



(19)
Bundesrepublik Deutschland
Deutsches Patent- und Markenamt

(10) DE 600 35 910 T2 2008.05.21

(12)

Übersetzung der europäischen Patentschrift

(97) EP 1 395 916 B1

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

(21) Deutsches Aktenzeichen: 600 35 910.7

(86) PCT-Aktenzeichen: PCT/EP00/06082

(96) Europäisches Aktenzeichen: 00 952 979.3

(87) PCT-Veröffentlichungs-Nr.: WO 2001/003002

(86) PCT-Anmeldetag: 29.06.2000

(87) Veröffentlichungstag

der PCT-Anmeldung: 11.01.2001

(97) Erstveröffentlichung durch das EPA: 10.03.2004

(97) Veröffentlichungstag

der Patenterteilung beim EPA: 08.08.2007

(47) Veröffentlichungstag im Patentblatt: 21.05.2008

(30) Unionspriorität:

347214 02.07.1999 US

(84) Benannte Vertragsstaaten:

DE, FR, GB

(73) Patentinhaber:

Koninklijke Philips Electronics N.V., Eindhoven,
NL

(72) Erfinder:

VAITHILINGAM, Gandhimathi, NL-5656 AA
Eindhoven, NL; ABDEL-MOTTALEB, Mohamed,
NL-5656 AA Eindhoven, NL

(74) Vertreter:

Volmer, G., Dipl.-Ing., Pat.-Anw., 52066 Aachen

(54) Bezeichnung: META-DESKRIPTOR FÜR MULTIMEDIA-INFORMATION

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

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

Beschreibung

Hintergrund der Erfindung

[0001] Die vorliegende Erfindung bezieht sich auf inhaltsbasierte Verarbeitung von Multimedia-Daten, und insbesondere auf die Schaffung und die Verwendung von Attributen von Multimedia-Daten, die für den Inhalt derselben beschreibend sind.

[0002] Multimedia-Information gibt es typischerweise in mehreren inhomogenen Formen, einschließlich beispielsweise digital, analog (beispielsweise VCR-Magnetband und Audiomagnetband), optisch (beispielsweise herkömmlicher Film), Bild (beispielsweise Bilder und Zeichnungen auf Papier) usw. Die Fähigkeit, die Multimedia-Information ausfindig zu machen, ist in der modernen Gesellschaft wichtig und ist insbesondere in vielen professionellen und Konsumentenapplikationen wie beispielsweise Ausbildung, Journalismus (beispielsweise das Aufsuchen von Reden eines bestimmten Politikers unter Verwendung seines Namens, seiner Stimme oder seines Gesichtes), Tourismusinformation, Kulturdiensten (beispielsweise historische Museen, Kunstmuseen usw.), Unterhaltung (beispielsweise das Suchen nach einem Spiel oder nach Karaoke-Titeln), Untersuchungsdiensten (beispielsweise menschliche Erkennung von Charakteristiken und Spurenabsicherung), geographischen Informationssystemen, Fernabtastung (beispielsweise Kartographie, Ökologie, Verwaltung natürlicher Quellen usw.), Überwachung (beispielsweise Verkehrsteuerung, Oberflächentransport, nicht zerstörerisches Testen in feindlicher Umgebung, usw.), Biomedizinischen Applikationen, Einkaufen (beispielsweise Suche nach Kleidung, die man mag), Architektur, Grundbesitz, Innenarchitektur, Soziales (beispielsweise Partnersuche) und Film-, Video- und Rundfunkarchiven wichtig. Bedauerlicherweise sind die heutigen Systeme nicht sorgfältig, schnell oder effizient in der Suche von Multimedia-Information; siehe beispielsweise "International Organisation for Standardisation ISO/IEC JTC1/SC29/WG11 Coding of Moving Pictures and Audio, MPEG-7 Applications Document V.8", Nr. N2728, März 1999.

[0003] Ein wichtiger Schritt in der Unterstützung der Suche nach Multimedia-Information ist die Darstellung derselben in einer Form, die unter Anwendung moderner Computersysteme suchbar ist. Sehr interessiert war man an der Entwicklung von Formen von audiovisueller Informationsdarstellung, die über die einfachen Wellenform- oder abtastbasierten Darstellungen, die kompressionsbasierten Darstellungen wie MPEG-1 und MP-WG-2, und die objektbasierten Darstellungen, wie MPEG-4 hinaus geht, und einer Anordnung oder einem Computercode zugeführt werden kann oder worauf von einer Anordnung oder einem Computercode zugegriffen werden kann. Viele geeignete Lösungen wurden zum beschreiben von Multimedia-Inhalt und zum Extrahieren der Darstellungen und zum Abfragen der Ergebnissammlungen der Darstellungen entwickelt, aber diese haben nur noch mehr heterogene Multimedia-Information gewuchert und haben die Schwierigkeiten der Durchführung schneller und effizienter Suchen nach Multimedia-Information noch weiter verschlimmert.

[0004] Ein "Deskriptor" ist eine Darstellung eines Merkmals, wobei ein "Merkmal" eine bestimmte Charakteristik von Multimedia-Information ist, ungeachtet der Medien oder der Technologie der Multimedia-Information und ungeachtet wie die Multimedia-Information gespeichert, codiert, wiedergegeben und übertragen wird. Da Deskriptoren, die in verschiedenen firmeneigenen Multimedia-Informationserfassungssystemen verwendet werden, nicht unbedingt kompatibel sind, hat das Schaffen eines Standards zum Beschreiben von Multimedia-Inhalt Interesse erweckt, was die Betriebsanforderungen von Rechensystemen unterstützt, die Multimedia-Information schaffen, austauschen, erfassen und/oder neu verwenden. Beispiele umfassen Rechensysteme, entworfen zum Bildverständnis (beispielsweise Überwachung, intelligenter Vorstellung, kluge Kameras) Media-Umwandlung (beispielsweise Sprache in text, Bild in Sprache, Sprache in Bild), und Informationserfassung (schnell und effizient Suchen nach mehreren Typen von Multimediadokumenten, die für den Benutzer interessant sind) und Filterung (zum Empfangen nur derjenigen Multimedia-Daten, die den Vorlieben des Benutzers entsprechen) in einem Strom von Audiovisueller Inhaltsbeschreibung.

[0005] In "Multimedia content description in the Infopyramid" von Chung-Sheng Li u. a., "Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing", Seattle, Wiedergabebeordnung, USA wird ein Multimedialinhaltbeschreibungssystem beschrieben, das viele Inhaltsdeskriptoren gruppiert.

[0006] Auf entsprechende Weise gibt es ein Bedürfnis nach einem Standard zur Beschreibung von Multimedialinhalt, die diese Betriebsanforderungen sowie andere Betriebsanforderungen, die noch entwickelt werden sollen, unterstützt.

Zusammenfassung der Erfindung

[0007] Auf entsprechende Weise ist es u. a. eine Aufgabe der vorliegenden Erfindung, wie dies in speziellen Ausführungsformen verwirklicht worden ist, die Effizienz der Erfassung von Multimedia-Information aus einem Aufbewahrungsort zu verbessern.

[0008] Es ist eine andere Aufgabe der vorliegenden Erfindung, wie in speziellen Ausführungsformen verwirklicht, die Geschwindigkeit der Erfassung von Multimedia-Information aus einem Aufbewahrungsort zu verbessern.

[0009] Noch eine andere Aufgabe der vorliegenden Erfindung ist es, wie in speziellen Ausführungsformen verwirklicht, einen Standarddarstellung eines Merkmals von Multimedia-Information zu schaffen.

[0010] Die vorliegende Erfindung ist durch die Hauptansprüche 1, 7 und 11 definiert. Die Unteransprüche definieren vorteilhafte Ausführungsformen.

[0011] Der Vollständigkeit halber sei auf WO 00/45341 verwiesen, die in Übereinstimmung mit den Artikeln 54(3) und 54(4) EPC als Stand der Technik betrachtet wird. Dieses Dokument beschreibt die Verwendung eines Bezugsbildes mit Merkmalen, denen Gewichtungen zugeordnet worden sind. Das Bezugsbild wird in das Suchsystem eingegeben zum Suchen ähnlicher Bilder. Die Suchergebnisse werden von einem Benutzer betrachtet. Wenn der Benutzer autorisiert ist, wird Benutzerrückkopplung in Bezug auf die Relevanz der Suchergebnisse verwendet zum Aktualisieren der Gewichtungen. Die aktualisierten Gewichtungen für dieses Bezugsbild werden für eine nächste Suche beibehalten. Auf entsprechende Weise lehrt das Dokument, dass ein Eingangsbezugsbild Gewichtungen erhalten hat für Merkmale, damit es ermöglicht wird, einen Ähnlichkeitsgrad zwischen dem Bezugsbild und anderen Bildern, die gesucht und durchstöbert werden, zu definieren.

Kurze Beschreibung der Zeichnung

[0012] [Fig. 1](#) ist eine Darstellung mehrerer Formen von Multimedia-Daten an einem Aufbewahrungsort.

[0013] [Fig. 2](#) ist ein Flussdiagramm eines Metadeskriptorerzeugungsprozesses und eines Multimediaqualitätsprozesses nach der vorliegenden Erfindung.

[0014] [Fig. 3](#) eine Tabelle von Daten für eine illustrative relationale Datenbank nach der vorliegenden Erfindung.

[0015] [Fig. 4](#) ist ein Flussdiagramm eines Prozesses zum Verfeinern von Metadeskriptoren für Multimedia an einem Aufbewahrungsort nach der vorliegenden Erfindung.

Detaillierte Beschreibung der bevorzugten Ausführungsform

[0016] Beispiel der jeweiligen Formen, in denen Multimedia-Daten bestehen können, sind in [Fig. 1](#) dargestellt und umfassen mehrere Auftritte von beispielsweise Graphiken **10**, Standbildern **12**, Video **14**, Film **16**, Musik **18**, Sprache **20**, Geräusche **22** und anderer Medien **24**. Diese vielen Auftritte können verschiedenartig gespeichert, codiert oder übertragen werden, es gibt sie auf verschiedenen Median, oder sie sind mit verschiedenen Technologien erzeugt worden. Die Multimedia-Daten können an einer einzigen Stelle gespeichert oder über die Welt verteilt sein; so können beispielsweise digitalisierte Multimedia-Daten, die für einen Benutzer interessant sind, in einer abgeschlossenen relationalen oder objektorientierten Datenbank gespeichert sein, oder in einzelnen unabhängigen Datenbanken, implementiert in verschiedenen Technologien und gespeichert an verschiedenen betriebseigenen Computer, verteilt über die ganze Welt und nur über Interner zugreifbar. Tatsächlich können nicht digitale Multimedia, die für einen Benutzer interessant sind, in einer Sammlung gespeichert sein, und zwar unter Ansteuerung einer einzigen Entität, oder weiter verbreitet in verschiedenen Sammlungen unter Ansteuerung verschiedener Entitäten. Ungeachtet der Form, in der die Multimedia-Information existiert und wie diese gespeichert ist, würde der Benutzer bevorzugen, dass er die Sammlung von Multimedia-Information als einen einzigen Aufbewahrungsort betrachten könnte, wie durch das Bezugszeichen **1** angegeben, und zwar zwecks der effizienten Suche nach bestimmten Multimedia-Daten.

[0017] Es hat sich herausgestellt, dass Multimedia-Informationserfassung, die nebst Deskriptoren Meta-Deskriptoren verwenden, nicht nur bei dem Identifizieren von Multimedia-Information effizient ist, sondern auch imstande ist, Multimedia-Information zu identifizieren, die in einer verschiedenartig dargestellt ist. Ein "Deskrip-

tor" ist eine Darstellung eines Merkmals, wobei ein "Merksmal" eine bestimmte Charakteristik von Multimedia-Information ist, während ein "Meta-Deskriptor" Information über den Deskriptor ist. Ein Meta-Deskriptor ist abweichend von aber relativiert an dem allgemeinen Konzept von Meta-Daten, was eine durchaus bekannte Art und Weise ist, zusätzliche Information einzubetten. So können beispielsweise Meta-Daten in Dokumenten das Format der Bilder in dem Dokument umfassen, und Meta-Daten in einer Datenbank können Wertbeschränkungen oder statistische Information für ein Attribut in einer Beziehung enthalten. Insbesondere identifiziert ein Meta-Deskriptor für ein Item von Multimedia-Information diejenigen Teile eines Deskriptors, für dieses Item der Multimedia-Information, welche die nützlichste Information zum Identifizieren dieses Items der Multimedia-Information enthalten. Das Konzept des Meta-Deskriptors basiert auf der Voraussetzung, dass ein bestimmtes Multimediainformationsitem am besten qualifiziert wird, zu wissen, was es am besten beschreibt, und das Angenommen dieser Information das inhaltsbasierte Suchen stark verbessert. Auf vorteilhafte Weise ermöglichen Meta-Deskriptoren, dass computerisierte Suchvorgänge für Multimedia-Information schneller durchgeführt werden, und zwar wegen der im Allgemeinen geringeren Größe von Meta-Deskriptoren, wie effizienter durchgeführt werden, und zwar wegen der Eliminierung weniger relevanter Information

[0018] Obschon viele Multimediaerfassungstechniken an die Verwendung von Meta-Deskriptoren angepasst werden können, ist eine bevorzugte Technik zum Suchen nach Multimediainformation unter Verwendung von Meta-Deskriptoren das Szenario Suche, oder Abfrage anhand eines Beispiels. Bei der Suche entsprechend Abfrage durch Beispiel von beispielsweise Bildern, wird ein bestimmtes Standbild als Basis zur Abfrage am Aufbewahrungsort spezifiziert. Die Abfrage wird entweder durch den Auslöser der Abfrage spezifiziert, der ein menschlicher Benutzer oder beispielsweise ein automatischer Prozess sein kann, oder durch Erfassungsalgorithmen, die in dem Erfassungsprozess angewandt werden. Die Abfrage wird gebildet auf Basis eines Merkmals oder auf Basis von Merkmalen, wovon in den Meta-Deskriptoren der Multimedia-Information an dem Aufbewahrungsort angegeben ist, dass sie wichtig sind. Deskriptoren für das Merkmal oder die Merkmale werden aus der Abfrage-Multimedia-Information extrahiert und mit Deskriptoren verglichen, die aus der Aufbewahrungsort-Multimedia-Information extrahiert wurden zum Erhalten von Ähnlichkeitsmaßen, die angewandt werden zum selektieren eines oder mehrerer "passenden" Multimediainformationsitems. In einigen Fällen können Meta-Deskriptoren für die Aufbewahrungsort-Multimedia-Information unmittelbar eine wahrscheinliche und große Unähnlichkeit im Inhalt angeben, wobei die Berechnungen zum Durchführen einer Extraktion von Deskriptoren und ein Vergleich von Merkmalen für die Abfrage und die betreffende Aufbewahrungsort-Multimedia-Information vermieden werden. Wenn der Benutzer nicht ein Informationsabfrageexperte ist, ist das betreffende Merkmal oder sind die betreffenden Merkmale, die in der Erfassung übereinstimmender Multimedia-Information verwendet werden, für den Benutzer der Einfachheit der Verwendung wegen, transparent.

[0019] [Fig. 2](#) ist ein Flussdiagramm, das Prozesse zum Erzeugen von Meta-Deskriptoren für Aufbewahrungsort-Multimedia-Information und zum Durchführen einer Abfrage des Aufbewahrungsortes zeigt. Obschon Meta-Deskriptoren für jeden beliebigen Multimedia-Informationstyp verwendet werden können, basiert das in [Fig. 2](#) dargestellte Beispiel auf Standbildern zur Erleichterung der Beschreibung. Ein Verfahren **110** zum Erzeugen von Meta-Deskriptoren ist durch die Hauptschritte **111–114** illustriert und ein Verfahren **130** zum Abfragen eines Multimediaaufbewahrungsortes, wie in [Fig. 1](#) dargestellt, ist durch die Hauptschritte **131–137** illustriert. Das Meta-Deskriptorerzeugungsverfahren **110** ist ein nicht überwachtes oder automatisches Verfahren zur Maschinenlernung, obschon Meta-Deskriptoren auch durch Formalisierung von Benutzereingabe durch einen Menschen oder durch ein Hybrid halbautomatischer Techniken erzeugt werden können. Das Abfrageverfahren **130** wird vorzugsweise automatisiert, ausgenommen die Benutzerformulierung einer Abfrage. Es sind mehrere Aspekte von Deskriptorerzeugung und Multimedia-Informationserfassung durchaus bekannt und sind in mehreren Veröffentlichungen beschrieben worden, einschließlich beispielsweise Yong Rui, Thomas S. Huang, und Shih-Fu Chang, "Image Retrieval: Past, Present, and Future", "Journal of Visual Communication and Image Representation", 10, 1-23(1999); Sharad Mehrotra, Yong Rui, Micheal Ortega-Binderberger, und Thomas S. Huang: "Supporting Content-based Queries over Images in MARS", "Proceedings of the IEEE International Conference an Multimedia Computing and Systems", Juni 3–6, 1997, Chateau Laurier, Ottawa, Ontario, Canada, 1997, Seiten 632–633; Sharad Mehrotra, Yong Rui, Kaushik Chakrabarti, Michael Ortega-Binderberger und Thomas S. Huang: "Multimedia Analysis and Retrieval System", "Proceedings of the 3rd International Workshop an Information Retrieval Systems", Como, Italien, September 25–27, 1997, Seiten 39–45; und Patrick M Kelly, Michael Cannon und Donald R. Hush: "Query by Image Example: the CANDID Approach" in "SPIE" Heft 2420 "Storage and Retrieval for Image and Video Databases III", 1995, Seiten 238–248.

[0020] Der erste Schritt **111** in dem Meta-Deskriptorerzeugungsverfahren **110** ist Extraktion von Deskriptoren aus Multimedia-Information (MM). Deskriptorextraktionsalgorithmen und Software, sowie Indexierungsalgorithmen und Software sind in dem betreffenden technischen Bereich durchaus bekannt, und Beispiele sind in den oben genannten Veröffentlichungen beschrieben. Mit Ausnahme für Beschränkungen, die durch betreffen-

de Standards auferlegt werden, kann jede beliebige Deskriptorextraktionstechnik oder jede beliebige Kombination von diesen Techniken nach Wunsch angewandt werden. Typischerweise umfassen Deskriptoren von Standbildern beispielsweise Farbe, Form, Struktur und Umriss, obschon der betreffende extrahierte Deskriptor zu jedem beliebigen Zeitpunkt von der in einem automatisierten Prozess angewandten Software, den Aktionen des Benutzers in einem manuellen Prozess, oder von den Interaktionen des Benutzers und der Software in einem hybriden Prozess abhängig ist. Automatisierte Prozesse sind günstig, da sie große Mengen Multimedia-Information an dem Aufbewahrungsort verarbeiten kann ohne dass dazu wesentliche menschliche Intervention erforderlich ist. Typischerweise arbeiten automatisierte Prozesse, die mit Standbildern arbeiten, nicht auf einem semantischen Pegel, was bedeutet, dass sie nicht ein Familienporträt in derartigen Termen beschreiben, sondern vielmehr Werte von Farbe, Form, Struktur und Umriss erzeugen für vielleicht das ganze Standbild, oder vielleicht für mehrere Blöcke, in die das Bild aufgeteilt worden ist.

[0021] Der nächste Schritt **112** ist, die Multimedia-Information auf Basis der Deskriptoren zu Cluster, obschon andere Techniken gewünschtenfalls angewandt werden können und Beispiele in den oben genannten Veröffentlichungen beschrieben worden sind. Im Wesentlichen ist Cluster das Gruppieren ähnlicher Multimedia-Information aus einem großen gemischten Datensatz – Cluster ist nicht notwendig für kleine Aufbewahrungsorte von Information gleichen Typs – auf Basis bestimmter Kriterien, angewandt auf die Deskriptoren. Ein Cluster ist ein Satz von Entitäten, die einander ähnlich sind, und Entitäten von verschiedenen Cluster sind einander nicht ähnlich. Extraktion von Meta-Deskriptoren kann durch beaufsichtigtes oder unbeaufsichtigtes Cluster erfolgen. Extraktion von Meta-Deskriptoren durch beaufsichtigtes Cluster betrifft das Cluster von Multimediainhalt auf Basis der Merkmale, was einen Satz von Clusterrepräsentativen ergibt, denen vorher ein Kennzeichen oder ein Deskriptor zugeordnet worden ist. Den Bildern in jedem Cluster werden danach das Kennzeichen oder der Deskriptor dieser Vertreter des Clusters zugeordnet. Extraktion von Meta-Deskriptoren durch unbeaufsichtigtes Cluster betrifft das Cluster von Multimediainhalt auf Basis jedes beschriebenen Merkmals. So hat beispielsweise ein Satz von Bildern eine erste Clusterdarstellung auf Basis der Farbmerkmale und er hat eine zweite Clusterdarstellung auf Basis der Strukturmerkmale. Auf Basis beispielsweise des Vergleichs der Cluster für jedes Merkmal, wobei programmierte Metrik angewandt wird um zu berechnen, wie deutlich definiert und kompakt die Cluster sind, wird ein einziges Merkmal oder werden einige Merkmale gefunden um andere beim Beschreiben eines bestimmten Bildes zu übertreffen. So kann beispielsweise ein bestimmtes Bild zu einem sehr kompakten und deutlich definierten Cluster in dem Satz von Cluster gehören, und zwar auf Basis von Farbmerkmalen, kann aber auch zu einem Cluster gehören mit einer weiten Verbreitung und Überlappung in dem Satz von Cluster auf Basis von Strukturmerkmalen. Das Farbmerkmal wird auf entsprechende Weise als Meta-Deskriptor für das Bild gewählt, da es das Bild besser klassifiziert als das Strukturmerkmal in den betreffenden Merkmalräumen.

[0022] Mathematisch ist ein Cluster eine Anhäufung von Punkten in dem Testraum, so dass der Abstand zwischen beliebigen zwei Punkten in dem Cluster kleiner ist als der Abstand zwischen jedem beliebigen Punkt in dem Cluster und einem beliebigen Punkt nicht in demselben. Siehe beispielsweise Anil K. Jain und Richard C. Dubes: "Algorithms for Clustering Data", Prentice Hall Advanced Reference Series, 1998, Seite 1. Clusteranalyse ist der Prozess der Klassifizierung von Objekten zu Subsätzen, die Bedeutung in dem Kontext eines bestimmten Problems haben. Die Objekte werden dabei zu einer effizienten Darstellung organisiert, welche die Besetzung, die abgetastet wird, kennzeichnet. Die Beziehung zwischen Objekten wird in einer Annäherungsmatrix dargestellt, in der Reihen und Spalten Objekten entsprechen. Wenn die Objekte als Muster, oder Punkte in einem d-dimensionalen metrischen Raum gekennzeichnet werden, kann die Nähe ein Abstand zwischen Paaren von Punkten sein, wie ein Euklidischer Abstand. Typischerweise ist die Nähematrix die einzige Eingabe zu einem Clusteralgorithmus. Die Objekte, die geclustert werden, könnten verschiedene Pflanzenarten, Pixel in einem digitalen Bild oder Dokumente über verschiedene Themen sein. Clusteranalyse findet verschiedene Applikationen in der Mustererkennung und Bildverarbeitung. Clusteranalyse wird zur Merkmalselektion angewandt, in Applikationen, die unbeaufsichtigtes Lernen betreffen, und als Hilfe bei Maschinenlernen und bei der Darstellung von Kenntnissen in künstlicher Intelligenz.

[0023] Deskriptoren. Extrahiert aus Standbildern, sind typischerweise Vektoren von mehrdimensionalen Nummern, die eine Sammlung von Punkten im Raum darstellen. In dem einfachen Fall vergleicht der Clusteralgorithmus Abstände in einer Sammlung von Punkten in einem zweidimensionalen Raum um zu ermitteln, wie nahe sich die Punkte liegen. Bei komplizierteren Algorithmen ist das Konzept auf den mehrdimensionalen Raum erweitert. Um Konsistenz zu gewährleisten wird der Clusterprozess vorzugsweise nur auf Multimedia-Information angewandt, die mit demselben Extraktionsalgorithmus verarbeitet worden ist. Für Standbilder basiert das Clustern auf Ähnlichkeit von typischerweise Tiefpegelmerkmalen, wobei bestimmte Bilder derselben deutliche Muster zusammenfügen für eine bestimmte Farbe und Struktur (beispielsweise aus einer Strandszene erzeugt werden würde), aber können mit sonst undefinierten Ergebnissen für Umriss und Form als in demsel-

ben Cluster vorhanden betrachtet werden. Unbeaufsichtigte Clusteralgorithmen wirken typischerweise iterativ, wodurch die Ergebnisse verfeinert werden, bis ein Schwellenpunkt, spezifiziert durch den Benutzer erreicht wird.

[0024] Der nächste Schritt **113** ist jedem Cluster Meta-Deskriptoren zuzuordnen. Dort, wo beispielsweise ein Muster nur für Farbe entsteht, wird dem Cluster ein Meta-Descriptor von Farbe 100% zugeordnet. Dort, wo ein Muster nur für Umriss entsteht, was für monochromatische technische Zeichnungen erwartet werden könnte, wird beispielsweise ein Meta-Descriptor des Umrisses 100 dem Cluster zugeordnet. Für das bestimmte Cluster von Standbildern des vorhergehenden Beispiels, in dem deutliche Muster nur für eine bestimmte Farbe und Struktur entstehen, wird dem Cluster beispielsweise ein Meta-Descriptor des Umrisses 100% zugeordnet. Für das bestimmte Cluster von Standbildern des vorhergehenden Beispiels, in denen deutliche Muster nur für eine bestimmte Farbe und Struktur entstehen, wird dem Cluster ein Meta-Descriptor von Farbe 50% und Struktur 50% zugeordnet. Wenn Prozessautomatisierung erwünscht ist, können Gewichtungen unter Anwendung heuristischer Regeln zugeordnet werden, die auf statistischer Information von vorhergehenden Routinen mit Multimedia-Information basieren. Auf alternative Weise können Meta-Deskriptoren, gewünschtenfalls, von Hand zugeordnet werden, oder, gewünschtenfalls, auf halbautomatische Weise mit menschlicher Interaktion.

[0025] Meta-Deskriptoren können jede beliebige Form annehmen, die für den Programmierer bequem ist. In einer bestimmten kompakten Form ist der Meta-Descriptor ein binärer Vektor X , wobei jedes Bit x_i die Relevanz ($x_i = \text{das i. Merkmal ist relevant}$) eines Merkmals, dem eine feste Anzahl geordneter Merkmale für diese Kategorie von Multimedia-Inhalt gegeben worden ist, angibt. Im Falle eines Standbildes beispielsweise ist ein geeigneter Vektor ein Vier-Bit-Vektor, in dem eine binäre 1 oder 0 die Bedeutung bzw. die Irrelevanz von Farbe, Form, Struktur und Umriss beim beschreiben der Multimedia-Information angibt. Wenn nur Farbe in einem bestimmten Standbild wichtig ist, ist ein geeigneter Meta-Descriptor **1000**.

[0026] Eine Notierung, die imstande ist, spezifische Gewichtungen zuzuordnen, benutzt einen gewichteten Vektor X , wobei jedes Element des Vektors x_i die dem i. Merkmal zugeordnete Gewichtung angibt, wenn eine feste Anzahl geordneter Merkmale für diese Kategorie von Multimedia-Inhalt gegeben ist. Wenn, sagen wir, Farbe und Umriss beide wichtig sind, aber je eine andere Gewichtung haben, ist ein geeigneter Meta-Descriptor dieses Typs "70,0,030" was angibt, dass Farbe eine 70% Gewichtung hat und Umriss eine 30% Gewichtung.

[0027] Eine andere Form ist eine Reihennotierung, die imstande ist, nicht nur verschiedene Gewichtungen, sondern auch verschiedene Extraktionsalgorithmen und Aufteilung des Standbildes in mehrere Blöcke zu verarbeiten. Weiterhin ermöglicht eine Reihendarstellung dass neue Merkmale später in dem Meta-Descriptor betrachtet werden. In der Reihennotierung gibt jedes Charakter oder jede Gruppe von Charakteren in der Reihe die Relevanz eines Merkmals an, wenn ein Satz von Merkmalen entsprechend einer vorher definierten Standbildnotierung gegeben ist. Es wird beispielsweise ein Satz von vier gültigen Farb-Descriptortypen betrachtet: (1) ein einziges Histogramm für ein ganzes Bild in dem RGB Farbraum; (2) fünfundzwanzig Histogramme für das Bild in dem RGB Farbraum, der in ein 5×5 Gitter aufgeteilt ist, wobei jeder der resultierenden fünfundzwanzig Blöcke durch ein Histogramm dargestellt ist; (3) ein einziges Histogramm für das ganze Bild in YUV Farbraum; und (4) fünfundzwanzig Histogramme für das Bild in YUV Farbraum, der in ein 5×5 Gitter aufgeteilt ist, wobei jeder der resultierenden fünfundzwanzig Blöcke durch ein Histogramm dargestellt wird. Es wird nun vorausgesetzt, dass diese Descriptortypen numerisch geordnet sind, und zwar von 1 bis "n", wobei n die Anzahl gültiger Descriptortypen ist, in dem vorliegenden Fall vier. Eine geeigneter Reihe-Meta-Descriptor für ein Standbild, das am besten durch beispielsweise den ersten und den vierten Descriptor beschrieben wird, ist "C214", was Folgendes bedeutet: das Farbmerkmal ist relevant (C) und wird mit zwei (2) Farbdarstellungen aus einem vordefinierten Satz von Farbdarstellungen erhalten, und zwar der ersten und der vierten (14) Farbdarstellung aus dem vordefinierten Satz von Farbdarstellungen. Die Reihennotierung ist insbesondere flexibel, wodurch es nicht nur möglich wird, dass verschiedene Farträume (beispielsweise RGB und YUV) identifiziert werden, sondern auch dass es ermöglicht wird, dass jeder Farbraum anders berechnet wird (beispielsweise als einen einzigen Block, als Satz von zehn Blöcken, als Satz von 100 Blöcken usw.). Erweiterungen der Reihennotierung können auch verschiedene Extraktionsalgorithmen durch geeignete vordefinierte Codes verarbeiten.

[0028] Meta-Deskriptoren verschiedener Formen können für verschiedene Multimedia-Informationen verwendet werden, und jede beliebige Information, die nicht in der betreffenden Form des Meta-Descriptor vor gesehen ist, kann standardmäßig geliefert werden. Wenn beispielsweise das standardmäßige Descriptortextraktionsverfahren und der standardmäßige Farbraum verwendet werden, ist ein binärer Meta-Descriptor ausreichend.

[0029] Der nächste Schritt **114** ist, das Anhängen von Meta-Deskriptoren an Multimedia-Information auf Basis von Clusterinformation. Eine Varietät verschiedener "Anhängetechniken" sind durchaus bekannt und können zur Anwendung auf Basis des Multimediatyps und der Art des Zugriffs darauf selektiert werden, und es können verschiedene Anhängetypen innerhalb einer bestimmten Datenbank von Meta-Deskriptoren verwendet werden. Die Deskriptoren selber können ggf. vorhanden sein, obschon, wenn sie nicht vorhanden sind, das System wissen soll, wie sie berechnet werden müssen, entweder standardmäßig oder durch einen Wert in der Meta-Deskriptornotierung. Vorzugsweise werden wenigstens die Meta-Deskriptoren und ihre Anhängseldaten werden in dem Speicher **120** gespeichert, der jeder beliebige Typ von Datenbank sein kann, worauf das System zugreifen kann. Deskriptoren können in dem Speicher **120** gespeichert werden oder mit der Multimedia-Information gespeichert werden, aus der sie extrahiert wurden. Die Multimedia-Information selber verbleibt an einem Aufbewahrungsort ([Fig. 1](#)), der ebenso spezifischen sein kann wie ein anderer Speicherraum in der Speicheranordnung **120**, oder so divers wie das Internet, oder sogar so divers, dass die Speicherung nicht elektronische Speicherformen, wie Papier, umfasst.

[0030] Wenn ein Merkmal einmal als in dem Meta-Deskriptor vorhanden gewählt worden ist, sagen wir für ein Bild, kann der Meta-Deskriptor gewünschtenfalls das Vorhandensein anderer Merkmale aus einem vordefinierten Satz erlauben. Zum Reduzieren der Anzahl Merkmale, die getestet werden sollen, kann ein Satz mit Assoziationsregeln, hergeleitet aus einem benannten Trainingssatz, verwendet werden, fass erforderlich. So kann beispielsweise ein bestimmter Aufbewahrungsort Multimedia-Information enthalten, die nicht gut beschrieben wird durch den Umriss, so dass das Erfassungssystem den Umriss zu verwenden nicht für notwendig hält.

[0031] Während der Extraktion können, gewünschtenfalls, Merkmale aus dem Multimedia-Inhalt kombiniert werden, und zwar mit semantischer Information höheren Pegels zum Ermitteln der Werte in dem Meta-Deskriptor. Dort, wo beispielsweise der Multimedia-Inhalt ein Bild ist, werden Merkmale aus dem Bild mit semantischer Information aus dem Titel, der mit dem Bild assoziiert ist, kombiniert um die Werte in dem Meta-Deskriptor für das Bild zu bestimmen.

[0032] [Fig. 3](#) ist ein Beispiel einer einfachen Multimedia-Informationstabelle für eine relationale Datei einer Datenbank (jeder andere beliebige Typ einer Datenbank ist auch geeignet), die mehrere Anhängetechniken benutzt. Auf illustrative Weise hat die Tabelle fünf Felder, ein Multimedia-Informationsnummerfeld MM_INFO_NO, ein Deskriptorwertfeld DV, ein Meta-Deskriptorwertfeld MDV, ein Multimediateilequellenfeld MM_SOURCE, und ein Kommentarfeld COMMENT. Das MM_INFO_NO-Feld ist ein primäres Schlüsselfeld. Das DV- und MDV-Feld sind Charakterfelder zum Enthalten von beispielsweise Reihenvektoren. Das MM_SOURCE-Feld ist ein OLE Datentyp, der OLE Objekte, wie digitalisierte Dokumente, Zeichnungen, Bilder, Geräusche usw. koppelt oder einbettet. Das COMMENT-Feld ist ein Memodatentypfeld. Gewünschtenfalls kann in der Tabelle auf das DV-Feld verzichtet werden, vorausgesetzt, dass die Deskriptoren entweder in die Multimedia-Information eingebettet sind oder dass das Erfassungssystem Deskriptoren aus der Multimedia-Information extrahieren kann. Die Tatsache aber, dass Deskriptoren in einer örtlichen Datenbank vorhanden sind, wird es ermöglichen, dass das Erfassungssystem schneller arbeitet.

[0033] Die Aufzeichnung, die durch den primären Schlüssel MM01 identifiziert ist, enthält den Deskriptorwert DV01 und den Meta-Deskriptorwert MDV01 und ist an ein digitalisiertes Bild angehängt, das in FILE01 gespeichert ist, und zwar durch eine Kopplung in dem OLE Datentypfeld. Die durch den primären Schlüssel MM02 identifizierte Aufzeichnung enthält den Meta-Deskriptorwert MDV02, und ist an ein digitalisiertes Standbild angehängt, das durch eine Kopplung in dem OLE Datentypfeld in FILE02 gespeichert ist. Der Deskriptorwert wird aus dem Inhalt von FILE02 extrahiert, und zwar während des Prozesses der Abfrage der Multimedia-Information in dem Aufbewahrungsort. Die durch den primären Schlüssel MM03 identifizierte Aufzeichnung enthält den Deskriptorwert DV03 und den Meta-Deskriptorwert MDV03 für einen Block der Multimedia-Information in einem Bild, und ist an ein digitalisiertes Standbild angehängt, das durch eine Kopplung in dem OLE Datentypfeld in FILE03 gespeichert ist. Die durch den primären Schlüssel MM04 identifizierte Aufzeichnung enthält den Deskriptorwert DVO4 und den Meta-Deskriptorwert MDVO4 für einen anderen Block der Multimedia-Information in demselben Bild, und ist an ein digitalisiertes Standbild angehängt, das in FILE03 gespeichert ist, und zwar durch eine Kopplung in dem OLE Datentypfeld. Die durch den primären Schlüssel MM05 identifizierte Aufzeichnung ist an ein digitalisiertes Standbild angehängt, das in FILE05 gespeichert ist, und zwar durch eine Kopplung in dem OLE Datentypfeld. Das DV-Feld und das MDV-Feld sind für diese Aufzeichnung Null, da der DV und der MDV in die gekoppelte Datei eingebettet sind und auf einfache Art und Weise daraus ausgelesen werden können. Die durch den primären Schlüssel MM06 identifizierte Aufzeichnung ist an ein digitalisiertes Standbild angehängt, das in einer anderen Datenbank gespeichert ist, auf die übers Internet durch eine URL-Kopplung in dem OLE Datentypfeld zugegriffen werden kann. Das DV-Feld und das MDV-Feld sind für diese Aufzeichnung Null, da DV und MDV in die gekoppelte Datei eingebettet sind und daraus ausgelesen wer-

den können. Die durch den primären Schlüssel MM07 identifizierte Aufzeichnung enthält den Meta-Deskriptorwert MDV07 und ist an ein VCR-Band angehängt. Das DV-Feld ist für diese Aufzeichnung Null da der Deskriptorwert in das Vertikal-Austastintervall auf dem VCR-Band eingebettet ist und daraus ausgelesen werden kann. Das MMSOURCE-Feld ist für diese Aufzeichnung Null. Es sei denn, dass das Erfassungssystem aus dem Meta-Deskriptor detektiert, dass es zwischen dem Abfragemultimedia und dem VCR-Band eine deutliche und große Ungleichheit im Inhalt gibt, muss das Band montiert und der Deskriptor während einer Abfrage aus dem VCR-Band ausgelesen werden. Die durch den primären Schlüssel MM08 identifizierte Aufzeichnung enthält den Deskriptorwert DV08 und ist an ein Standbild angehängt, das auf Fotopapier abgedruckt und in dem Schubladen 08 abgelegt ist. Das MM SOURCE-Feld ist für diese Aufzeichnung Null.

[0034] Das Verfahren 130 zur Abfrage eines Multimediaaufbewahrungsortes, wie in [Fig. 1](#) dargestellt, ist durch die Grundschritte 131–137 dargestellt. Der Schritt 131 ist die Bildung einer Abfrage durch den Benutzer unter Anwendung jedes beliebigen bequemen Verfahrens, in dem vorliegenden Fall Abfrage anhand eines Beispiels. In der Abfrage anhand eines Beispiels selektiert der Benutzer ein Item der Multimedia-Information und wünscht sich, alle übereinstimmenden Items der Multimedia-Information aus dem Aufbewahrungsort zu finden. In dem Schritt 132 werden Deskriptoren und Meta-Deskriptoren für ein Item der Multimedia-Information an dem Aufbewahrungsort erfasst, illustrativ aus dem Speicher 120. In dem Fall der Deskriptoren können diese stattdessen in der Multimedia-Information gespeichert sein, und deswegen können sie aus der Multimedia-Information erfasst werden, oder es kann sein, dass sie nicht verfügbar sind und wieder auf Basis der Werte in den Meta-Deskriptoren extrahiert werden müssen. Wenn der Deskriptor für das Multimedia-Informationsitem von einem Typ ist, der vorher noch nicht in der Abfrage 130 (Schritt 133 – YES) verarbeitet worden ist, wird ein entsprechender Deskriptor aus dem Abfrage-Multimedia-Item (Schritt 134) extrahiert, und zwar durch Anwendung des Extraktionsverfahrens und von Gewichtungen, angegeben durch den Meta-Deskriptor für das Aufbewahrungsort-Multimedia-Informationsitem. Danach wird zwischen dem Abfragedeskriptor und dem Deskriptor für das Aufbewahrungsort-Multimedia-Informationsitem ein Vergleich (Schritt 135) angestellt. Merkmale, die keine Gewichtung in dem Meta-Deskriptor für das Aufbewahrungsort-Multimedia-Informationsitem geben, brauchen nicht für Meta-Deskriptorextraktion verarbeitet zu werden. Der Vergleich wird für alle Cluster in der Datenbank (Schritte 136 – NO) wiederholt, und der Satz mit den besten Übereinstimmungen aus jedem Cluster wird entsprechend bewertet, und zwar mit in dem betreffenden technischen Bereich durchaus bekannten geeigneten Mitteln, und für den Benutzer wiedergegeben (Schritt 137).

[0035] Eine Technik zum Optimieren von Meta-Deskriptoren, die sich mit Formalisierung von Benutzereingabe durch menschliche Kenntnisse befasst, ist in [Fig. 4](#) dargestellt. Der Deutlichkeit der Beschreibung halber ist ein einfacher Aufbewahrungsort von Standbildern vorgeschlagen. Derartige Bilder sind typischerweise in nur einigen Kategorien klassifizierbar, beispielsweise menschlichen Figuren, Landschaften, Textilien, und ermöglichen eine inhaltsbasierte Erfassung durch einige Methoden, wie Farbe, Form, Struktur und Umriss. Eine bestimmtes Bild in der Datenbank wird am besten durch ein oder mehrere dieser Merkmale beschrieben und wird nur mäßig durch andere Merkmale beschrieben. So wird eine menschliche Figur am besten durch die Form beschrieben, Pflanzen werden am besten durch Farbe und Struktur beschrieben, und Landschaften werden am besten nur durch Struktur beschrieben. Eine Datenbank mit diesen Charakteristiken ist darauf trainiert, menschliche Eingaben wie folgt zu verwenden. In dem Schritt 301 wird ein Meta-Deskriptorerzeugungsprozess durchgeführt, und zwar unter Verwendung aller Merkmale, illustrativ sind Farbe-, Form-, Struktur-, und Umriss-Schritte 112–114 nach [Fig. 2](#) illustrativ für ein derartiger Prozess. In dem Schritt 302 wird ein Multimedia-Abfrageprozess durchgeführt, und zwar unter Verwendung der Merkmale, illustrativ sind Farbe-, Form-, Struktur- und Umriss-Schritte 131–136 nach [Fig. 2](#) illustrativ für einen derartigen Prozess. In dem Schritt 303 werden die Ergebnisse durch das Erfassungssystem bewertet und für den menschlichen Sachverständigen wiedergegeben. Aus den besten Übereinstimmungen ermittelt der Benutzer, welches Verfahren am besten zu dem Abfragebild passt, oder wenn mehr als ein Verfahren zu dem Abfragebild passt, bestimmt der Benutzer die Gewichtungen für jedes der geeigneten Merkmale, unter Anwendung eines geeigneten Kriteriums. Der Benutzer gibt auch alle anderen Bilder in den Sätzen der besten Übereinstimmungen an, die dieselbe Gewichtung für die geeigneten Merkmale haben sollen. In dem Schritt 306 aktualisiert das Erfassungssystem den Wert der Meta-Deskriptoren, beispielsweise durch Zuordnung neuer Gewichtungen für Merkmale, und zwar auf Basis der Eingabe des menschlichen Sachverständigen. Jede beliebige iterative Lerntechnik kann angewandt werden. Einem Bild, das noch nicht im Training betrachtet worden ist, können für alle Merkmale gleiche Gewichtungen zugeordnet werden. Die Schritte 301, 302, 303 und 306 werden wiederholt, bis der menschliche Sachverständige mit dem Ergebnis zufrieden ist, wobei dann die Meta-Deskriptoren optimiert werden und der Prozess 300 endet.

[0036] Meta-Deskriptoren sind am effektivsten, wenn in ein System von Standards für Deskriptoren, Deskriptorschemen und Deskriptordefinitionssprachen einverlebt, obschon das betreffende Standardisierungssche-

ma nicht kritisch ist. Eine illustrative Darstellung davon, was ein Deskriptor ist, und wie er in einem Multimedia-Informationserfassungssystem funktioniert, ist beschrieben in: "International Organisation for Standardisation ISO/IEC JTC1/SC29/WG11 Coding of Moving Pictures and Audio, MPEG-7 Requirements Dokument V.8, No. N2727", März 1999; und "International Organisation for Standardisation ISO/IEC JTC1/SC29/WG11 Coding of Moving Pictures and Audio, MPEG-7: Context, Objectives and Technical Roadmap, V.11, No. N2729", März 1999. Entsprechend dem in diesem Dokument ausgedrückten Gesichtspunkt erfolgt die Suche nach Multimedia-Information durch einen Vergleich von "Deskriptoren" und deren Instantiierungen ("Deskriptorwerte"), wobei ein Deskriptor eine Darstellung eines "Merkmals" der Multimedia-Information ist und wobei ein Merkmal eine bestimmte Charakteristik der Multimedia-Information ist, die einem etwas kund tut. Ein Deskriptor definiert den Syntax und die Semantik der Merkmalsdarstellung. Gewünschtenfalls können verschiedene Deskriptoren verwendet werden um ein einziges Merkmal darzustellen, wie wenn verschiedene relevante Anforderungen erfüllt werden müssen. So sind beispielsweise mögliche Deskriptoren für das Farbkennzeichen: das Farbhistogramm, der Mittelwert der Frequenzanteile, das Bewegungsfeld, der Text des Titels, usw. Deskriptorwerte werden über den Mechanismus eines "Deskriptorschemas" kombiniert um eine "Deskription" zu bilden. Insbesondere spezifiziert ein Deskriptionsschema (DS) die Struktur und die Semantik der Beziehungen zwischen den Komponenten, was Deskriptoren sowie Deskriptionsschemen sein können, und eine Deskription besteht aus einer DS (Struktur) und dem Satz von Deskriptorwerten (Instatiierungen), welche die Multimedia-Daten beschreiben. Eine Deskriptionsdefinitionssprache ("DDL") ist eine Sprache, welche die Schaffung neuer Deskriptionsschemen ermöglicht und möglicherweise auch die Schaffung von Deskriptoren. Sie ermöglicht auch die Erweiterung und die Modifikation bestehender Beschreibungsschemen. Die Tabelle 1, die dem oben genannten MPEG-7 Anforderungen-Dokument V.8 entnommen ist (zwar geändert um ein Umrissmerkmal einzuschließen), erläutert den Unterschied zwischen einem Merkmal und den Deskriptoren.

TABELLE 1

Merkmal		Deskriptor		
Merkmarkategorie	Typ	Merkmal	Merkmal Etikett	Datentyp
Räumlich		Struktur eines Objektes Form eines Objektes		Ein Satz von Wellenkoeffizienten Ein Satz von Polygoneckpunkten
Zeitlich		Trajekt von Objekten		Kettencode
Objektiv		Farbe eines Objekts		Farbhistogramm rgb Vektor
		Form eines Objektes		Ein Satz von Polygoneckpunkten Ein Satz von Momenten
		Struktur eines Objektes		Ein Satz von Wellenkoeffizienten Der Satz von Kontrast, Körnigkeit und Direktionalität
		Umriss eines Objektes		Satz von Rändern
		Audiofrequenzinhalt		Mittelwert der Frequenzanteile
Subjektiv		Emotion (Glück, Angst, Bosheit)		Ein Satz von eigenface Parameter, Text
		Stil		Text
		Kommentar		Text
Produktion		Autor		Text
		Produzent		Text
		Regisseur		Text, usw.
Komposition Information		Szenenkomposition		Baumgraphik
Konzepte		Ereignis		Text
		Aktivität		Text, ein numerischer Wert

[0037] Allgemeine Anforderungen für Deskriptoren und Deskriptionsschemen, wie in dem oben genannten MPEG-7 Anforderungendokument genannt, werden von geeigneten Deskriptoren und deren Meta-Deskriptoren unterstützt. Mehrere Typen von Merkmalen – geeignete Deskriptoren und deren Meta-Deskriptoren unterstützen Multimediaskriptionen unter Anwendung mehrere Typen von Merkmalen wie: N-bemessene räumlich-zeitliche Struktur (beispielsweise die Dauer eines Musikfragmentes), objektive Merkmale (beispielsweise die Anzahl Bette in einem Hotel, Farbe eines Objektes, Form eines Objektes, Audiotonhöhe, usw.) subjektive Merkmale (beispielsweise wie schön, glücklich oder dick jemand ist, Thema, Stil, usw.), Produktionsmerkmale (beispielsweise Information über Dokumenterschaffung, wie Antragsdatum, Produzent, Regisseur, Darsteller, Rollen, Produktionsgesellschaft, Produktionsgeschichte, jede mögliche Nicht-IPMP Produktionsinformation), Kompositionsinformation (beispielsweise wie eine Szene zusammengestellt ist. Editierinformation, die Vorlieben des Benutzers, usw.) und Konzepte (beispielsweise Ereignis, Aktivität). Abstraktionspegel für das Multimediamaterial – Es werden hierarchische Mechanismen zum Beschreiben von Multimediodokumenten auf verschiedenen Abstraktionspegeln unterstützt, worin die Wünsche des Benutzers nach Information auf verschiedenen Abstraktionspegeln untergebracht sind, wie beispielsweise die Zusammensetzung von Objekten aus Subobjekten, eine sequenzweise Analyse von Bewegung in einem Video, und die Handlungsstruktur eines Videos. Kreuzmodalität – Audio, visuell oder andere Deskriptoren und deren Meta-Deskriptoren, die Abfragen auf Basis visueller Deskriptionen ermöglichen um Audiodaten zu erfassen und umgekehrt, werden unterstützt, wobei beispielsweise die Abfrage ein Ausschnitt der Stimme von Pavarotti ist und das Ergebnis eine Erfassung von Videoclips, in denen Pavarotti singt und wobei Pavarotti anwesend ist). Mehrere Deskriptionen – die Fähigkeit, mehrere Deskriptionen desselben Materials auf verschiedenen Stufen des Produktionsprozesses zu verarbeiten wird unterstützt, ebenso wie Deskriptionen, die für mehrere Kopien desselben Materials gelten. Deskriptionsschemenbeziehungen – geeignete Deskriptionsschemen drücken die Beziehungen zwischen Deskriptoren und deren Meta-Deskriptoren aus, um deren Verwendung in mehr als nur einem Deskriptionsschema zu verwenden. Die Fähigkeit gleichwertige Beziehungen zwischen Deskriptoren und deren Meta-Deskriptoren in verschiedenen Deskriptionsschemen zu kodieren, wird unterstützt. Deskriptorprioritäten – die Priorisierung von Deskriptoren und deren Meta-Deskriptoren wird durch die Deskriptionsschemen derart unterstützt, dass Abfragen effizienter verarbeitet werden können. Die Prioritäten können Vertrauenspegel oder Zuverlässigkeitsspegel reflektieren. Deskriptorhierarchie – geeignete Deskriptionsschemen unterstützen die hierarchische Darstellung verschiedener Deskriptoren, damit Abfragen in aufeinander folgenden Pegeln effizienter verarbeitet werden können, wobei N Pegeldeskriptoren (N-i) Pegeldeskriptoren ergänzen. Deskriptorskaliertbarkeit – geeignete Deskriptionsschemen unterstützen skalierbare Deskriptoren mit ihren Meta-Deskriptoren, damit Abfragen in aufeinander folgenden Deskriptionsschichten effizienter verarbeitet werden können. Deskription von zeitlichem Bereich – Assoziation von Deskriptoren mit ihren Meta-Deskriptoren zu verschiedenen zeitlichen Bereichen werden unterstützt, und zwar hierarchisch (Deskriptoren mit ihren Meta-Deskriptoren werden mit allen Daten oder mit einem zeitlichen Subsatz davon assoziiert) sowie sequentiell (Deskriptoren mit ihren Meta-Deskriptoren werden nacheinander mit aufeinander folgenden Zeitperioden assoziiert). Direkte Datenmanipulation – Deskriptoren und ihre Meta-Deskriptoren, die als Griffe arbeiten, die sich unmittelbar auf die Daten beziehen, werden unterstützt, damit eine Manipulation des Multimediamaterials ermöglicht wird. Sprache von textbasierten Deskriptoren – geeignete Deskriptoren mit ihren Meta-Deskriptoren spezifizieren die in der Deskription verwendeten Sprache und unterstützen alle natürlichen Sprachen. Übersetzungen in Textdeskriptionen – geeignete Textdeskriptionen schaffen eine Möglichkeit, Übersetzung in eine Anzahl verschiedene Sprachen zu enthalten, um die Beziehung zwischen der Deskription in den verschiedenen Sprachen zu transportieren.

[0038] Funktionelle Anforderungen für Deskriptoren und Deskriptionsschemen, wie in dem oben genannten MPEG-7 Anforderungendokument vorgeschlagen, werden durch geeignete Deskriptoren und ihre Meta-Deskriptoren unterstützt. Abfrageeffektivität – die effektive Abfrage von Multimediamaterial wird unterstützt. Abfrageeffizienz – die effiziente Abfrage von Multimediamaterial wird unterstützt. Abfrage auf Basis der Ähnlichkeit – Deskriptionen, die es ermöglichen, Datenbankinhalt nach Ähnlichkeitsgrad mit der Abfrage zu bewerten, werden unterstützt. Assozierte Information – die Assoziation anderer Information mit den Daten wird unterstützt. Gestreamte und gespeicherte Deskriptionen – getreamte (synchronisiert mit dem Inhalt) und nicht gestreamte Datenbeschreibungen werden unterstützt. Verteilte Multimediodatenbanken – die simultane und transparente Abfrage von Multimedia-Daten in verteilten Datenbanken wird unterstützt. Bezugnahme auf analoge Daten – die Fähigkeit, auf Multimediodokumente in analogem Format Bezug zu nehmen und diese zu beschreiben, wird unterstützt (beispielsweise das Schaffen zeitlicher Bezugnahmen auf Sequenzen in einem VHS Band). Interaktive Abfragen – Mechanismen um interaktive Abfragen zu ermöglichen, werden unterstützt. Kopplung – Mechanismen, die es ermöglichen, dass Quellendaten in Raum und in Zeit geortet werden, werden unterstützt, einschließlich Kopplungen mit relativer Information. Priorisierung relativer Information – Mechanismen, welche die Priorisierung relativer Information ermöglichen, oben stehend erwähnt bei Kopplung, werden unterstützt. Browsing – Deskriptionen, die es ermöglichen, Informationsinhalt vorher zu sehen, um Benutzern behilflich zu sein beim Überwinden ihrer Unvertrautheit mit der Struktur und/oder den Typen der Information, oder

zum Verdeutlichen ihrer unentschiedenen Bedürfnisse, werden unterstützt. Assoziierte Beziehungen – Beziehungen zwischen Komponenten einer Deskription werden unterstützt. Interaktivitätsunterstützung – Mittel, die eine Spezifikation der Interaktivität in Bezug auf eine Deskription ermöglichen, werden unterstützt (beispielsweise Tele-Wahl in Bezug auf ausgestrahlte Ereignisse). Information über geistiges Eigentum – das Einschließen von Copyright Lizenzierungs- und Beurkundungsinformation in Bezug auf DS, DSs und Deskriptionen, wird unterstützt.

[0039] Visuelle spezifische Anforderungen für Deskriptoren und Deskriptionsschemen, wie in dem oben genannten MPEG-7 Anforderungendokument vorgeschlagen, werden durch geeignete Deskriptoren und deren Meta-Deskriptoren unterstützt. Merkmaltyp – visuelle Deskriptionen ermöglichen die nachfolgenden Merkmale (vorwiegend in Bezug auf den Informationstyp, verwendet in den Abfragen) werden unterstützt: Farbe, visuelle Objekte, Struktur, Umriss, Form, Standbilder und Bewegtbilder (beispielsweise Daumennägel), Lautstärke, räumliche Beziehungen, Bewegung, Verformung, Quelle visuellen Objektes und der Charakteristiken (beispielsweise das Quellenobjekt, das Quellenereignis, die Quellenattribute, Ereignisse, Ereignisattribute und typische assoziierte Szenen und Modelle (beispielsweise MPEG-4 SNHC)). Datenvisualisierung unter Verwendung der Deskription – ein Gebiet von Multimedia-Datenbeschreibungen mit zunehmenden Fähigkeiten in Termen von Visualisierung wird unterstützt (ermöglicht eine mehr oder weniger skizzenmäßige Visualisierung der indizierten Daten). Visuelle Datenformate – Beschreibung der nachfolgenden visuellen Datenformate wird unterstützt: Digitalvideo und Film, wie MPEG-1, MPEG-2 oder MPEG-4; Analogvideo und Film; Standbilder in Elektronik, wie JPEG, Papier oder anderen Formaten; Graphik, wie CAD; dreidimensionale Modelle, insbesondere VRML; und Kompositionssdaten, assoziiert mit Video. Beschreibung anderer visueller Datenformate, die noch definiert werden müssen, ist möglich. Visuelle Datenklassen – Beschreibungen, insbesondere aus die nachfolgenden Klassen visueller Daten anwendbar, werden unterstützt: natürliches Video, Standbilder, Graphik, Animation (zweidimensional), dreidimensionale Modelle, und Kompositionssinformation.

[0040] Audiospezifische Anforderungen für Deskriptoren und Deskriptionsschemen, wie in em oben genannten MPEG-7 Anforderungendokument erwähnt, werden durch geeignete Deskriptoren und deren Meta-Deskriptoren unterstützt. Merkmaltyp – Audiodeskriptionen, welche die nachfolgenden Merkmale ermöglichen (vorwiegend bezogen auf den in den Abfragen verwendeten Informationstyp): Frequenzumriss (allgemeiner Trend, melodischer Umriss), Audioobjekte, Timbre, Harmonie, Frequenzprofil, Amplitudenumhüllende, Zeitstruktur (einschließlich Rhythmus), Textinhalt (typischerweise Sprache oder Lyrik) akustische Annäherungen (Vokalisierung eines akustischen Schaubildes durch beispielsweise Summen einer Melodie oder Brummen eines Schalleffektes), prototypische Ton (typischerweise Abfrage durch Beispiel), Raumstruktur (anwendbar auf Mehrkanalquellen, Stereo, 5.1-Kanal und binaurale Geräusche mit je einer eigenen Mapping), Schallquelle und deren Charakteristiken (beispielsweise das Quellenobjekt, das Quellenereignis, Quellenattribute, Ereignisse, Ereignisattribute, und typische assoziierte Szenen), und Modelle (beispielsweise MPEG-4 SAOL). Datenvertonung unter Verwendung der Deskription – ein Gebiet von Multimedia-Datendeskriptionen mit zunehmenden Fähigkeiten in Termen von Vertonung wird unterstützt. Auditive Datenformate – die Deskription der nachfolgenden Typen auditiver Daten werden unterstützt: digitales Audio (beispielsweise MPEG-1 Audio, CD), analoges Audio (beispielsweise Vinyl Schallplatten, Magnetbandmedia), MIDI einschließlich General MIDI und Karaokeformate, modellbasiertes Audio (beispielsweise MPEG-4 strukturierte Audio Orchestersprache – SAOL), und Produktionsdaten. Auditiven Datenklassen – Beschreibungen, typischerweise auf die nachfolgenden Subklassen auditiver Daten anwendbar, werden unterstützt (natürliche Audioszene), Musik, atomare Schalleffekte (beispielsweise Krachen), Sprache, symbolische Audiodarstellungen (MIDI, SNHC Audio), und Mischnformation (einschließlich Effekte).

[0041] Codierungsanforderungen für Deskriptoren und Deskriptionsschemen, wie in dem oben genannten MPEG-7 Anforderungendokument erwähnt, werden durch geeignete Deskriptoren und deren Meta-Deskriptoren unterstützt. Deskriptionseffiziente Darstellung – die effiziente Darstellung von Datenbeschreibungen wird unterstützt. Deskriptionsextraktion – die Verwendung von Deskriptoren und Deskriptionsschemen, die auf einfache Weise aus nicht komprimierten und komprimierten Daten, extrahiert werden können, entsprechen verschiedenen weit verbreiteten Formaten, wird von den Meta-Deskriptoren unterstützt. Robustheit gegenüber Informationsfehler und Informationsverlust – Mechanismen, die ansprechendes Verhalten des Systems im Falle von Übertragungsfehlern gewährleisten, werden unterstützt.

[0042] Während textspezifische Anforderungen für Deskriptoren und Deskriptionsschemen nicht in dem oben genannten MPEG-7 Anforderungendokument vorgeschlagen sind, unterstützen geeignete Deskriptoren und ihre Meta-Deskriptoren die Fähigkeit, dass Multimedia-Inhalt nebst audiovisueller Information Text enthält oder darauf hinweist, unter der Voraussetzung, dass die Textbeschreibungen und die Schnittstelle Abfragen auf Basis audiovisueller Beschreibungen ermöglichen um Textdaten zu erfassen und umgekehrt, und dass die Be-

schreibungen von Nur-Text-Dokumenten und zusammengesetzten Dokumenten mit Text die gleichen sind.

[0043] Während in einigen Situationen, in denen Meta-Deskriptoren verwendet werden, die Suchmaschine oder der Filteragent (Benutzerseite) den genauen Merkmalextraktionsalgorithmus kennen soll, der von dem Metadeskriptionserzeugungsprozess angewandt wird, da der betreffende zur Merkmalextraktion während des Deskriptionserzeugungsprozesses verwendete Algorithmus sonst für den Metadeskriptionserzeugungsprozess nicht relevant ist. Folglich ist der Metadeskriptionsprozess imstande, technologische Entwicklungen in der Merkmalextraktion und Bewerbungsentwicklung entgegen zu kommen.

[0044] Die Beschreibung der vorliegenden Erfindung und deren Anwendung, wie hier beschrieben, ist illustriativ und soll den Rahmen der vorliegenden Erfindung, wie in den nachfolgenden Patentansprüchen beschrieben, nicht beschränken. Variationen und Modifikationen der hier beschriebenen Ausführungsformen sind möglich, und praktische Alternativen der betreffenden Elemente der Ausführungsformen sind dem Fachmann durchaus bekannt.

Patentansprüche

1. Verfahren zum Darstellen eines Multimedia-Informationsitems, das zu einer bestimmten Kategorie von Multimediainhalt (10, 12, 14, 16, 18, 20, 22, 24) gehört, wobei dieses Verfahren die nachfolgenden Verfahrensschritte umfasst:

- das Erwerben (111) einer Anzahl Deskriptoren, die eine Darstellung von Merkmalen des Multimedia-Informationsitems sind; wobei die Deskriptoren von einem vordefinierten Satz verschiedener Deskriptortypen sind; wobei das Verfahren durch die nachfolgenden Verfahrensschritte gekennzeichnet wird:
- das Ermitteln der Relevanz jedes der Deskriptoren zum Erfassen des Multimedia-Informationsitems; und
- das Erzeugen (113) wenigstens eines Meta-Deskriptors für die vielen Deskriptoren, wobei der Meta-Deskriptor eine Gruppe von Datenelementen aufweist, wobei jedes der Datenelemente eine relative Gewichtung eines betreffenden Deskriptors angibt, und zwar entsprechend der bestimmten Relevanz des genannten Deskriptors; und
- das Anhängen (114) des Meta-Deskriptors an das Multimedia-Informationsitem.

2. Verfahren nach Anspruch 1, wobei die Multimedia-Information kollektiv von mehreren Inhaltstypen ist, wobei das Verfahren weiterhin Folgendes umfasst: das Erzeugen (112) von Cluster der Deskriptoren, wobei:

- der Meta-Deskriptorerzeugungsschritt das Erzeugen (113) von Meta-Deskriptoren für die Cluster umfasst; und
- der Meta-Deskriptoranhängeschritt das Anhängen (114) des Meta-Deskriptoren für die Cluster an Multimedia-Informationsitems aufweist, wie von den Deskriptoren in den Cluster beschrieben.

3. Verfahren nach Anspruch 2, wobei der Meta-Deskriptorerzeugungsschritt das Erzeugen betreffender Gruppen von Datenelementen für jedes der Cluster umfasst, welche die Relevanz der Deskriptoren darin angeben.

4. Verfahren nach Anspruch 3, wobei:

- wenigstens einige Deskriptoren Darstellungen von Merkmalen eines Multimedia-Informationsitems sind, das zu einer Kategorie von Multimediainhalt gehört, wobei die Merkmale einen geordneten Satz von Merkmalen aufweisen, die Farbe, Struktur, Form und Umriss aufweisen, und die Kategorie des Multimediainhalts ein Standbild ist, und
- der Meta-Deskriptorerzeugungsschritt wenigstens Folgendes umfasst:
- das Erzeugen betreffender binärer Vektoren für jedes der Cluster, welche die Relevanz der Deskriptoren darin angeben,
- das Erzeugen betreffender Gruppen von Gewichtungswerten für jedes der Cluster, welche die betreffenden Gewichtungen für die Deskriptoren darin angeben, und
- das Erzeugen betreffender Zeichenreihen für jedes der Cluster, die wenigstens ein relevantes Merkmal mit einem vorbestimmten Satz von Darstellungstypen identifizieren, und wenigstens eines der Darstellungstypen aus dem vorbestimmten Satz mit Darstellungstypen identifizieren.

5. Verfahren nach Anspruch 2, das weiterhin das Indexieren der Deskriptoren zu der Multimedia-Information umfasst.

6. Verfahren nach Anspruch 1, wobei:

- wenigstens einige Deskriptoren Darstellung von Merkmalen eines Multimedia-Informationsitems sind, das zu

einer Kategorie von Multimedainhalt gehört, wobei die Merkmale einen geordneten Satz von Merkmalen aufweisen, die Farbe, Struktur, Form und Umriss umfassen, und die Kategorie von Multimedia-Inhalt ein Standbild ist; und

- der Meta-Deskriptorerzeugungsschritt wenigstens Folgendes umfasst:
- das Erzeugen binärer Vektoren, welche die Relevanz der Deskriptoren darin angeben;
- das Erzeugen einer Gruppe von Gewichtungswerten, welche die Gewichtungen für die deskriptoren darin angeben, und
- das Erzeugen einer Zeichenreihe, die wenigstens ein relevantes Merkmal, das einen vorbestimmten Satz von Darstellungstypen hat, und wenigstens einer der Darstellungstypen aus dem vorbestimmten Satz von Darstellungstypen, identifiziert.

7. Verfahren zum Suchen eines Multimedia-Informationsitems (**10, 12, 14, 16, 18, 20, 22, 24**) an einem Aufbewahrungsort unter Verwendung eines Abfragemultimedia-Informationsitems; wobei Multimedia-Information in dem Aufbewahrungsort durch eine Anzahl verschiedener Typen Aufbewahrungsortdeskriptoren beschrieben wird; wobei die Deskriptoren von Merkmalen der Multimedia-Information von einem vorbestimmten Satz verschiedener Deskriptortypen sind, wobei das Verfahren Folgendes umfasst:

- das Erwerben (**132**) von Meta-Deskriptoren der Aufbewahrungsortdeskriptoren, wobei ein Meta-Deskriptor eine Gruppe von Datenelementen umfasst, wobei jedes der Datenelemente eine relative Gewichtung des betreffenden Aufbewahrungsortdeskriptors entsprechend einer vorbestimmten Relevanz des genannten Deskriptors angibt zur Erfassung des Multimedia-Informationsitems,
- das Selektieren (**131**) der Abfragemultimedainformation,
- das Extrahieren (**134**) wenigstens eines Abfragedeskriptors aus der Abfragemultimedia-Information auf Basis der Meta-Deskriptoren zum Erhalten wenigstens eines Abfragedeskriptors;
- das Vergleichen (**135**) des Abfragedeskriptors mit den Aufbewahrungsortdeskriptoren; und
- das Eingliedern (**137**) wenigstens einiger Items der Multimedia-Information in den Aufbewahrungsort entsprechend dem Vergleichsschritt.

8. Verfahren nach Anspruch 7, wobei die Aufbewahrungsortdeskriptoren Instanzen einer Anzahl verschiedener Deskriptortypen sind, und wobei:

- der Schritt der Extraktion wenigstens eines Deskriptors aus der Abfragemultimedia-Information das Extrahieren (**133, 134**) einer Anzahl Deskriptoren umfasst, und zwar auf Basis der Meta-Deskriptoren zum Erhalten einer Anzahl Abfragedeskriptoren, wobei die Abfragedeskriptoren betreffende Instanzen der Deskriptortypen sind, und
- der Vergleichsschritt das Vergleichen (**135, 136**) der Abfragedeskriptoren mit betreffenden Aufbewahrungsortdeskriptoren mit denselben Deskriptortypen umfasst.

9. Verfahren nach Anspruch 7, wobei der Schritt der Erwerbung der Meta-Deskriptoren Folgendes umfasst:

- das Extrahieren (**111**) von Aufbewahrungsortdeskriptoren aus der Multimedia-Information in dem Aufbewahrungsort,
- das Erzeugen (**112**) von Cluster der Aufbewahrungsortdeskriptoren,
- das Erzeugen (**113**) von Meta-Deskriptoren für die Cluster, wobei ein Meta-Deskriptor eine Gruppe von Datenelementen enthält, welche die Relevanz der Aufbewahrungsortdeskriptoren darin angeben, und
- das Anhängen (**114**) der Meta-Deskriptoren für die Cluster an die betreffende Multimedia-Information in den Cluster.

10. Verfahren nach Anspruch 9, wobei die Aufbewahrungsortdeskriptoren Instanzen einer Anzahl verschiedener Deskriptortypen sind und wobei:

- der Schritt der Extraktion wenigstens eines Deskriptors aus der Abfragemultimedia-Information das Extrahieren (**133, 134**) einer Anzahl Deskriptoren auf Basis der Meta-Deskriptoren umfasst zum Erhalten einer Anzahl Abfragedeskriptoren, wobei die Abfragedeskriptoren betreffende Instanzen der Deskriptortypen sind, und
- der Vergleichsschritt das Vergleichen (**135, 136**) der Abfragedeskriptoren mit betreffenden Aufbewahrungsortdeskriptoren mit denselben Deskriptortypen umfasst.

11. Datenstruktur zur Darstellung eines Multimedia-Informationsitems, das zu einer bestimmten Kategorie von Multimedainhalt (**10, 12, 14, 16, 18, 20, 22, 24**) gehört, wobei die genannte Multimedia-Information durch eine Anzahl Deskriptoren beschrieben wird, die Darstellungen von Merkmalen der Multimedia-Information sind; wobei die Deskriptoren aus einem vordefinierten Satz verschiedener Deskriptortypen sind; wobei die genannte Datenstruktur dadurch gekennzeichnet ist, dass sie eine Gruppe von Datenelementen aufweist, wobei jedes der Datenelemente eine relative Gewichtung eines betreffenden Deskriptors entsprechend einer bestimmten Relevanz des genannten Deskriptors zur Erfassung des Multimedia-Informationsitems angibt.

12. Datenstruktur nach Anspruch 11, wobei:

- die Kategorie des Multimediainhalts ein Standbild ist (**12**);
- die Merkmale eines geordneten Satzes von Merkmalen Farbe, Struktur, Form und Umriss umfasst; und
- die Datenelemente Bits eines binären Vektors sind.

13. Datenstruktur nach Anspruch 11, wobei:

- die Kategorie des Multimediainhalts ein Standbild ist (**12**);
- die Merkmale ein geordneter Satz von Merkmalen Farbe, Struktur, Form und Umriss umfasst; und
- die Datenelemente Zeichen einer Reihe sind.

14. Datenstruktur nach Anspruch 13, wobei die Reihe Folgendes umfasst:

- ein erstes Zeichen, das ein relevantes Merkmal identifiziert, wobei das relevante Merkmal einen vorbestimmten Satz von Darstellungstypen aufweist;
- ein zweites Zeichen, das angibt, wie viele der Darstellungstypen aus dem vorbestimmten Satz von Darstellungstypen verwendet werden zum Erhalten des relevanten Merkmals; und
- eine Anzahl dritter Zeichen, die bestimmte Zeichen der Darstellungstypen aus dem vorbestimmten Satz von Darstellungstypen identifizieren.

Es folgen 2 Blatt Zeichnungen

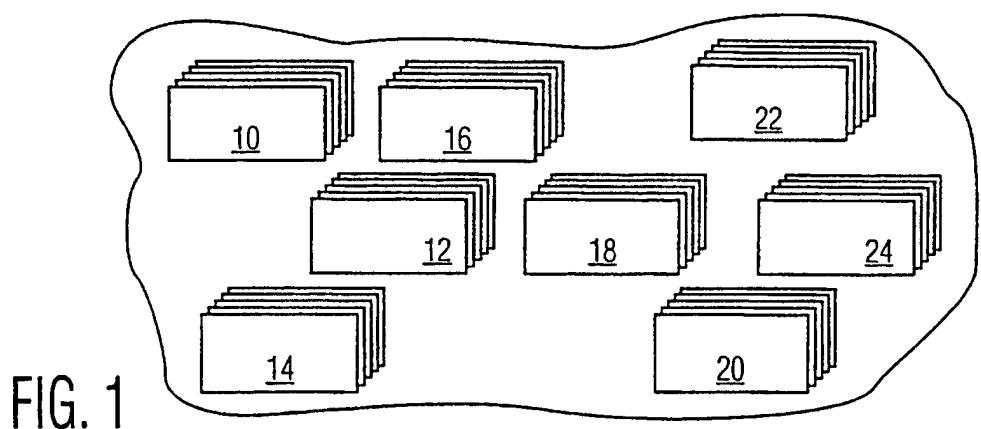
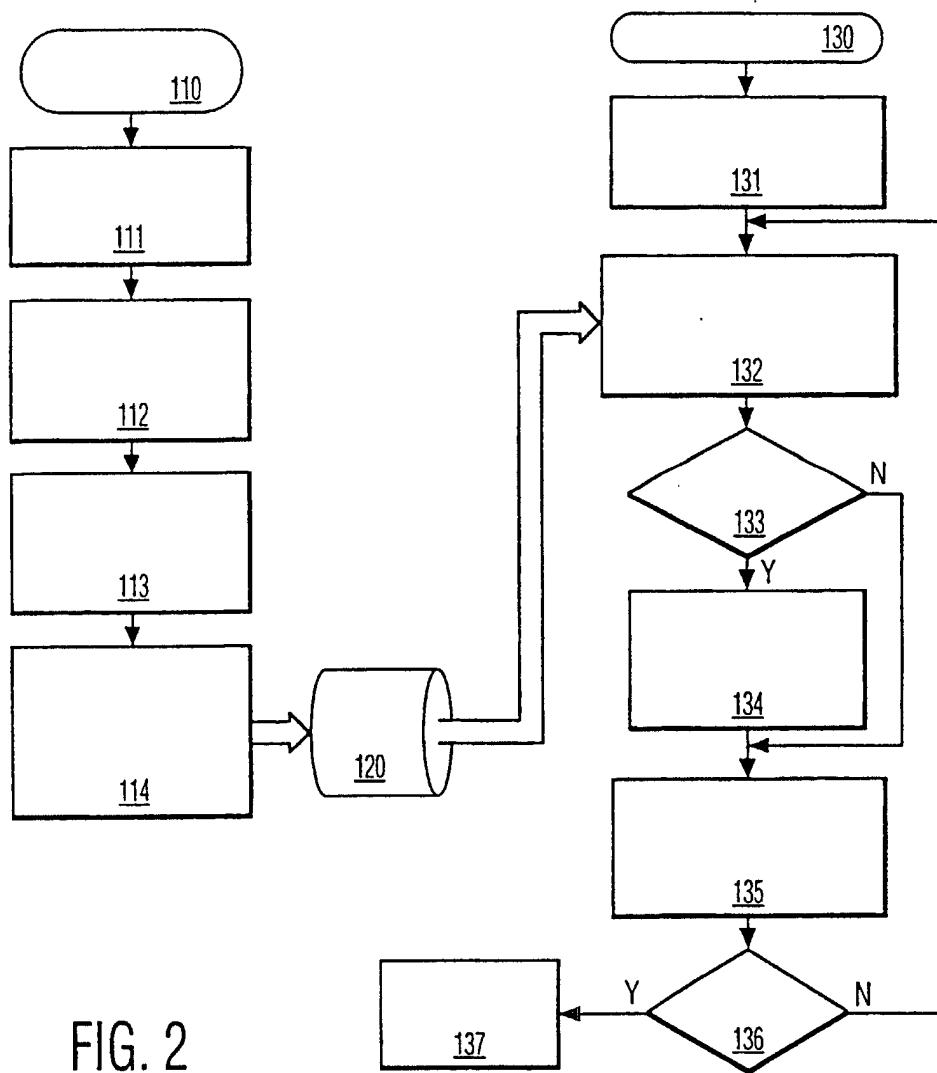


FIG. 1



MM_INFO_NO	DV	MDV		
MM01	DV01	MDV01		
MM02		MDV02		
MM03	DV03	MDV03		
MM04	DV04	MDV04		
MM05				
MM06				
MM07		MDV07		
MM08	DV08	MDV08		

FIG. 3

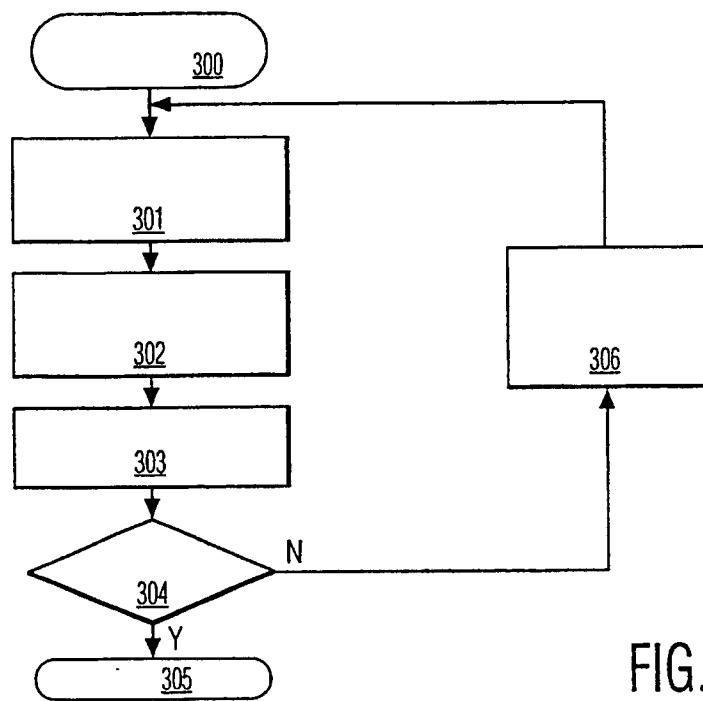


FIG. 4