll eingesetzt worden. Der Cluster ist so ausgelegt, dass die Kreise nicht alle in einer Ebene liegen; dies ermöglicht eine eindeutige Berechnung der Kameraposition aus einem einzelnen Cluster und erleichtert die automatische Messung der Clusterpositionen. Jede Marke innerhalb des Clusters hat einen Streifencode, um diese eindeutig zu identifizieren.

[0057] Das Bild wird analysiert, um die exakte Position von jeder Marke innerhalb jedes Clusters zu lokalisieren, woraus bei Kenntnis der Positionen der Marken in dem Studio die Kameraposition und -orientierung (drei Positionskoordinaten und drei Rotationswinkel) berechnet werden kann. Die internen Kameraparameter (einschließlich der Brennweite und Pixelgröße) müssen bekannt sein. Wenn die Kamera ein Zoomobjektiv aufweist, können diese mit Hilfe von Sensoren auf dem Objekt und/oder mit Hilfe von bekannten Verfahren zum Detektieren von Zoom- und Fokuseinstellungen eines Objektivs aus dem Bild bestimmt werden. Falls ein Festobjektiv verwen-

det wird, können die Einstellungen mit Hilfe einer separaten Anfangskalibrierungsprozedur bestimmt werden, weil diese für eine gegebene Wahl des Objekts konstant bleiben.

[0058] Die Cluster brauchen nicht alle dieselbe Anzahl von einzelnen Marken enthalten und einzelne Marken können auch bei bekannten Positionen innerhalb des Bildes angeordnet sein.

[0059] Die einzige Beschränkung, die dieses System der Positionierung der Markencluster auferlegt, besteht darin, dass zumindest einer und idealerweise ein Minimum von etwa drei gleichzeitig sichtbar sein müssen. So viele Cluster, wie erforderlich sind, um den Aktionsbereich abzudecken, können verwendet werden und so platziert werden, so dass diese nicht die Bewegung der Schauspieler stören. Dieser Ansatz ist inhärent auf sehr große Studios erweiterbar und lässt Szenen zu, die eine beträchtliche Schauspieler- und Kamerabewegung involvieren, wie beispielsweise verfolgende Schauspieler, die lange Korridore entlang rennen.

[0060] Weil die Cluster erneut positioniert werden können, um zu jeder aufgenommenen Szene zu passen, ist es wichtig, ein Verfahren vorzusehen, um deren Positionen mit hoher Genauigkeit automatisch zu messen. Dies kann dadurch bewerkstelligt werden, dass eine Anzahl von Bildern erfasst wird, die jeweils zwei oder mehr Cluster zeigen. Die Identitäten der Marken in jedem Bild können automatisch bestimmt werden. Eine globale Anpassung mit mittlerem quadratischem Fehler bestimmt die Relativposition und -orientierung von sämtlichen Marken über die verschiedenen Bilder.

[0061] Das System ist vorzugsweise ausgelegt, um in Echtzeit unter Verwendung einer normalen Workstation zu laufen, die mit einem Frame-Grabber bzw. einer Bildauslesekarte ausgestattet ist. Um dies zu erzielen, ist es wünschenswert, die Position von Marken von einem Bild zu dem nächsten zu verfolgen, so dass nur ein kleiner Teil des Bildes (um die vorhergesagten Markenpositionen herum) gesucht werden müssen. Außerdem erlaubt dies die Verwendung eines einfachen Algorithmus, um die Mitte von jedem Kreis aufzufinden. Bei Verwendung eines solchen Ansatzes hat sich herausgestellt, dass es möglich ist, die Kameraposition mit einer Frequenz von 25 Hz zu messen (wenn 20 Kreise verfolgt werden), wenn ein PC verwendet wird, der auf einem 200 MHz Pentium Pro basiert. Wenn das System jedoch gestartet wird, muss das gesamte Bild durchsucht werden, was einige Sekunden dauern kann.

[0062] Obwohl die beschriebenen Techniken ideal zur Bestimmung der Position einer Videokamera geeignet sind, können diese beispielsweise auf die Bestimmung der Position von Roboterarmen oder für

Virtual Reality angewendet werden, wo die Position eines Kopfhörers bzw. Headsets bestimmt werden soll, der bzw. das von einem Benutzer getragen wird.

Patentansprüche

1. Verfahren zur Bestimmung der Position eines Objekts, relativ zu einer Vielzahl von Referenzpositionen, basierend auf der Identifizierung einer entsprechenden Vielzahl von Markierungen an jenen Referenzpositionen, wobei das Verfahren die Anwendung von Korrekturfaktoren auf gespeicherte oder gemessene Positionen der Markierungen relativ zu dem Objekt umfasst, um entsprechende, korrigierte Markierungspositionen zu erzeugen, und um die Objektposition relativ zu den Referenzpositionen, basierend auf den korrigierten Referenzpositionen, zu ermitteln, wobei die Korrekturfaktoren auf der Basis der Differenz zwischen gemessenen Markierungspositionen und erwarteten Markierungspositionen berechnet werden, so dass die korrigierten Markierungspositionen korrigiert werden, um gegenseitig selbstkonsistente Werte zu werden, wobei die Korrekturfaktoren zeitlich Tiefpass gefiltert werden, um die Rate der Änderungen von Korrekturfaktoren zu reduzieren.

2. Verfahren nach Anspruch 1, wobei das zeitliche Tiefpass-Filtern die Rate der Änderung von Korrekturfaktoren auf eine vorherbestimmte Höhe reduziert, die so gewählt wurde, dass Übergänge bzw. Wechsel in der bestimmten Objektposition über eine Zeitspanne von mindestens 2 Sekunden geglättet werden, die aus einer oder mehreren identifizierten, unbrauchbar gewordenen Markierungen oder brauchbar gewordenen neuen Markierungen resultieren.

3. Verfahren nach irgendeinem der vorhergehenden Ansprüche, bei welchem ein Maß für die Genauigkeit einer Positionsbestimmung erzielt wird.

4. Verfahren nach Anspruch 3, wobei besagtes Maß auf der Bestimmung besagter Korrekturfaktoren basiert.

5. Verfahren nach irgendeinem der vorhergehenden Ansprüche, bei welchem die Markierungen in unterschiedlichen Entfernungen von einer Referenz-Oberfläche angebracht sind, wobei die gespeicherten Informationen ein Maß für die Entfernung der Markierung von der Referenz-Oberfläche enthalten.

6. Verfahren nach irgendeinem der vorhergehenden Ansprüche, das ferner ein Bild von mindestens einer Teilmenge der besagten Vielzahl von Markierungen einer Kamera aufweist, die mit dem Objekt in Verbindung gebracht wird.

7. Verfahren nach Anspruch 6, bei welchem das besagte Objekt eine Videokamera ist, die ein Bildfeld

bzw. Blickfeld aufweist, welches vom Bildfeld bzw. Gesichtsfeld der besagten Kamera, die mit dem Objekt in Verbindung gebracht wird, abweicht.

8. Verfahren nach Anspruch 7, weiterhin umfassend eine Bereitstellung von zumindest einer zusätzlichen Markierung an einer vorgegebenen Position, so dass diese für die Videokamera sichtbar ist, wobei die oder jede Markierung Identifizierungsinformationen aufweist, die darin in binärer Form als Folge von konzentrischen Ringen von zwei Farbtönen kodiert sind, wobei die Farbtöne so gewählt werden, dass die Markierung durch Einblendung bzw. Chroma-Keying aus dem Videokamerabild ausgetastet werden kann, wobei die Bestimmung der Position der Videokamera auf der Position der zusätzlichen Markierung, falls diese vorhanden ist, in dem Videokamerabild basiert.

Es folgen 6 Blatt Zeichnungen

Anhängende Zeichnungen

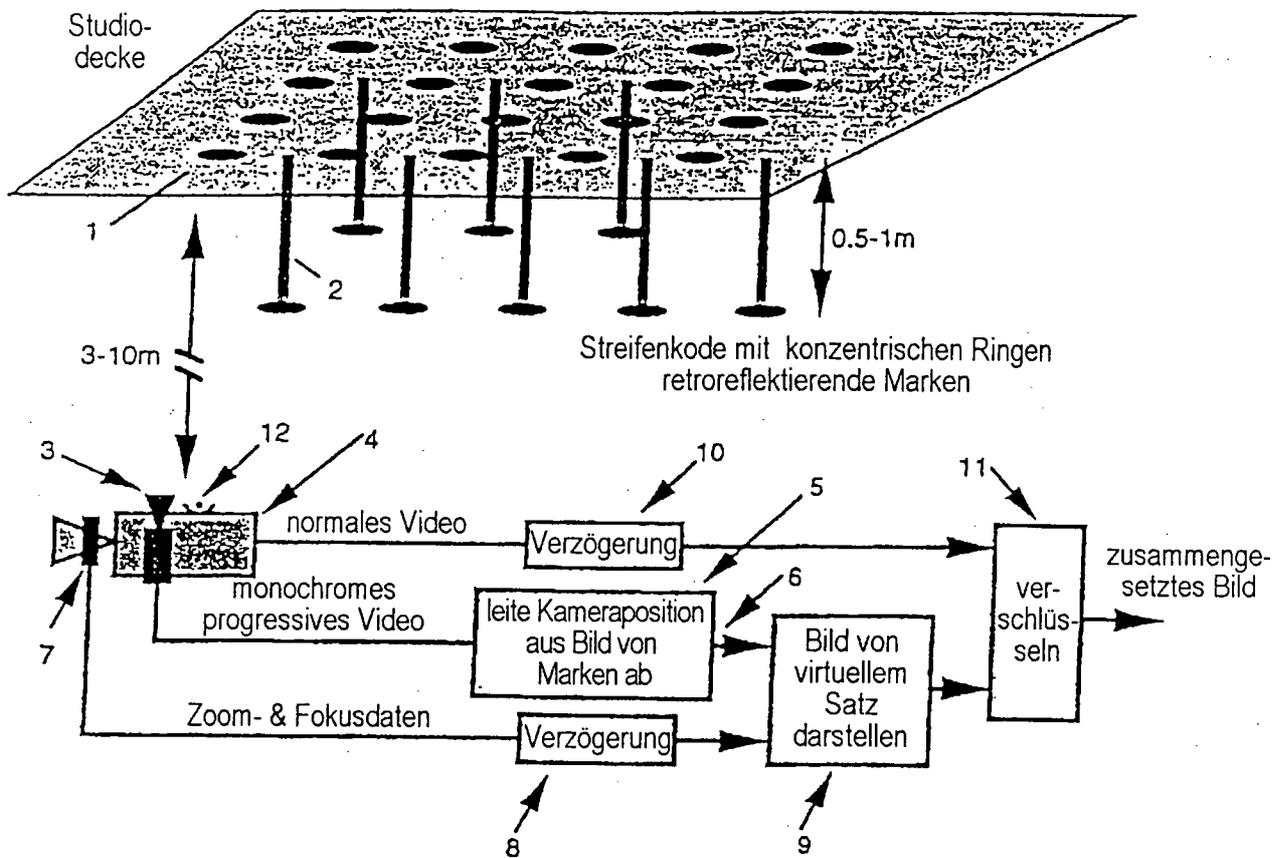


Fig. 1

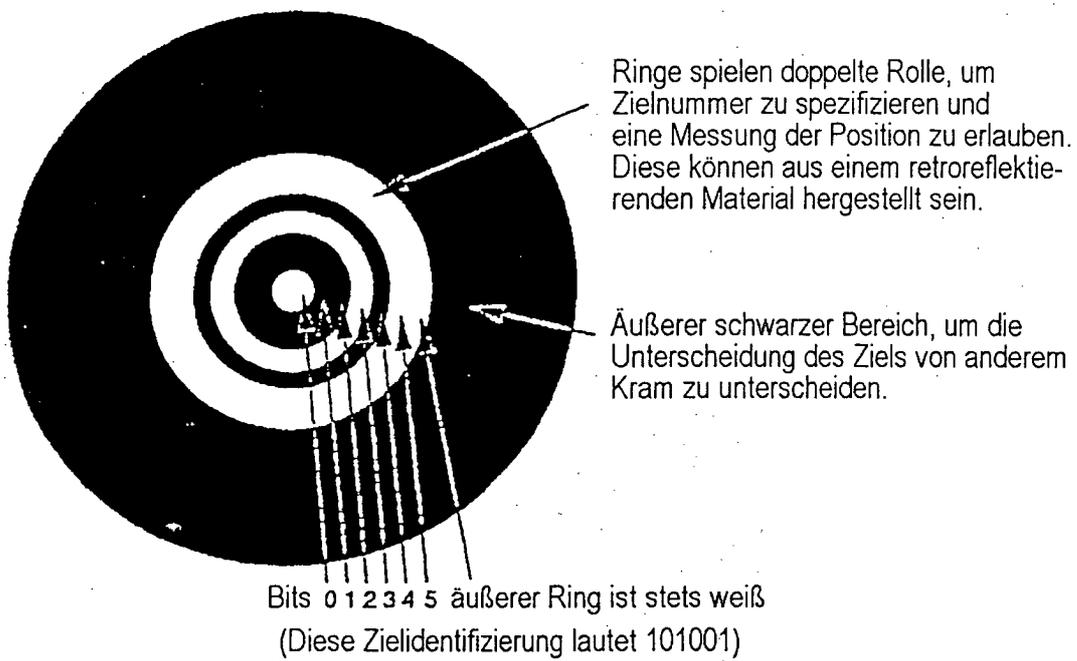


Fig. 2

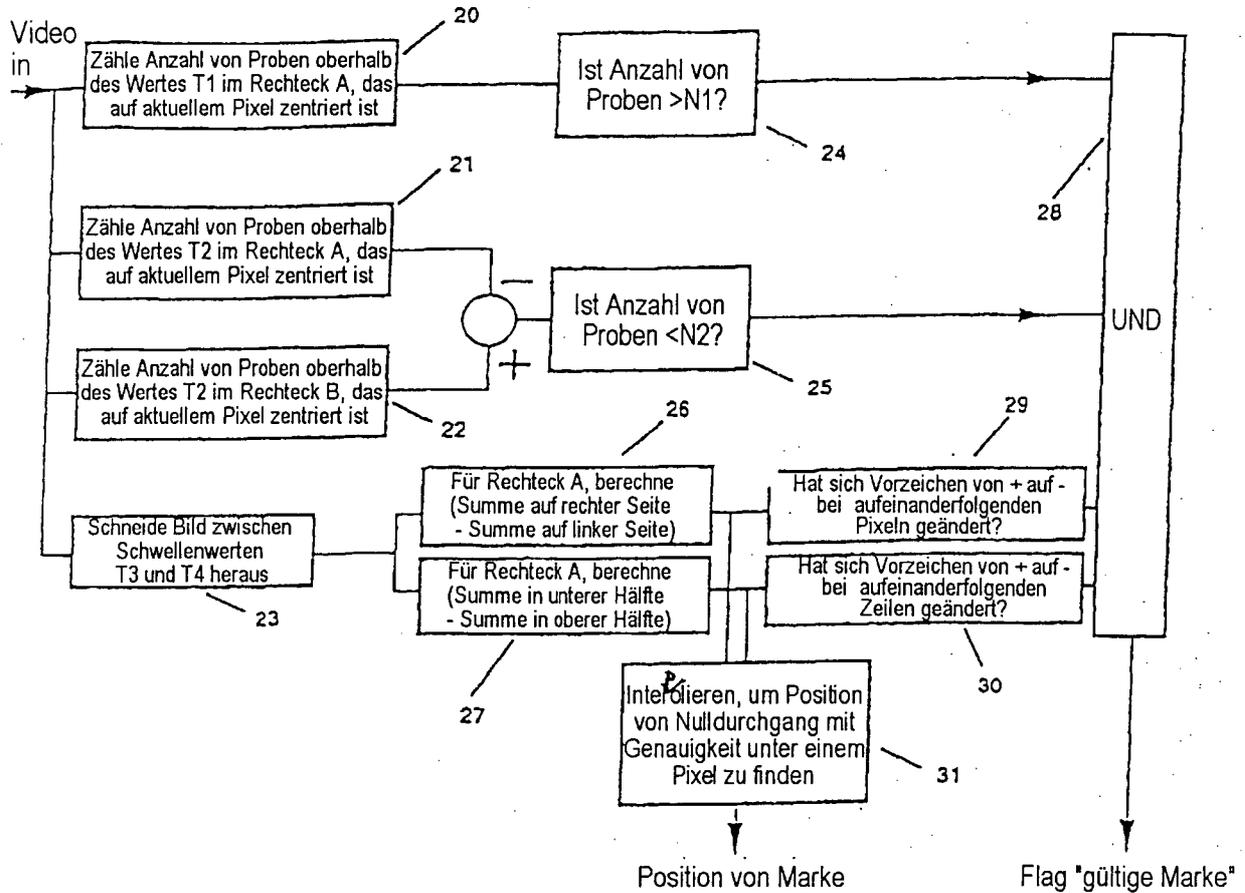


Fig. 3

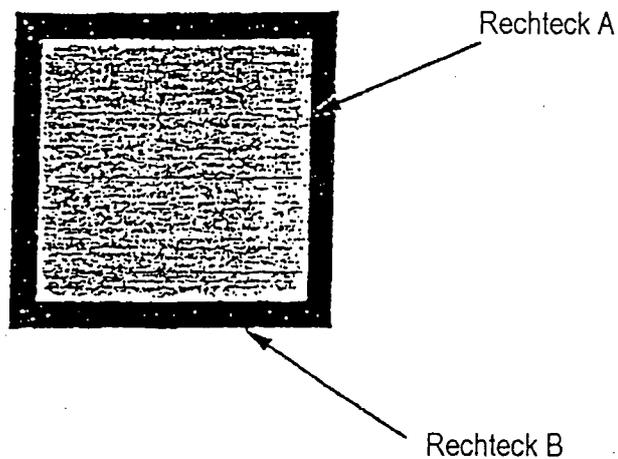


Fig. 4

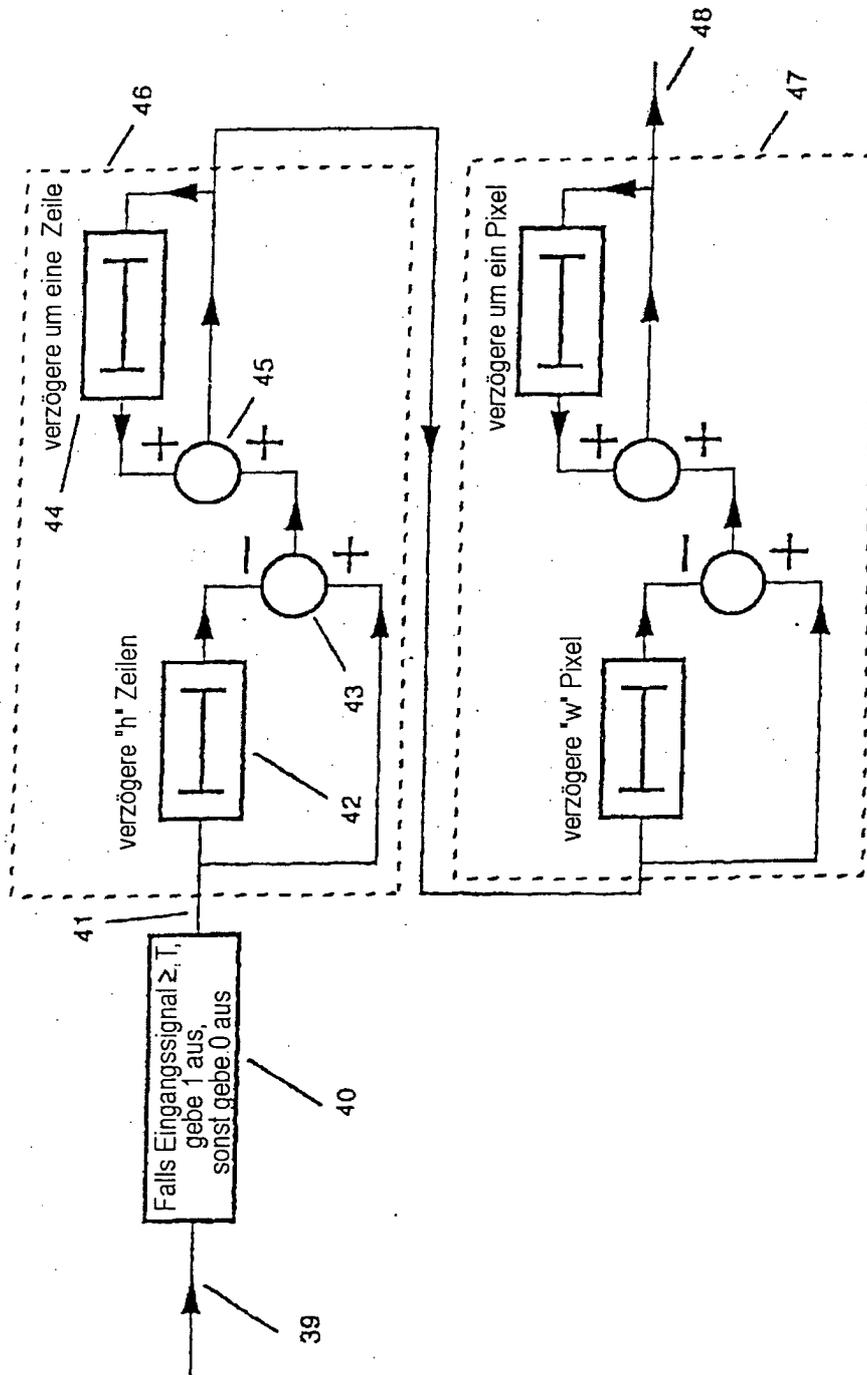


Fig. 5

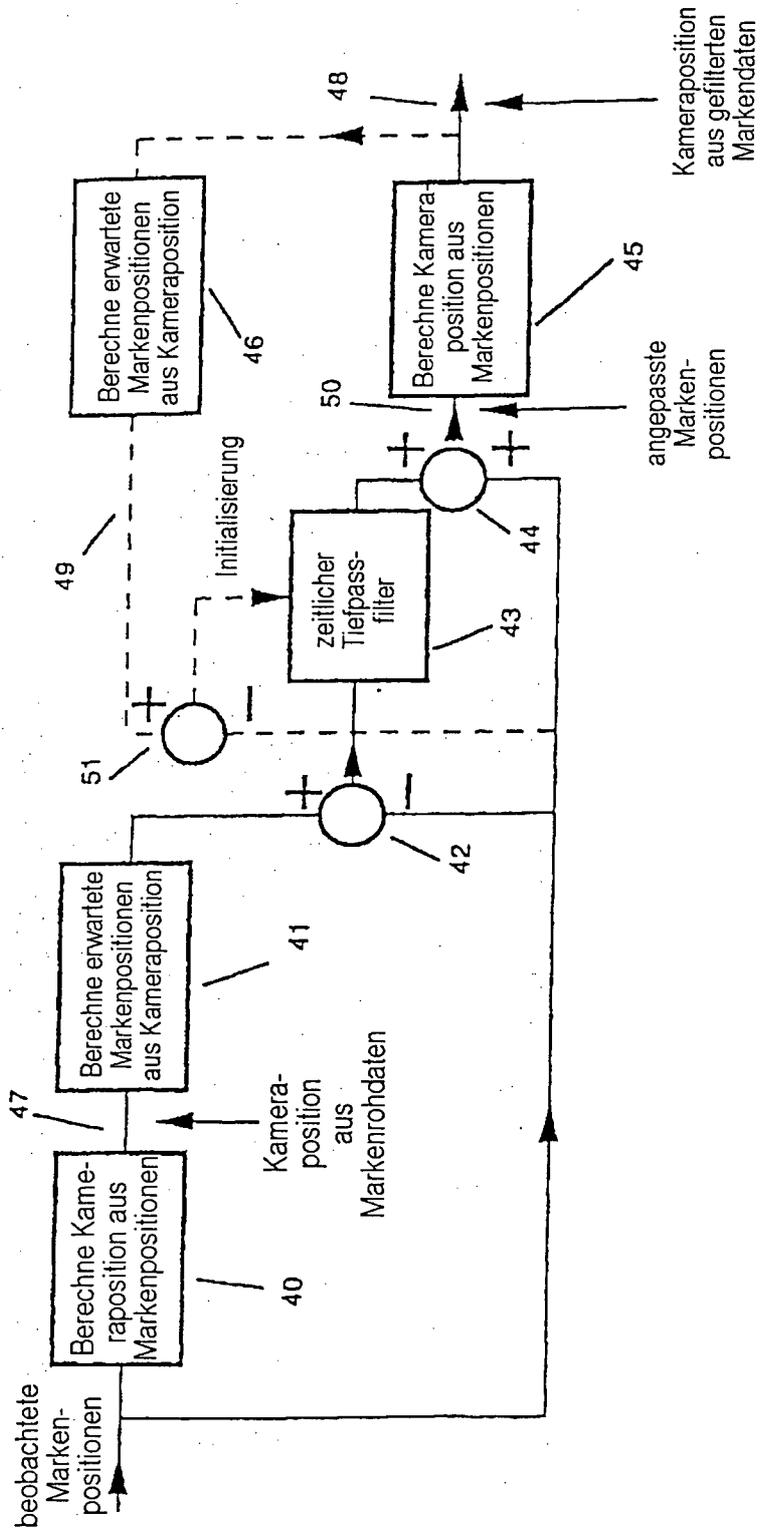


Fig. 6

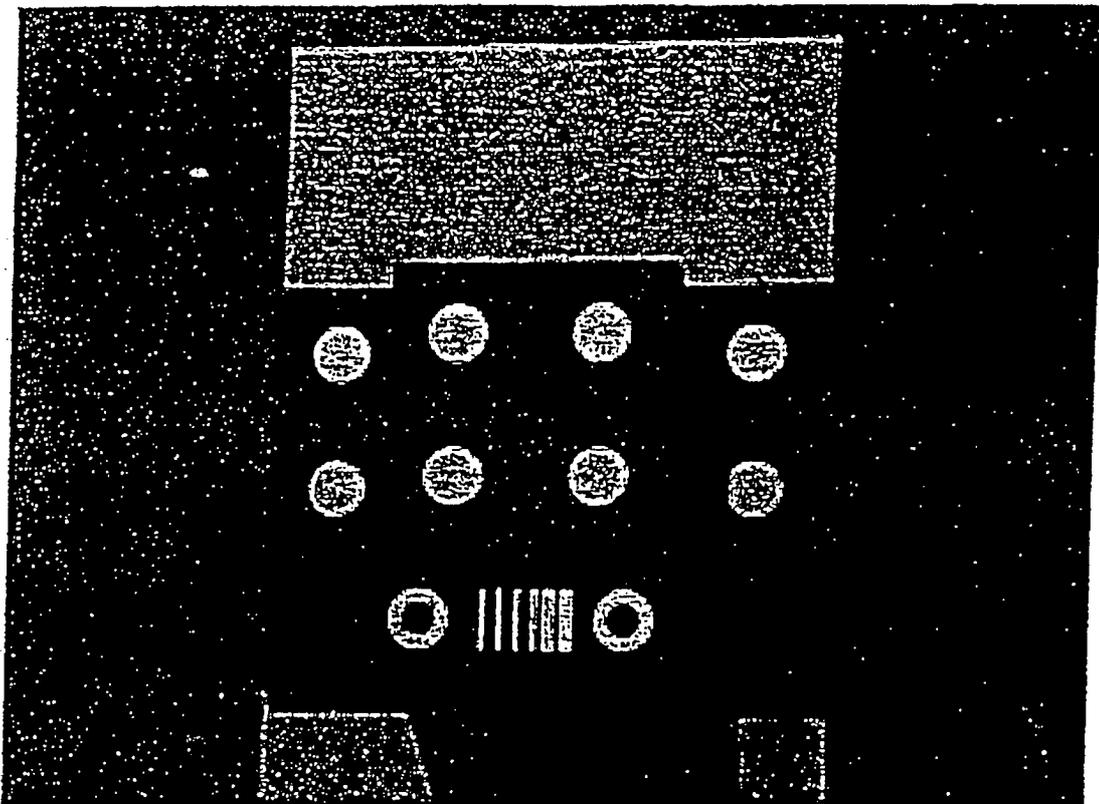


Fig. 7