



(19)
Bundesrepublik Deutschland
Deutsches Patent- und Markenamt

(10) **DE 603 17 053 T2** 2008.08.07

(12)

Übersetzung der europäischen Patentschrift

(97) **EP 1 494 132 B1**

(21) Deutsches Aktenzeichen: **603 17 053.6**

(96) Europäisches Aktenzeichen: **03 254 263.1**

(96) Europäischer Anmeldetag: **04.07.2003**

(97) Erstveröffentlichung durch das EPA: **05.01.2005**

(97) Veröffentlichungstag

der Patenterteilung beim EPA: **24.10.2007**

(47) Veröffentlichungstag im Patentblatt: **07.08.2008**

(51) Int Cl.⁸: **G06F 17/30** (2006.01)
G06T 7/40 (2006.01)

(73) Patentinhaber:

Mitsubishi Denki K.K., Tokyo, JP

(74) Vertreter:

**Rummler, F., Dipl.-Ing.Univ., Pat.-Anw., 80802
München**

(84) Benannte Vertragsstaaten:

DE, FR, GB, IT

(72) Erfinder:

Cieplinski, Leszek, Guildford, Surrey GU1 2UF, GB

(54) Bezeichnung: **Verfahren und Vorrichtung zur Darstellung einer Bildgruppe**

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 Erfindung betrifft ein Verfahren und eine Vorrichtung zum Darstellen einer Gruppe von Bildern, insbesondere hinsichtlich der Farbe und zum Suchen nach und Abrufen von Bildern.

[0002] Es gibt verschiedene bekannte Techniken zum Darstellen eines Bilds unter Verwendung visueller Merkmale wie der in dem Bild erscheinenden Farbe. In einer bekannten Technik wird beispielsweise jedem Pixel ein Farbwert zugewiesen und ein Farbhistogramm wird abgeleitet, indem Bins für eine Anzahl von Farbwertbereichen eingestellt werden und die Anzahl von Pixeln in einem Bild gezählt wird, welche einen Farbwert in jedem der Bereiche aufweisen. Das Farbhistogramm wird dann als eine Darstellung des Bilds verwendet. In einer weiteren bekannten Technik wird eine oder mehrere dominante Farbe in einem Bild identifiziert und die dominanten Farben werden zur Darstellung des Bildes verwendet.

[0003] Die vorliegende Erfindung betrifft insbesondere Bildgruppen. Die Bildgruppen können beispielsweise eine Bildsequenz (Frames oder Felder) aus einem Video sein oder jede Bildgruppe aus jeder Quelle, in der die Bilder auf eine Weise zusammengeordnet sind. Die Bildgruppe könnte beispielsweise eine einzelne Szene oder Aufnahme in einem Video betreffen. Ein in dem Gebiet der Technik bekannter Ausdruck für eine solche Bildgruppe ist Framegruppe/Bildgruppe. Im Folgenden wird dieser Ausdruck als GoFGoP bezeichnet. In dieser Beschreibung wird der Begriff Bild verwendet, um einen Frame/ein Bild in einer Gruppe zu beschreiben, unabhängig davon, ob es sich um einen Videoframe oder -feld oder ein stilles Bild handelt. Die Begriffe Bild und Bildbereich werden außerdem austauschbar verwendet, es sei denn, es geht aus dem Kontext anders hervor.

[0004] Ein Ansatz zur Darstellung einer Bildgruppe ist die Auswahl eines einzelnen Bilds aus der Bildgruppe und das Behandeln des einzelnen Bilds als Darstellung der gesamten Gruppe. Das einzelne Bild wird dann unter Verwendung einer bekannten Technik zur Darstellung eines einzelnen Bilds dargestellt. Das einzelne Bild kann beispielsweise das erste oder letzte Bild, das in einer Sequenz erscheint, sein oder die Bildgruppe kann analysiert werden, um ein Bild in der Gruppe zu identifizieren, das hinsichtlich des interessierenden visuellen Merkmals die Gruppe besonders darstellt.

[0005] Ein weiterer Ansatz ist die Aggregation der Bildgruppe. Der bestehende visuelle Standard MPEG-7 (ISO/IEC 15938-3) ermöglicht die Beschreibung von Farbe in einem Videosegment oder einer Bildgruppe unter Verwendung eines GoFGoP-Farbbeschreibers. Dies wird beispielsweise ausführlich in dem Buch beschrieben: Introduction to MPEG-7 Mul-

timedia Content Description Interface, herausgegeben von Manjunath, Salembier und Sikora, ISBN 0-471-48678-7, Abschnitt 13.5. Es werden drei Techniken zum Aggregieren einer Bildgruppe beschrieben: Durchschnitt, Median und Querschnitt.

[0006] In jeder Technik wird ein Farbhistogramm für jedes Bild in der Gruppe abgeleitet wie oben beschrieben. In der Durchschnittstechnik werden die Farbhistogramme akkumuliert und dann wird jeder akkumulierte Bin-Wert durch N geteilt, wobei N die Anzahl von Bildern in der Gruppe ist, um ein Durchschnittshistogramm zu erzeugen. In der Mediantchnik werden für jeden Binwert die Histogrammwerte der Gruppe in ansteigender/absteigender Reihenfolge angeordnet und der Medianwert wird dem jeweiligen Bin zugewiesen. Das Querschnittshistogramm wird erhalten, indem für jeden Bin der Mindesthistogrammwert aus den Histogrammen für die Gruppe genommen wird. Anders ausgedrückt, stellt das Querschnittshistogramm die Anzahl von Pixeln einer bestimmten Farbe oder Farbbereichs dar (entsprechend einem Bin), welche in allen Bildern erscheinen.

[0007] In jedem Fall wird das aggregierte Histogramm (Durchschnitt, Median oder Querschnitt) dann unter Verwendung eines skalierbaren Farbbeschreibers (Abschnitt 13.4 des oben genannten Buches) dargestellt, der das Anlegen eines Codierschemas auf Basis einer Haar-Transformation auf die Werte in dem Farbhistogramm enthält.

[0008] Das Dokument „Automatic Video Scene Extraction by Shot Grouping“ von Tong Lin und Hong-Jiang Zhang betrifft das Gruppieren von Aufnahmen in Szenen. Eine Framegruppe, die eine Aufnahme bildet, wird analysiert, um die dominanten Farbobjekte in jedem Frame zu bestimmen und dann die dominanten Farbobjekte, die durch die gesamte Framegruppe anhalten, zu bestimmen. Dies erzeugt ein dominantes Farbhistogramm für eine Aufnahme. Dominante Farbhistogramme für verschiedene Aufnahmen werden verglichen. Wenn die Korrelation zwischen zwei Aufnahmen hoch ist, werden die Aufnahmen in eine Szene gruppiert.

[0009] T. Lin, C.-W. Ngo, H.-J. Zhang, Q.-Y. Shi, Integrating color and spatial features for content-based video retrieval, Proceedings 2001 International Conference On Image Processing, Icip 2001, Thessaloniki, Griechenland, 7.-10. Oktober 2001, International Conference on Image Processing, New York, NY:IEEE, US (07.10.2001), 1 von 3, CONF, 8, 592-595, bezieht sich auf einen Beschreiber für eine Framegruppe unter Verwendung dominanter Farbhistogramme, die temporäre Informationen aufnehmen, indem sie Histogrammbins entsprechend jeder dominanten Farbe bezüglich der relativen Dauer gewichten.

[0010] EP-A-1302865 (Mitsubishi EIT), 16. April 2003, (16.04.2003) betrifft ein Verfahren zur Bildsuche unter Verwendung von Beschreibern mit zwei oder mehr dominanten Farben, so dass die Beschreiber Subbeschreiber haben und jeder Subbeschreiber sich auf mindestens eine dominante Farbe bezieht, und mittels Abgleich unter Betrachtung einer Untergruppe dominanter Farben unter Verwendung von Subbeschreibern.

[0011] Aspekte der Erfindung werden in den beigefügten Ansprüchen dargelegt.

[0012] Eine Ausführungsform der Erfindung wird unter Bezugnahme auf die beigefügten Zeichnungen beschrieben, in denen:

[0013] [Fig. 1](#) ein Blockdiagramm eines Systems gemäß einer Ausführungsform der Erfindung ist;

[0014] [Fig. 2a](#), [Fig. 2b](#) und [Fig. 3–Fig. 5](#) Graphen sind, die ein Verfahren zum Zusammenführen von Bildbeschreibern gemäß einer Ausführungsform der Erfindung darstellen;

[0015] [Fig. 6](#) das Kombinieren von Bildern zur Bildung eines Superbilds darstellt.

[0016] In [Fig. 1](#) wird ein System gemäß einer Ausführungsform der Erfindung gezeigt. Das System enthält eine Steuereinheit **12**, wie einen Computer zum Steuern des Betriebs des Systems, wobei die Steuereinheit **12** mindestens einen Speicher und einen Prozessor, eine Anzeigeeinheit **14**, wie einen Monitor, der mit der Steuereinheit **12** verbunden ist, zum Anzeigen von Ausgaben einschließlich Bilder und Text und eine Zeigevorrichtung **16**, wie eine Maus zum Eingeben von Anweisungen an die Steuereinheit **12**, enthält. Das System enthält auch eine Bilddatenbank **18**, die digitale Versionen einer Vielzahl von Bildgruppen speichert und eine Beschreiberdatenbank **20**, die Beschreiberinformationen speichert, welche im Folgenden ausführlicher beschrieben werden, für jede der Bildgruppen, die in der Bilddatenbank **18** gespeichert sind. In diesem Beispiel entspricht jede Bildgruppe einer Aufnahme aus einer Videosequenz. Es gibt verschiedene bekannte Techniken zum Aufteilen einer Videosequenz in Aufnahmen, welche hier nicht ausführlich beschrieben werden. Die Erfindung kann auf jede Bildgruppe oder Bildbereiche, einschließlich Bereichen in Bildern, angewendet werden. Die Bilder können in Gruppen gespeichert werden oder es kann beispielsweise einen Identifizierer geben, der anzeigt, welche Bilder zu derselben Gruppe gehören.

[0017] Jede der Bilddatenbank **18** und der Beschreiberdatenbank **20** ist mit der Steuereinheit **12** verbunden. Das System enthält auch eine Suchmaschine **22**, bei der es sich um ein Computerprogramm unter

der Steuerung der Steuereinheit **12** handelt und welches an der Beschreiberdatenbank **20** betrieben wird.

[0018] In dieser Ausführungsform werden die Elemente des Systems an einer einzigen Stelle bereitgestellt, wie beispielsweise einer Bildbibliothek, wobei die Komponenten des Systems ständig verknüpft sind.

[0019] Die Beschreiberdatenbank **20** speichert Beschreiber aller in der Bilddatenbank gespeicherten Bilder und zusätzlich aller Bildgruppen in der Bilddatenbank. Die Bild- und Gruppenbeschreiber werden wie unten beschrieben abgeleitet.

[0020] Jedem Bild ist ein Bildbeschreiber zugeordnet, welcher das jeweilige Bild hinsichtlich der dominanten Farbe oder Farben in dem Bild ausdrückt. In dieser Ausführungsform werden die Bildbeschreiber im Wesentlichen wie in unserer gleichzeitig anhängigen Anmeldung WO 00/67203 beschrieben abgeleitet.

[0021] Es folgt eine kurze Zusammenfassung des Ableitens eines Bildbeschreibers. Jedes Bild weist eine Vielzahl von Pixeln auf und jedem Pixel ist ein Farbwert in dem relevanten Farbraum wie beispielsweise RGB zugewiesen.

[0022] Die Farbwerte und die entsprechenden Pixel bilden in dem Farbbereich Cluster, um die dominanten Farben und welche Farben einer jeweiligen dominanten Farbe entsprechen, zu bestimmen. Dies erfolgt unter Verwendung geeigneter Clusteralgorithmen, wie den verallgemeinerten Lloydalgorithmus, wie in Abschnitt 13.3.1 des oben genannten MPEG-7-Buchs beschrieben.

[0023] Die sich aus der Clusterprozedur ergebenden Clusterschwerpunkte werden als dominante Farbwerte verwendet, und die Pixelgruppen, die die jeweiligen Cluster bilden, werden zur Berechnung zusätzlicher Felder (Gewicht und Farbabweichung) wie unten erläutert verwendet.

[0024] Alternativ können die dominanten Farben unter Verwendung eines Histogrammansatzes abgeleitet werden, wie in WO 00/67203 beschrieben.

[0025] In diesem Fall wird ein Farbhistogramm für ein Bild abgeleitet, indem eine vorgegebene Anzahl von Farbwerten oder Farbwertbereichen in dem jeweiligen Farbraum ausgewählt wird, und die Anzahl von Pixeln in dem Bild, die jeden Farbwert oder Wert in dem relevanten Bereich aufweisen, wird gezählt.

[0026] Allgemein wird das Histogramm eine oder mehrere Spitzen aufweisen und die Spitzen (oder eine Untergruppe davon, wie beispielsweise eine

vorgegebene Anzahl der höchsten Spitzen) werden als die dominanten Farben ausgewählt und die Farbwerte/Pixel werden in dem Farbbereich hinsichtlich der dominanten Farben geclustert.

[0027] Nach Clusterung der Farbwerte wird ein Farbabweichungswert für jede dominante Farbe bestimmt, welcher für jede dominante Farbe die Abweichung der Farbwerte für das jeweilige auf der dominanten Farbe zentrierte Cluster ausdrückt. Die dominante Farbe kann als ein Mittelwert für die Farbverteilung in dem relevanten Cluster angesehen werden.

[0028] Die Berechnung der Abweichung kann unter Verwendung der folgenden Gleichung ausgedrückt werden:

$$CV_j = \frac{1}{N} \sum_{k=0}^{M_j-1} (m_{kj} - p_{kj})^2$$

wobei j die Farbkomponente indiziert, m_{kj} die j -te Komponente der dominanten Farbe ist, p_{kj} die j -te Komponente des k -ten Pixelwerts ist und die Summierung über N Pixel entsprechend der betrachteten dominanten Farbe erfolgt.

[0029] Der Beschreiber enthält auch einen Gewichtswert für jede dominante Farbe, bei dem es sich um ein Maß der relativen Signifikanz jeder dominanten Farbe in dem Bild handelt. In diesem Beispiel ist das Gewicht das Verhältnis der Anzahl von Pixeln in dem Cluster, die dem dominanten Farbwert entsprechen, zu der Gesamtanzahl von Pixeln in dem Bild. Das Gewicht kann als Prozentsatz ausgedrückt werden.

[0030] Die dominanten Farbwerte und ihre jeweiligen Abweichungen und Gewichte werden kombiniert, um einen Farbbeschreiber des Bilds zu bilden. Der Beschreiber kann auch andere Komponenten aufweisen, wie beispielsweise einen Grad n , der die Anzahl dominanter Farben anzeigt. Der Beschreiber kann auch Kovarianzwerte C_{ij} enthalten, wobei i und j Farbkomponenten in dem relevanten Farbraum für jede dominante Farbe und Cluster darstellen, sowie Abweichungswerte enthalten.

[0031] Der Farbbeschreiber für jedes Bild wird in der Beschreiberdatenbank gespeichert.

[0032] Es ist wichtig darauf hinzuweisen, dass ein dominanter Farbbeschreiber nicht das gleiche wie eine Histogrammdarstellung eines Bilds ist. Ein dominanter Farbbeschreiber enthält die Werte der dominanten Farbe in einem Bild, welche in vorherigen Verarbeitungsschritten bestimmt werden konnten. Eine Histogrammdarstellung eines Bilds enthält die dominanten Farben, aber es gibt keine Identifizierung der

dominanten Farben. Der dominante Farbbeschreiber kann auch andere Werte enthalten wie beispielsweise Werte, die der Abweichung der Farbverteilung hinsichtlich der oder jeder dominanten Farbe entsprechen, aber eine Histogrammdarstellung enthält keine Berechnung oder Bestimmung der Abweichung oder anderer Werte. Zu weiteren Komponenten eines dominanten Farbbeschreibers kann beispielsweise ein Gewicht gehören, welches die Einwirkung der dominanten Farbe in dem Bild, die Anzahl dominanter Farben in dem Bild und die räumliche Homogenität von Pixeln anzeigt, die den dominanten Farben in dem Bild entsprechen.

[0033] Ein Gruppenbeschreiber für eine Bildgruppe wird gemäß einer ersten Ausführungsform wie folgt abgeleitet.

[0034] Die Bildbeschreiber für jedes Bild in der Bildgruppe werden aus der Beschreiberdatenbank abgerufen. Die Bildbeschreiber werden dann zur Bildung eines Gruppenbeschreibers kombiniert.

[0035] Der Gruppenbeschreiber hat ein ähnliches Format wie ein Bildbeschreiber (dominante Farben, Abweichungen, Gewichte usw.). Vorzugsweise beträgt die Anzahl dominanter Farben in dem Gruppenbeschreiber zwischen 1 und 8. Die Anzahl dominanter Farben kann jedoch unbeschränkt sein oder auf ein vorgegebenes Maximum eingestellt sein. Die Anzahl dominanter Farben in dem Bildbeschreiber kann auch unbeschränkt sein oder auf ein vorgegebenes Maximum beschränkt sein. Das Maximum der Bildbeschreiber und das Maximum für den Gruppenbeschreiber muss nicht notwendigerweise dasselbe sein, so dass beispielsweise die Bildbeschreiber mehr dominante Farben aufweisen können als die Gruppenbeschreiber.

[0036] In diesem Beispiel gibt es zwei Bilder in einer Gruppe und die beiden jeweiligen Bildbeschreiber werden wie folgt kombiniert.

[0037] Allgemein ausgedrückt, werden die Bildbeschreiber kombiniert, indem Cluster in den Bildern basierend auf der Nähe der Cluster im Farbraum zusammengeführt werden.

[0038] [Fig. 2a](#) und [Fig. 2b](#) sind abstrakte Darstellungen von Clustern im Farbraum der beiden Bilder. [Fig. 2a](#) zeigt das erste Bild und [Fig. 2b](#) zeigt das zweite Bild. In jedem Fall stellen die Kreise einen Cluster für eine jeweilige dominante Farbe dar. Der Einfachheit halber wird der Farbraum in zwei Dimensionen gezeigt, obwohl der Farbraum normalerweise drei Dimensionen aufweisen wird. Die Cluster müssen außerdem nicht notwendigerweise Kreisen im Farbraum entsprechen, sondern werden so der Einfachheit halber gezeigt, und die Kreise geben keine Angabe über das Gewicht der Cluster. Im Folgenden

wird eine Darstellung, wie sie in [Fig. 2a](#) und [Fig. 2b](#) gezeigt wird (das heißt, eine Darstellung hinsichtlich der dominanten Farben und jeweiligen Cluster), als Clusterbeschreiber beschrieben.

[0039] Die beiden Clusterbeschreiber aus [Fig. 2a](#) und [Fig. 2b](#) werden kombiniert, um einen Clustersuperbeschreiber zu bilden. Dies wird in [Fig. 3](#) dargestellt.

[0040] Als nächstes wird der Abstand zwischen jedem Clusterpaar (wie oben definiert) im Superbeschreiber unter Verwendung einer geeigneten Abstandsmessung im Farbraum beschrieben. In diesem Beispiel ist der Abstand der euklidische Abstand im Farbraum.

[0041] In diesem Beispiel gibt es zwei dominante Farben und zwei Cluster im ersten Bild und drei dominante Farben und drei Cluster im zweiten Bild, die jeweils **1** bis **5** nummeriert sind. Jede dominante Farbe entspricht einem Punkt im RGB-Farbraum, welcher in [Fig. 2–Fig. 5](#) durch ein Kreuz dargestellt wird. Der Abstand zwischen einem Clusterpaar ist der Abstand im 3-D RGB-Raum zwischen den Schwerpunkten der Cluster. Der Abstand zwischen jedem Paar dominanter Farben in dem ersten und zweiten Bild wird berechnet einschließlich in diesem Beispiel der Cluster von demselben Bild. Das Paar dominanter Farben, die den geringsten Abstand geben, wird ausgewählt.

[0042] Als nächstes werden die den beiden dominanten Farben entsprechenden Cluster zusammengeführt.

[0043] In diesem Beispiel sind die beiden nächsten Cluster, wie in [Fig. 3](#) gezeigt, die Cluster **1** und **3** und diese werden wie folgt zusammengeführt.

[0044] Der dominante oder repräsentative Farbwert des zusammengeführten Clusters ist ein gewichteter Mittelwert der dominanten Farben der beiden Cluster, wobei das Gewicht wie oben definiert ist. Somit weist die zusammengeführte dominante Farbe m für die beiden dominanten Farben m_1 , m_2 und die jeweiligen Gewichte W_1 und W_2 den folgenden Wert auf:

$$m = w_1 m_1 + w_2 m_2$$

wobei w_1 , w_2 die relativen Gewichte, $W_1/W_1 + W_2$ und $W_2/W_1 + W_2$, sind.

[0045] Es wird auch die Abweichung des zusammengeführten Clusters unter Verwendung der Abweichungen der beiden zusammengeführten Cluster berechnet. In diesem Beispiel wird jede Farbkomponente unabhängig behandelt und es wird angenommen, dass die Abweichung des zusammengeführten Clusters eine gewichtete Summe zweier Gaußschen Ver-

teilungen ist. Dies führt zu der folgenden Formel für die Abweichung des zusammengeführten Clusters:

$$\sigma^2 = w_1 \sigma_1^2 + w_2 \sigma_2^2 + w_1 w_2 (m_1 - m_2)^2$$

wobei σ_1^2 , σ_2^2 die Abweichungen für die Komponentencluster, m_1 , m_2 ihre Mittelwerte und w_1 , w_2 ihre relativen Gewichte wie oben definiert sind.

[0046] Das Gewicht W für den zusammengeführten Cluster ist $W_1 + W_2$.

[0047] Das zusammengeführte Cluster wird als ein neues Cluster behandelt mit dem Gewicht, dem Mittelwert und der Abweichung wie oben beschrieben.

[0048] Dies wird in [Fig. 4](#) dargestellt, wo die Cluster **1** und **3** zusammengeführt werden, um ein neues Cluster **6** zu bilden.

[0049] Als nächstes gibt es eine weitere Wiederholung der Zusammenführungsschritte. Die beiden Cluster **1** und **3**, die in der ersten Wiederholung zusammengeführt worden sind, werden von der weiteren Betrachtung ausgeschlossen und durch das zusammengeführte Cluster **6** wie in [Fig. 4](#) gezeigt ersetzt.

[0050] Die Zusammenführungsschritte werden dann wiederholt, indem das nächste Clusterpaar einschließlich des zusammengeführten Clusters **6** im Farbraum identifiziert wird und sie wie oben dargelegt zusammengeführt werden.

[0051] In diesem Beispiel sind in der zweiten Wiederholung die Cluster **2** und **4** das nächste Clusterpaar. Diese werden zusammengeführt, um ein neues Cluster **7** zu erzeugen, mit der dominanten Farbe, dem Gewicht und der Abweichung, die wie oben dargelegt aus den dominanten Farben, den Gewichten und den Abweichungen von Clustern **2** und **4** abgeleitet werden. Das zusammengeführte Cluster **7** ersetzt die Cluster **2** und **4** wie in [Fig. 5](#) gezeigt.

[0052] Die Zusammenführungen werden wiederholt, bis ein vorgegebener Zustand eingehalten wird. Der vorgegebene Zustand kann beispielsweise sein, dass das Zusammenführen fortgeführt wird, bis eine vorgegebene Anzahl von Gesamtclustern übrig bleibt (Summe zusammengeführter Cluster und verbleibende Originalcluster in dem ersten und dem zweiten Bild). Alternativ kann der vorgegebene Zustand sein, dass das Zusammenführen fortgeführt wird, bis der Abstand zwischen jedem übrig bleibenden Clusterpaar größer als ein gegebener Wert ist. Alternativ kann das Verfahren eine vorgegebene Anzahl von Wiederholungen enthalten. Mehr als einer der vorgegebenen Zustände kann kombiniert werden.

[0053] Im vorliegenden Fall wird das Zusammen-

führen wiederholt, bis eine vorgegebene Anzahl (drei) von Clustern übrig bleibt.

[0054] Im obigen Beispiel werden die zusammengeführten Cluster bei weiteren Wiederholungen berücksichtigt. Sie können jedoch auch von weiteren Wiederholungen ausgeschlossen werden. Im obigen Beispiel können Cluster auch mit anderen Clustern in demselben Bild zusammengeführt werden, aber alternativ können die Cluster nur mit Clustern zusammengeführt werden, die in einem weiteren Bild erscheinen, in der ersten und/oder jeder nachfolgenden Wiederholung. Dies verringert die Anzahl von Abstandsmessungen.

[0055] Obwohl die Beschreibung hinsichtlich Clustern erfolgte, versteht sich, dass das Zusammenführen an den Werten der Beschreiber, das heißt, der dominanten Farbe, der Abweichung und dem Gewicht funktioniert und es nicht notwendig ist, die Cluster selbst zu analysieren.

[0056] Nach Abschluss der Wiederholungen werden die verbleibenden Cluster verwendet, um einen Gruppenbeschreiber zu bilden. Genauer ausgedrückt, gibt es für jeden Endcluster eine Darstellung oder eine dominante Farbe, eine jeweilige Abweichung und ein jeweiliges Gewicht. Diese werden zusammen mit einem Grad m , der die Anzahl von Endclustern angibt, kombiniert, um einen Gruppenbeschreiber zu bilden. Der Gruppenbeschreiber kann auch andere Faktoren wie eine Angabe des Farbraums oder die in der Darstellung verwendeten Farbquantisierung enthalten. Der Gruppenbeschreiber ist ein GoFGoP-dominanter Farbbeschreiber.

[0057] Im obigen Beispiel gibt es in der Bildgruppe nur zwei Bilder. Das Verfahren lässt sich jedoch auch auf Gruppen, die mehr als zwei Bilder enthalten, anwenden. Die Clusterbeschreiber für jedes Bild in der Gruppe könnten alle kombiniert werden, um wie oben einen Superbeschreiber zu bilden. Alternativ könnte die Bildgruppe in Untergruppen kombiniert werden, beispielsweise in Gruppen von zwei oder drei, die gegebenenfalls aufeinanderfolgen, und dann könnten die Beschreiber für die Untergruppen auf ähnliche Weise kombiniert werden.

[0058] Wenn es in einer Gruppe eine große Anzahl von Bildern gibt, enthält das obige Verfahren potentiell eine große Anzahl von Clustern und eine große Anzahl von Berechnungen von Abständen zwischen den Clustern.

[0059] Angesichts des Obigen berücksichtigt eine Variante des obigen Verfahrens den Umstand, dass die meisten Bilder in einer Videoaufnahme oder einer Sammlung und die entsprechenden Beschreiber sehr ähnlich sein werden. Das bedeutet, dass die meisten Cluster ohne signifikanten Verlust an Genau-

igkeit auf einer Framebasis zusammengeführt werden können.

[0060] Genauer ausgedrückt, berücksichtigt die Variante die Bilder einer Bildgruppe in einer Sequenz. Wie in dem oben beschriebenen Verfahren wurden die Bildbeschreiber bereits abgeleitet und in der Beschreiberdatenbank gespeichert. Die Clusterbeschreiber für ein erstes und zweites Bild in einer Bildgruppe werden abgerufen. Als nächstes werden die Abstände zwischen jedem Cluster in dem Bildpaar bestimmt. Wenn der Abstand zwischen jedem Clusterpaar unter einem vorgegebenen Schwellwert liegt, wird das Clusterpaar zusammengeführt. Die zusammengeführten Cluster und alle verbleibenden Cluster in dem Bildpaar werden in einem Superbeschreiber wie in dem obigen Beispiel gesammelt. Der Clusterbeschreiber für das nächste Bild wird dann abgerufen und mit dem Superbeschreiber für die ersten beiden Bilder auf dieselbe Weise zusammengeführt, indem Cluster, die nah sind, zusammengeführt werden und einen neuen Superbeschreiber bilden. Wenn die Clusterbeschreiber für alle Bilder betrachtet worden sind, wird der sich ergebende Superbeschreiber unter Verwendung aller verbleibender Cluster, wie in dem ersten Beispiel dargelegt, zusammengeführt.

[0061] In der obigen Beschreibung wurden die Clusterbeschreiber bereits abgeleitet. Alternativ kann das Bild abgerufen oder zugeführt und der Beschreiber aus dem Bild abgeleitet werden, bevor das Clusterzusammenführen ausgeführt wird.

[0062] In der ersten oben beschriebenen Ausführungsform wird ein Beschreiber eines GoFGoP abgeleitet, indem die Beschreiber für jedes Bild in der Gruppe aggregiert werden.

[0063] In einem ersten Beispiel werden die Bilder in dem Bild oder Pixelbereich aggregiert und dann wird ein dominanter Farbbeschreiber aus dem aggregierten Bild abgeleitet, um einen GoFGoP-dominanten Farbbeschreiber zu erzeugen. Wenn es in der Gruppe N Bilder **8** gibt, wobei jedes Bild $m \times m$ Pixel enthält, kann das Superbild **9** als $(N \times m) \times m$ -Array von Pixeln betrachtet werden, wie in [Fig. 6](#) gezeigt.

[0064] Im Gegensatz zur ersten Ausführungsform verwendet das erste Beispiel nicht die Bildbeschreiber für jedes Bild, sondern arbeitet direkt mit den Bildern.

[0065] Nach dem Aggregieren der Bilder wird aus dem Superbild unter Verwendung der oben in Bezug auf die erste Ausführungsform beschriebenen Technik ein dominanter Farbbeschreiber abgeleitet.

[0066] Ein Vorteil des ersten Beispiels ist, dass es keinen Verlust an Genauigkeit im Extraktionsprozess gibt. Zum Betrachten aller Bilder kann jedoch ein gro-

ßes Maß an Komplexität, insbesondere Speicher, erforderlich sein. Um das Problem der Komplexität zu überwinden, können die Bilder temporär und/oder räumlich abgetastet werden.

[0067] Im Obigen werden die Bilder in dem Bild- oder Pixelbereich aggregiert und dann wird ein dominanter Farbbeschreiber aus dem aggregierten Bild abgeleitet, um einen GoFGoP-dominanten Farbbeschreiber zu erzeugen. Alternativ könnte jedes Bild in dem Farbbereich aggregiert werden (zum Beispiel in Form eines Histogramms für jedes Bild) und der dominante Farbbeschreiber könnte aus den aggregierten Histogrammen abgeleitet werden. Ein Farbhistogramm für jedes Bild könnte beispielsweise abgeleitet werden oder aus einem Speicher abgerufen werden. Als nächstes werden die Farbhistogramme kombiniert, indem sie zur Bildung eines Superbildhistogramms addiert werden, welches durch die Anzahl von Bildern, die das Superbild bilden, normalisiert werden kann. Schließlich wird unter Verwendung der oben in Bezug auf die erste Ausführungsform beschriebenen Technik ein dominanter Farbbeschreiber aus dem Superbildhistogramm abgeleitet. Anders ausgedrückt, werden die Spitzen (dominante Farben) des Superbildhistogramms ausgewählt, sowie die jeweiligen Abweichungen und Gewichte.

[0068] Andere Techniken zum Aggregieren der Bilder vor dem Ableiten des dominanten Farbbeschreibers können verwendet werden. Statt die Histogramme für jedes Bild zu addieren, kann beispielsweise der Durchschnitt, Median oder Querschnitt der Bildgruppe berechnet werden. Der dominante Farbgruppenbeschreiber wird dann aus dem sich ergebenden Durchschnitts-, Median- oder Querschnittshistogramm abgeleitet.

[0069] In einer Variante des ersten Beispiels, welches für die Aggregation im Bild/Pickelbereich und im Farb/Histogrammbereich gilt, wird die Bildgruppe temporär abgetastet, wie unten beschrieben. In diesem Beispiel werden die Clusterbeschreiber für jedes Bild abgeleitet oder abgerufen. Auf der Basis der Ähnlichkeit der Clusterbeschreiber wird bestimmt, welche Bilder in der Gruppe beim Ableiten des GoFGoP-Beschreibers verwendet werden.

[0070] Das erste Bild in einer Sequenz aus der Bildgruppe bildet das Anfangssuperbild. Die folgenden Bilder werden verworfen, bis die Ähnlichkeit zwischen dem Clusterbeschreiber eines Bildes und dem letzten Bild, welches dem Superbild zugeführt wird (anfänglich das erste Bild) einen vorgegebenen Entscheidungszustand erfüllt. Wenn ein Bild den Zustand erfüllt, wird es dem Superbild zugefügt. Die folgenden Bilder in der Gruppe werden dann mit dem Bild verglichen, welches zuletzt dem Superbild zugefügt wurde, bis der vorgegebene Zustand wieder erfüllt wird usw., bis alle Bilder in der Gruppe betrachtet

worden sind. Der dominante Farbbeschreiber wird dann aus dem sich ergebenden Superbild abgeleitet.

[0071] Ein mögliches Entscheidungskriterium für die Bildähnlichkeit ist der Wert der Abgleichfunktion zwischen den jeweiligen dominanten Farbbeschreibern, wie in unserer gleichzeitig anhängigen Anmeldung WO 00/67203 beschrieben oder unter Verwendung einer Abgleichfunktion, wie in dem oben erwähnten MPEG-7 Buch beschrieben. Ein weiteres Kriterium könnte das Ergebnis der oben beschriebenen „Online“-Zusammenführung sein. Das „Entscheidungskriterium“ würde in diesem Fall erfüllt werden, wenn alle Cluster in den bestehenden Beschreiber zusammengeführt worden sind. Ein weiterer Ansatz, der das Extrahieren der dominanten Farbe für alle Bilder vermeiden würde, wäre die Berechnung eines rohen Farbhistogramms und die Verwendung des Histogrammabgleichfunktionswerts als das Kriterium.

[0072] Beide Kriterien erfordern das Festlegen zusätzlicher Parameter: einen Schwellwert, unter dem der Abgleichfunktionswert im ersten Fall als klein betrachtet wird und einem Zusammenführungsschwellwert im zweiten Fall. Ein alternativer Ansatz, der sich insbesondere im Fall eines beschränkten Speichers eignen würde, wäre die Anpassung des Schwellwertes, so dass die Anzahl von Bildern, die gesammelt werden, eine festgelegte Grenze nicht übersteigt.

[0073] Ein zweites Beispiel leitet einen GoFGoP-dominanten Farbbeschreiber aus der Gruppe dominanter Farbbeschreiber für die Bildgruppe ab. Genauer ausgedrückt, werden die jeweiligen dominanten Farbbeschreiber für eine Bildgruppe aus der Beschreiberdatenbank abgerufen (oder abgeleitet, wenn sie noch nicht abgeleitet worden sind).

[0074] Für jeden dominanten Farbbeschreiber wird der Abstand zwischen dem Beschreiber und jedem der verbleibenden Beschreiber für die Gruppe gemessen. Dies führt zu einer Gruppe von Abstandsmessungen für jeden Beschreiber, welche addiert werden können, um eine Gesamtstreckmessung für jeden Beschreiber zu ergeben. Der Beschreiber mit der kleinsten Gesamtstreckmessung wird als ein repräsentativer Beschreiber ausgewählt und als der GoFGoP-dominante Farbbeschreiber behandelt.

[0075] Es können andere Arten zum Auswählen des repräsentativen Beschreibers verwendet werden, die vorzugsweise Tests oder Vergleiche enthalten, welche mindestens einige der Bildbeschreiber für die Bilder in der Gruppe umfassen. Als weiteres Beispiel könnte dies, basierend auf dem Verzerrungsmaß erfolgen, wie in MPEG-7 definiert, wenn dies für alle Beschreiber erfolgt.

[0076] Es ist möglich, sehr nahe Beschreiber vorher

zu verwerfen, um die Berechnung zu verringern.

[0077] Oben werden verschiedene Verfahren zum Ableiten eines GoFGoP-Beschreibers, insbesondere eines GoFGoP-dominanten Farbbeschreibers, dargelegt.

[0078] Es gibt unterschiedliche Anwendungen für den GoFGoP-Beschreiber, wie beispielsweise beim Suchen nach und Abrufen von Bildgruppen. Beispielsweise kann ein Nutzer wünschen, nach Bildgruppen entsprechend einem Eingabebild oder einer Bildgruppe zu suchen.

[0079] Im Folgenden wird ein Suchverfahren umrissen.

[0080] Unter Bezugnahme auf [Fig. 1](#) wird von einem Nutzer ein Abfragebild unter Verwendung geeigneter Mittel, wie eines Scanners oder einer Digitalkamera, eingegeben oder durch Auswählen eines Abfragebilds aus einem Bildbereich, der von dem Computer angezeigt wird, oder durch Auswählen eines Bereichs eines solchen Bilds ausgewählt. Ein dominanter Farbbeschreiber für das Bild wird abgeleitet wie oben beschrieben. Die GoFGoP-dominanten Farbbeschreiber liegen im selbem Format vor wie ein dominanter Farbbeschreiber eines einzelnen Bilds und so kann der Abgleich beispielsweise unter Verwendung einer Abgleichfunktion, wie in WO 00/67203 dargelegt, oder ähnlichem oder unter Verwendung einer Abgleichfunktion, wie in dem Abschnitt 13.3.2 des oben genannten MPEG-7 Buchs dargelegt, ausgeführt werden. Optional kann der Abfragebeschreiber auch mit Bildbeschreibern einzelner Bilder verglichen werden, die in der Datenbank gespeichert sein können.

[0081] Die Ergebnisse der Abgleichfunktion werden geordnet und die Bildgruppen, für die die Abgleichfunktion die nächste Übereinstimmung anzeigt, werden abgerufen. Es können ein oder mehrere Bilder aus den nächsten Übereinstimmungen angezeigt werden.

[0082] Es können andere Verfahren zum Stellen einer Anfrage verwendet werden. Eine Anfrage kann gestellt werden, indem eine Bildgruppe ausgewählt wird und ein GoFGoP-Beschreiber wie oben für die Gruppe beschrieben extrahiert wird. Die Gruppe kann explizit ausgewählt werden, beispielsweise indem ein Bereich von Frames ausgewählt wird oder implizit ausgewählt werden, indem beispielsweise ein Schlüsselframe in einem Video ausgewählt wird, indem eine „Aufnahme“, die den Schlüsselframe enthält, dann unter Verwendung eines geeigneten Algorithmus abgeleitet wird.

[0083] Ein erfindungsgemäßes System kann beispielsweise in einer Bildbibliothek bereitgestellt wer-

den. Alternativ können die Datenbanken räumlich entfernt von der Steuereinheit des Systems angeordnet sein und mit der Steuereinheit über eine temporäre Datenstrecke, wie beispielsweise eine Telefonleitung oder ein Netzwerk, wie dem Internet, verbunden sein. Die Bild- und Beschreiberdatenbanken können beispielsweise im permanenten Speicher oder auf einem tragbaren Datenspeichermedium, wie beispielsweise CD-ROMS oder DVDs bereitgestellt werden.

[0084] Das oben als eine Ausführungsform der Erfindung beschriebene System liegt in Form eines Computersystems vor. Das Computersystem kann ein Standardcomputer sein, der unter Verwendung geeigneter Programme zum Ausführen eines Verfahrens gemäß einer Ausführungsform der Erfindung programmiert worden ist. Die Programme können in jedem geeigneten Speichermedium einschließlich fester oder permanenter Speicher oder entfernter Speichermitel gespeichert sein. Das System kann unter Verwendung spezifischer Hardware und/oder Software einschließlich beispielsweise spezifischer Chips modifiziert sein. Die Erfindung kann auch in einer speziell angepassten Vorrichtung implementiert sein einschließlich spezifischer Hardware und/oder Software.

[0085] In der obigen Beschreibung wurden die Farbdarstellungen hinsichtlich roter, grüner und blauer Farbkomponenten beschrieben. Natürlich können auch andere Darstellungen verwendet werden, wie beispielsweise eine Darstellung unter Verwendung von Farbton, Sättigung und Intensität oder ein YUV-Koordinatensystem oder eine Untergruppe von Farbkomponenten in jedem Farbraum, zum Beispiel nur Farbton und Sättigung in HSI.

[0086] Die oben beschriebene Ausführungsform der Erfindung verwendet Beschreiber, die für Bilder und Bildgruppen abgeleitet worden sind. Die Bildbeschreiber können jedoch auch für Bildbereiche gelten und analog können GoFGoP-Beschreiber auf Bildbereichen basieren. Bereiche können rechteckige Blöcke oder Bereiche unterschiedlicher Formen und Größe sein. Alternativ können die Beschreiber für Bereiche in dem Bild abgeleitet werden, welche Objekten beispielsweise einem Auto, einem Haus oder einer Person entsprechen. In jedem Fall können die Beschreiber für das gesamte Bild oder nur einen Teil davon abgeleitet werden. Ein GoFGoP-Gruppenbeschreiber könnte außerdem für ein einzelnes Bild abgeleitet werden, indem die obigen Verfahren auf eine Vielzahl von Bereichen in einem Bild, die eine Bildbereichsgruppe bilden, angelegt werden.

[0087] Bei dem Suchverfahren kann der Nutzer statt eine einfache Farbanfrage einzugeben oder einen Bildblock auszuwählen beispielsweise die Anzeigevorrichtung verwenden, um einen Bereich in einem Bild zu beschreiben, beispielsweise durch Umkrei-

sen, woraufhin die Steuereinheit einen Beschreiber für diesen Bereich ableitet und ihn zum Suchen verwendet, auf ähnliche Weise wie oben beschrieben. Statt der Verwendung der bereits in der Bilddatenbank gespeicherten Bilder zum Initiieren einer Suche könnte außerdem ein Bild in das System eingegeben werden unter Verwendung beispielsweise eines Bildscanners oder einer Digitalkamera. Um eine Suche in einer solchen Situation auszuführen, leitet das System wiederum zunächst Beschreiber für das Bild oder Bildbereiche ab, entweder automatisch oder wie durch den Nutzer bestimmt.

[0088] Geeignete Aspekte der Erfindung können unter Verwendung von Hardware oder Software implementiert werden.

[0089] In den obigen Ausführungsformen werden die Clusterverteilungen für jede repräsentative Farbe unter Verwendung von Gaußschen Funktionen angenähert und die Mittelwerte, Abweichungen und Kovarianzen solcher Funktionen werden in den Beschreiberverwerten verwendet. Andere Funktionen oder Parameter können jedoch verwendet werden, um die Komponentenverteilungen anzunähern, beispielsweise unter Verwendung von Basisfunktionen wie Sinus und Kosinus mit Beschreibern, die auf solchen Funktionen basieren.

Patentansprüche

1. Verfahren zur Darstellung einer Gruppe von Bildern (**8**) umfassend die Bestimmung der Werte von einer oder mehreren dominanten Farben (**1–5**) für die Gruppe von Bildern und zur Ableitung einer dominanten Farbdarstellung, die die Gruppe von Bildern hinsichtlich eines oder mehrerer der dominanten Farbwerte wiedergibt, wobei mindestens einige der Bilder in der Gruppe von Bildern durch eine entsprechende dominante Farbdarstellung hinsichtlich eines oder mehrerer dominanter Farbwerte für das Bild dargestellt werden, und wobei jede der dominanten Farbdarstellungen über eine oder mehrere Komponenten verfügt, umfassend mindestens einen dominanten Farbwert, wobei das Verfahren ferner die Kombination mehrerer dominanter Farbdarstellungen oder die Auswahl einer der dominanten Farbdarstellungen als repräsentativ für die Gruppe umfasst, gekennzeichnet durch die Kombination von einer oder mehreren Komponenten der Darstellungen basierend auf der Nähe der dominanten Farben im Farbraum.

2. Verfahren nach Anspruch 1, wobei jede der dominanten Farbdarstellungen über eine oder mehrere Komponenten verfügt, optional für jeden dominanten Farbwert umfassend eine Abweichung, die die Farbverteilungsabweichung für das Bild hinsichtlich des dominanten Farbwerts angibt, ein Farbgewicht, das die Einwirkung der dominanten Farbe im Bild, die An-

zahl der dominanten Farben im Bild und die räumliche Homogenität der Pixel entsprechend der dominanten Farben im Bild angibt.

3. Verfahren nach einem der vorhergehenden Ansprüche, das ein zeitliches oder räumliches Subsampling der Gruppe von Bildern (**8**) einschließt.

4. Verfahren nach Anspruch 3, wobei Bilder in Abhängigkeit ihrer Ähnlichkeit zu anderen Bildern in der Gruppe weggelassen oder eingebunden werden.

5. Verfahren nach Anspruch 3 oder Anspruch 4, wobei das Subsampling die Schritte aus Anspruch 1 einschließt.

6. Vorrichtung, die zur Implementierung eines Verfahrens nach einem der vorhergehenden Ansprüche eingerichtet ist.

7. Vorrichtung nach Anspruch 6, die Speichermitel (**18, 20**) zum Speichern von Bilddaten und/oder Deskriptordaten, Verarbeitungsmittel (**22**) und Steuerungsmittel (**12**) einschließlich Anweisungen für die Implementierung eines Verfahrens nach einem der Ansprüche 1 bis 10 umfasst.

8. Computerprogramm zum Ausführen des Verfahrens nach einem der Ansprüche 1 bis 5.

9. Ein Computer-lesbares Speichermedium, das ein Computerprogramm zum Ausführen des Verfahrens nach einem der Ansprüche 1 bis 5 speichert.

Es folgen 3 Blatt Zeichnungen

Anhängende Zeichnungen

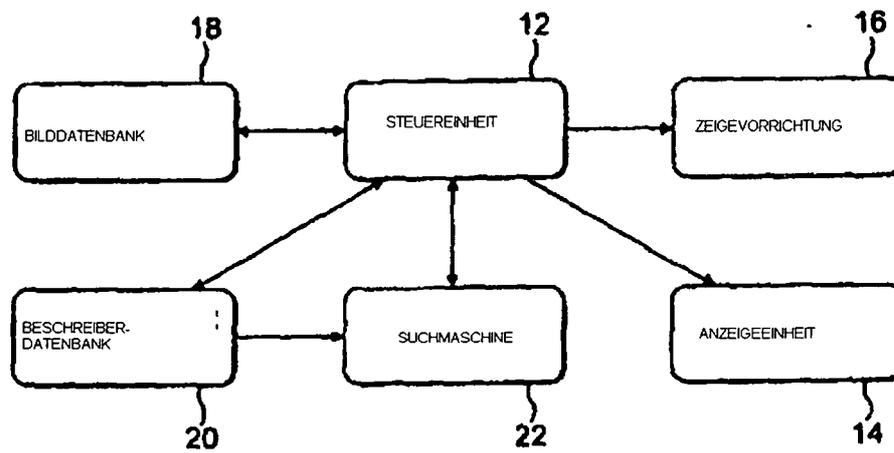


FIG. 1

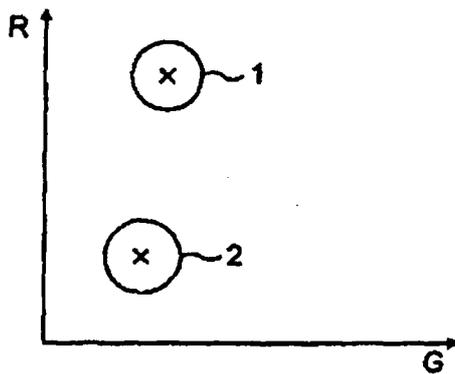


FIG. 2a

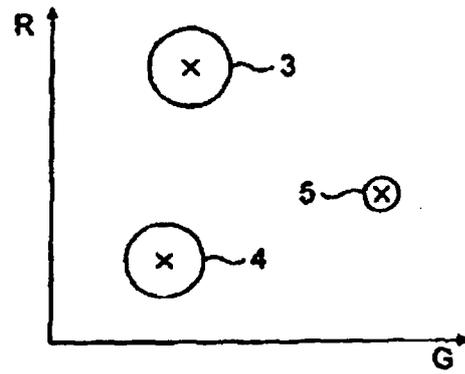


FIG. 2b

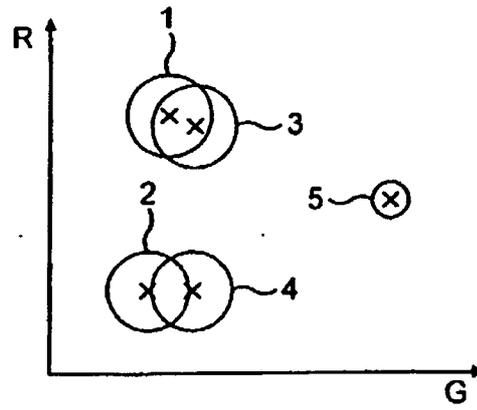


FIG. 3

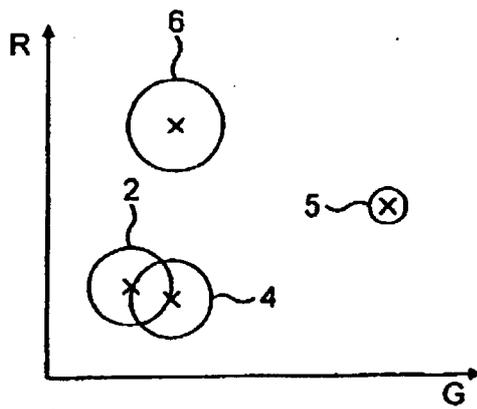


FIG. 4

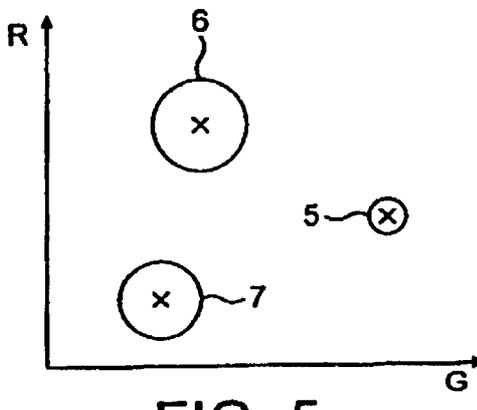


FIG. 5

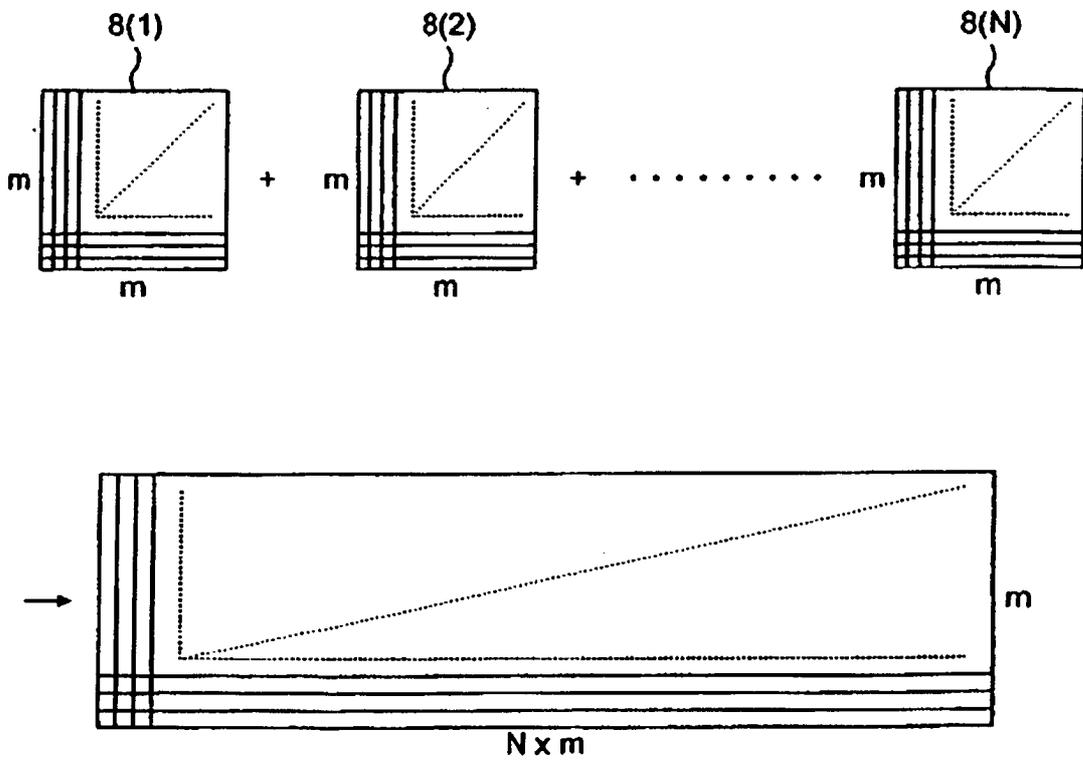


FIG. 6