

(19) 日本国特許庁(JP)

(12) 公開特許公報(A)

(11) 特許出願公開番号

特開2012-53855

(P2012-53855A)

(43) 公開日 平成24年3月15日(2012.3.15)

(51) Int.Cl.	F I	テーマコード (参考)
<b>G06F 17/30 (2006.01)</b>	G06F 17/30 380E	5B075
	G06F 17/30 210D	
	G06F 17/30 210A	

審査請求 未請求 請求項の数 7 O L (全 18 頁)

(21) 出願番号 特願2010-198132 (P2010-198132)  
 (22) 出願日 平成22年9月3日 (2010.9.3)

(71) 出願人 000006747  
 株式会社リコー  
 東京都大田区中馬込1丁目3番6号  
 (74) 代理人 100070150  
 弁理士 伊東 忠彦  
 (72) 発明者 新西 誠人  
 東京都大田区中馬込1丁目3番6号 株式  
 会社リコー内  
 Fターム(参考) 5B075 ND16 NK31 NR12

(54) 【発明の名称】 コンテンツ閲覧装置、コンテンツ表示方法及びコンテンツ表示プログラム

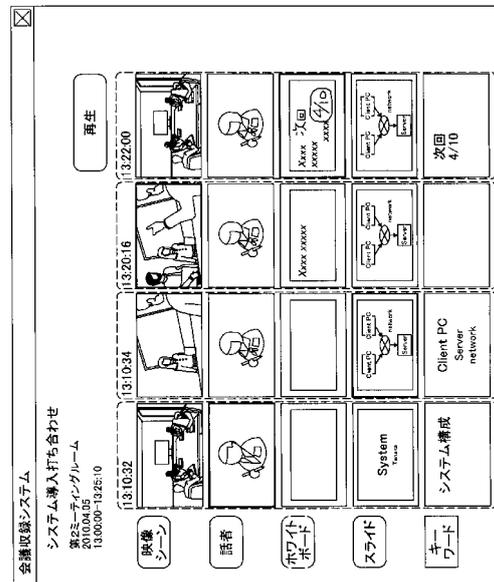
(57) 【要約】

【課題】映像コンテンツなど時間情報のある情報の振り返りにおいて、重要なコンテキストとその要点内容を容易に把握可能なコンテンツ閲覧装置等を提供する。

【解決手段】本発明におけるコンテンツ閲覧装置は、映像コンテンツ及び時間軸上でリンク付けされた音声、静止画像を有し、映像コンテンツからの映像シーンが格納され、また音声、静止画像から抽出されたキーワードと当該キーワード毎に解析された重要度とが対応し格納され、重要度が上位所定数のキーワードを読み出すとともに、読み出されたキーワードの抽出元となった音声又は静止画像と、同一の時間軸上でリンク付けされた映像シーンとを表示データとして読み出し、読み出された表示データのうち同一の時間軸上でリンク付けされた映像シーン、音声及び/又は静止画像、並びにキーワードのグループを表示するとともにグループが複数ある場合グループ毎に時系列に並べ表示させる。

【選択図】 図2

本実施形態に係る表示画面例を示す図



**【特許請求の範囲】****【請求項 1】**

コンテンツを表示手段に表示するコンテンツ閲覧装置であって、

映像コンテンツ及び前記映像コンテンツと時間軸上でリンク付けされた音声、静止画像のうち少なくともいずれかを有するデータ格納手段と、

前記映像コンテンツから抽出された映像シーンが格納され、また前記音声、前記静止画像のうち少なくともいずれかから抽出されたキーワードと当該キーワード毎に解析された重要度とが対応付けて格納された解析データ格納手段と、

前記解析データ格納手段から前記重要度が上位所定数のキーワードを読み出すとともに、読み出された前記キーワードの抽出元となった前記音声又は静止画像と、当該音声又は静止画像と同一の時間軸上でリンク付けされた映像シーンを表示データとして読み出す読出制御手段と、

前記読出制御手段により読み出された前記表示データのうち、同一の時間軸上でリンク付けされた映像シーン、音声及び / 又は静止画像、並びにキーワードのグループを表示するとともに、前記グループが複数ある場合にはグループ毎に時系列に並べ、前記表示手段に表示させる表示制御手段と、

を有することを特徴とするコンテンツ閲覧装置。

**【請求項 2】**

前記映像コンテンツから抽出された映像シーンと当該映像シーン毎に解析された重要度とが対応付けて格納された第 2 解析データ格納手段と、

前記第 2 解析データ格納手段から前記重要度が上位所定数の映像シーンを読み出すとともに、読み出された前記映像シーンと同一の時間軸上でリンク付けされた音声及び静止画像とを第 2 表示データとして読み出す第 2 読出制御手段と、

前記第 2 読出制御手段により読み出された前記表示データのうち、同一の時間軸上でリンク付けされた映像シーン、並びに音声及び / 又は静止画像のグループを表示するとともに、前記グループが複数ある場合にはグループ毎に時系列に並べ、前記表示手段に表示させる第 2 表示制御手段と、

を有することを特徴とする請求項 1 記載のコンテンツ閲覧装置。

**【請求項 3】**

前記表示制御手段は、前記抽出元となった音声又は静止画像は強調して表示し、

前記第 2 表示制御手段は、前記重要度が上位所定数の映像シーンは強調して表示すること、

を特徴とする請求項 2 記載のコンテンツ閲覧装置。

**【請求項 4】**

コンテンツを表示手段に表示するコンテンツ閲覧装置におけるコンテンツ表示方法であって、

前記コンテンツ閲覧装置は、

映像コンテンツ及び前記映像コンテンツと時間軸上でリンク付けされた音声、静止画像のうち少なくともいずれかを有するデータ格納手段と、

前記映像コンテンツから抽出された映像シーンが格納され、また前記音声、前記静止画像のうち少なくともいずれかから抽出されたキーワードと当該キーワード毎に解析された重要度とが対応付けて格納された解析データ格納手段と、

を備え、

前記解析データ格納手段から前記重要度が上位所定数のキーワードを読み出すとともに、読み出された前記キーワードの抽出元となった前記音声又は静止画像と、当該音声又は静止画像と同一の時間軸上でリンク付けされた映像シーンを表示データとして読み出す読出制御手段と、

前記読出制御手段により読み出された前記表示データのうち、同一の時間軸上でリンク付けされた映像シーン、音声及び / 又は静止画像、並びにキーワードのグループを表示するとともに、前記グループが複数ある場合にはグループ毎に時系列に並べ、前記表示手段

10

20

30

40

50

に表示させる表示制御手順と、  
を有することを特徴とするコンテンツ表示方法。

【請求項 5】

前記コンテンツ閲覧装置は、  
前記映像コンテンツから抽出された映像シーンと当該映像シーン毎に解析された重要度とが対応付けて格納された第 2 解析データ格納手段と、  
を備え、

前記第 2 解析データ格納手段から前記重要度が上位所定数の映像シーンを読み出すとともに、読み出された前記映像シーンと同一の時間軸上でリンク付けされた音声及び静止画像とを第 2 表示データとして読み出す第 2 読出制御手順と、

前記第 2 読出制御手順により読み出された前記表示データのうち、同一の時間軸上でリンク付けされた映像シーン、並びに音声及び / 又は静止画像のグループを表示するとともに、前記グループが複数ある場合にはグループ毎に時系列に並べ、前記表示手段に表示させる第 2 表示制御手順と、

を有することを特徴とする請求項 4 記載のコンテンツ表示方法。

【請求項 6】

前記表示制御手順は、前記抽出元となった音声又は静止画像は強調して表示し、

前記第 2 表示制御手順は、前記重要度が上位所定数の映像シーンは強調して表示すること、

を特徴とする請求項 5 記載のコンテンツ表示方法。

【請求項 7】

請求項 4 ないし 6 何れか一項記載のコンテンツ表示方法をコンピュータに実行させるためのコンテンツ表示プログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、コンテンツ閲覧装置、コンテンツ表示方法及びコンテンツ表示プログラムに関する。

【背景技術】

【0002】

従来、会議（ミーティング）の議事録といえば、例えば文書議事録による記録方式が主流であった。この方式では書記が会議に同席し議事録を作成していく。書記は会議中にあった発言の内容を逐一記録した議事録を作成する場合もあるが（特に重要な会議等）、大抵の場合、書記は会議の内容を理解しつつ要点を選択しながら要約された議事録を作成していく。後日、会議の内容を振り返る場合、要約された文書議事録は重要な事項がまとめられているため、短時間で容易に会議の内容を把握できる。

【0003】

近年デジタル技術の発達に伴い、会議（ミーティング）の内容は、容易に映像コンテンツとして記録できるようになった。会議の内容をビデオカメラに収録しておけば、後日会議の模様をそのまま再生できるが、ここで、ビデオカメラに収録された映像コンテンツはデータとして時間とともに流れる連続的な情報である。従って上述の要約された文書議事録とは違い、会議の内容（要約）を短時間で効率よく振り返りたい場合、早送りや巻戻しを繰り返しながら再生する必要があるので重要な要点（場面）を探すのだけでも長い時間を要してしまうことになる。

【0004】

そこで会議を撮影した映像コンテンツを後から振り返る目的で、映像コンテンツに検索のためのタグ付けを行い、また重要度を算出して、時間軸上、映像の重要な要点（場面）の位置を特定する技術が知られている。例えば特許文献 1 には、音声や画像などのマルチメディア情報を用いた会議システムにおいて、会議中の重要な項目を短時間で簡便に編集できる装置およびユーザインタフェースを提供する目的で、会議でのデータを取り込む動

10

20

30

40

50

画入力手段、静止画入力手段、音声入力手段、ペン入力手段、ポインティング手段およびキー入力手段のうち少なくとも1つ以上により入力し、入力されたデータをデータ格納手段に格納し、データの時間関係を解析して検索性ファイルを作成し、作成された検索性ファイルを格納し、格納された検索性ファイルの参照結果をもとに該当するデータをデータ格納手段から読み出し、データを表示および編集することにより会議録の作成を支援する構成が開示されている。

【発明の概要】

【発明が解決しようとする課題】

【0005】

しかしながら、従来の映像を振り返る技術（例えば特許文献1）は、映像コンテンツのストリームを時間軸に沿って表示し、表示された映像の中からコンテキストの重要度の高いものについては色分けなどで表示するものであるため、特に長時間に及ぶ会議映像の場合、重要な場面を探すのには依然相応の時間を要することになる。また色分けされた映像だけが表示されて重要な場面の位置は特定できたとしても、実際にその映像場面を再生して内容を確認しないことには、なかなか要点内容までを把握することは困難である。即ち従来の技術は、映像コンテンツなど時間情報のある情報の振り返りにおいてユーザフレンドリーの観点からそのコンテンツの表示方法に依然改良されるべき余地があった。

10

【0006】

本発明は、上記の点に鑑みてなされたものであって、映像コンテンツなど時間情報のある情報の振り返りにおいて、重要なコンテキストとその要点内容を容易に把握可能なコンテンツ閲覧装置、コンテンツ表示方法及びコンテンツ表示プログラムを提供することを目的とする。

20

【課題を解決するための手段】

【0007】

上記の目的を達成するために、本発明に係るコンテンツ閲覧装置は、コンテンツを表示手段に表示するコンテンツ閲覧装置であって、映像コンテンツ及び前記映像コンテンツと時間軸上でリンク付けされた音声、静止画像のうち少なくともいずれかを有するデータ格納手段と、前記映像コンテンツから抽出された映像シーンが格納され、また前記音声、前記静止画像のうち少なくともいずれかから抽出されたキーワードと当該キーワード毎に解析された重要度とが対応付けて格納された解析データ格納手段と、前記解析データ格納手段から前記重要度が上位所定数のキーワードを読み出すとともに、読み出された前記キーワードの抽出元となった前記音声又は静止画像と、当該音声又は静止画像と同一の時間軸上でリンク付けされた映像シーンを表示データとして読み出す読出制御手段と、前記読出制御手段により読み出された前記表示データのうち、同一の時間軸上でリンク付けされた映像シーン、音声及び/又は静止画像、並びにキーワードのグループを表示するとともに、前記グループが複数ある場合にはグループ毎に時系列に並べ、前記表示手段に表示させる表示制御手段と、を有することを特徴とする。

30

【0008】

また、上記の目的を達成するために、上記コンテンツ閲覧装置において、前記映像コンテンツから抽出された映像シーンと当該映像シーン毎に解析された重要度とが対応付けて格納された第2解析データ格納手段と、前記第2解析データ格納手段から前記重要度が上位所定数の映像シーンを読み出すとともに、読み出された前記映像シーンと同一の時間軸上でリンク付けされた音声及び静止画像とを第2表示データとして読み出す第2読出制御手段と、前記第2読出制御手段により読み出された前記表示データのうち、同一の時間軸上でリンク付けされた映像シーン、並びに音声及び/又は静止画像のグループを表示するとともに、前記グループが複数ある場合にはグループ毎に時系列に並べ、前記表示手段に表示させる第2表示制御手段と、を有することを特徴とする。

40

【0009】

また、上記の目的を達成するために、上記コンテンツ閲覧装置において、前記表示制御手段は、前記抽出元となった音声又は静止画像は強調して表示し、前記第2表示制御手段

50

は、前記重要度が上位所定数の映像シーンは強調して表示すること、を特徴とする。

【0010】

なお、本発明の構成要素、表現または構成要素の任意の組合せを、方法、装置、システム、コンピュータプログラム、記録媒体、などに適用したのも本発明の態様として有効である。

【発明の効果】

【0011】

本発明によれば、映像コンテンツなど時間情報のある情報の振り返りにおいて、重要なコンテキストとその要点内容を容易に把握可能なコンテンツ閲覧装置、コンテンツ表示方法及びコンテンツ表示プログラムを提供することができる。

10

【図面の簡単な説明】

【0012】

【図1】実施形態に係る会議収録システムの一連の処理の流れを示す図である。

【図2】本実施形態に係る表示画面例を示す。

【図3】会議収録システム1の一実施形態の主要構成を示すハードウェア構成図である。

【図4】本実施形態に係る会議収録システム1の一実施形態の主要機能を示す機能ブロック図である。

【図5】データ格納手段及び解析データ格納手段のDB構成例を示す図である。

【図6】映像コンテンツデータDB510構成例を示す。

【図7】音声データDB520構成例を示す。

20

【図8】ホワイトボード画像DB530構成例を示す。

【図9】スライド画像DB540構成例を示す。

【図10】映像コンテンツデータから切り出された映像シーンDB511構成例を示す。

【図11】映像シーンから抽出された場面転換DB512構成例を示す。

【図12】音声データから取り出された音声認識データDB521構成例を示す。

【図13】音声認識データから解析された話者DB522構成例を示す。

【図14】音声認識データ、ホワイトボード画像及びスライド画像から解析されたキーワードDB550構成例を示す。

【図15】データの読み出し及び表示処理を説明するフローチャートである。

【図16】読み出された解析データ例を示す。

30

【発明を実施するための形態】

【0013】

以下、本発明を実施するための最良の形態について図面を参照して説明する。

【0014】

[システム構成]

(概要)

本発明は、映像コンテンツなど時間情報のある情報の振り返りにおいて、重要なコンテキストとその要点内容を容易に把握可能にするコンテンツ閲覧装置を提供するものであるが、コンテンツ閲覧する前に、映像コンテンツを取得、解析する必要があることから、以下の実施形態では、映像コンテンツを取得、解析、表示(閲覧)、再生を含む一連の処理機能を備える会議収録システム(便宜上このように呼ぶ)に本発明を適用した例を示す。

40

【0015】

図1は、実施形態に係る会議収録システムの一連の処理の流れを示す図である。実施形態に係る会議収録システムは、映像コンテンツを取得、解析、表示(閲覧)、再生を含む一連の処理機能を備え、図に示されるように大きくS100~400のステップの流れでその処理を進める。

【0016】

S100:データ入力ステップである。本実施形態においてデータは会議に関する会議データをいい、具体的に映像コンテンツデータ、音声(映像コンテンツデータに含まれてもよい)、画像、マウス、キーボードからの入力指示データなどである。会議が進行され

50

るにつれ、会議収録システムは各入力手段を介しこれら会議データを入力し続ける。会議が終了するとデータ入力も終了する。

【 0 0 1 7 】

S 2 0 0 : データ解析ステップである。S 1 0 0 で入力されたデータは格納手段 ( 記憶手段 ) に格納されており、会議収録システムはこのデータに対してデータ解析を行う。データ解析は、例えば映像ストリームを分割したり、各シーンに対し重要度を算出したり、音声からテキストを抽出しその話者を特定したり、ホワイトボードやスライドを解析しキーワードを抽出したりまたその重要度の算出などを行う。データ解析結果は格納手段 ( 記憶手段 ) に格納される。

【 0 0 1 8 】

S 3 0 0 : データの読み出し及び表示ステップである。このステップはユーザが会議の映像コンテンツの振り返り閲覧を行う場面であり、例えばユーザは会議収録システムの表示手段から、1の会議を指定し所定操作を行うと、その会議コンテンツの中から、表示手段上表示可能なコマ数分の重要なコンテキストが表示される。また重要とされたコンテンツだけでなく、時間軸上同時刻の他のコンテキストも対応するように表示される。

【 0 0 1 9 】

図 2 は、本実施形態に係る表示画面例を示す。当会議では、入力データとして、映像コンテンツデータ、音声データ、ホワイトボード画像、スライド ( 電子スライド ) 画像が入力されたため、データ解析の結果、映像シーン、話者、ホワイトボード、スライド、及びキーワードというコンテキストに分解、解析される。このうち重要度の高いコンテキスト ( 強枠表示のもの ) が抽出されるとともに、その重要度の高いコンテキストと時間軸上同時刻の他のコンテキストも対応するよう並列して表示される。当会議では4つの重要度の高いコンテキストが抽出、表示されており、ユーザは当会議において特にこれら4つが重要な内容を持つ場面 ( 映像シーン ) であると認識するとともに、キーワードからより具体的にどのような内容であったかなどもあわせて認識できる。

【 0 0 2 0 】

S 4 0 0 : シーンを選択及び映像の再生ステップである。会議収録システムの表示手段上、いくつかの重要なコンテキストが表示されているので、ユーザはこの中から再生しようとする映像シーンを選択し再生できる。会議中、選択したシーンの時点から実際に映像コンテンツを再生し、ユーザは局所的に会議を振り返ることができる。例えば、いずれかのコンテキストを選択し、「再生」ボタンを押下すると、再生画面に切り替わりこの時刻から映像コンテンツが再生される。

【 0 0 2 1 】

( ハードウェア )

ここで、会議収録システム ( 会議収録装置 ) 1 のハードウェア構成について簡単に説明しておく。図 3 は、会議収録システム 1 の一実施形態の主要構成を示すハードウェア構成図である。会議収録システム 1 は、主要な構成として、CPU 1 0 1、ROM ( Read Only Memory ) 1 0 2、RAM ( Random Access Memory ) 1 0 3、補助記憶装置 1 0 4、記憶媒体読取装置 1 0 5、入力装置 1 0 6、表示装置 1 0 7、通信装置 1 0 8、及びインターフェース装置 1 0 9 を含む構成である。

【 0 0 2 2 】

CPU 1 0 1 は、マイクロプロセッサ及びその周辺回路から構成され、装置全体を制御する回路である。また、ROM 1 0 2 は、CPU 1 0 1 で実行される所定の制御プログラム ( ソフトウェア部品 ) を格納するメモリであり、RAM 1 0 3 は、CPU 1 0 1 が ROM 1 0 2 に格納された所定の制御プログラム ( ソフトウェア部品 ) を実行して各種の制御を行うときの作業エリア ( ワーク領域 ) として使用するメモリである。

【 0 0 2 3 】

補助記憶装置 1 0 4 は、汎用の OS ( Operating System )、各種プログラムを含む各種情報を格納する装置であり、不揮発性の記憶装置である HDD ( Hard Disk Drive ) などが用いられる。

10

20

30

40

50

## 【 0 0 2 4 】

入力装置 1 0 6 は、ユーザが各種入力操作を行うための装置である。入力装置 1 0 6 は、マウス、キーボード、表示装置 1 0 7 の表示画面上に重畳するように設けられたタッチパネルスイッチなどを含む。表示装置 1 0 7 は、各種データを表示画面に表示する装置である。例えば、LCD(Liquid Crystal Display)、CRT(Cathode Ray Tube)などから構成される。

## 【 0 0 2 5 】

通信装置 1 0 8 は、ネットワークを介して他の機器との通信を行う装置である。有線ネットワークや無線ネットワークなど含む各種ネットワーク形態に応じた通信をサポートする。

10

## 【 0 0 2 6 】

インターフェース装置 1 0 9 は、外部の周辺機器と接続するためのインターフェースである。本実施形態では、ビデオカメラ(動画映像用)、デジタルカメラ(静止画用)、マイクなど主にデータキャプチャ機器と接続され、このインターフェース装置 1 0 9 を介してデータが入力される。

## 【 0 0 2 7 】

(機能構成)

次に、本実施形態に係る会議収録システム 1 の主要機能構成についてそれぞれ簡単に説明する。図 4 は、本実施形態に係る会議収録システム 1 の一実施形態の主要機能を示す機能ブロック図である。図に示すように会議収録システム 1 は、主要な機能として、データ入力手段 2 0 1、計時手段 2 0 2、データ登録手段 2 0 3、データ格納手段 2 0 4、データ解析手段 2 0 5、解析データ格納手段 2 0 6、表示制御手段 2 0 7、表示データ読出制御手段 2 0 8、操作手段 2 0 9、表示手段 2 1 0、再生手段 2 1 1 を含み構成される。

20

## 【 0 0 2 8 】

データ入力手段 2 0 1 は、データを入力する手段である。例えば上述のインターフェース装置 1 0 9 によって実現され、入力されるデータは、例えばビデオカメラの映像コンテンツデータ、マイクからの音声データ、カメラによるホワイトボードのスクリーンキャプチャ(静止画像)、スライドの資料データ(静止画像)などである。

## 【 0 0 2 9 】

計時手段 2 0 2 は、時間を秒単位などで計測し、データ登録手段 2 0 3 からの要求に応じて、その時点での時刻を返答する。

30

## 【 0 0 3 0 】

データ登録手段 2 0 3 は、データ入力手段 2 0 1 から入力されたデータをデータ格納手段 2 0 4 に格納する。データ格納手段 2 0 4 にまだ登録されていないデータであれば、計時手段 2 0 2 に時刻を問い合わせ、その時刻情報を一緒にデータ格納手段 2 0 4 に登録する。すでに登録されている、例えば映像や音声のようなストリーミング情報の場合は、時刻を計時手段 2 0 2 に問い合わせない。また静止画像の場合も計時手段 2 0 2 に時刻を問い合わせ、その時刻情報を一緒にデータ格納手段 2 0 4 に登録する。

## 【 0 0 3 1 】

データ格納手段 2 0 4 は、データ登録手段 2 0 3 から登録要求のあったデータを格納する。またその時刻情報も格納する。またデータ解析手段 2 0 5 や表示データ読出制御手段 2 0 8 からの要求に応じて、格納されたデータを取り出す。データ格納手段 2 0 4 は例えば上述の補助記憶装置 1 0 4 によって実現され、HDD(ハードディスク)などであってよい。なおデータ格納手段 2 0 4 はDB(データベース)によって格納データを管理する(具体例後述)。

40

## 【 0 0 3 2 】

データ解析手段 2 0 5 は、データ格納手段 2 0 4 に格納されたデータを取り出し、特定の処理を行ってメタデータを取り出す。またメタデータに基づいて重要度を算出する。データ解析手段 2 0 5 は、例えば、場面転換抽出手段、話者識別手段、キーフレーム抽出手段、スライド抽出手段、音声認識手段、キーワード抽出手段などから構成される。

50

## 【 0 0 3 3 】

解析データ格納手段 2 0 6 は、データ解析手段 2 0 5 によって解析されたデータを格納する。また解析データ格納手段 2 0 6 は表示データ読出制御手段 2 0 8 からの要求に応じてデータを取り出す。なお解析データ格納手段 2 0 6 は D B によって格納データを管理する（具体例後述）。

## 【 0 0 3 4 】

表示制御手段 2 0 7 は、操作手段 2 1 0 によって操作された情報を基に表示データ読出制御手段 2 0 8 に対して読み出し要求を出す。また読み出した情報に従って表示手段 2 0 9 に対して表示要求を出す。

## 【 0 0 3 5 】

表示データ読出制御手段 2 0 8 は、表示制御手段 2 0 9 からの要求に応じて、解析データ格納手段 2 0 6 に対して解析データ（解析結果）の取り出しを要求する。またその解析データを比較演算し、データ格納手段 2 0 4 に対して必要なデータの取り出しを要求する。取り出したデータを表示データとして表示制御手段 2 0 7 に渡す。

## 【 0 0 3 6 】

表示手段 2 0 9 は、表示制御手段 2 0 7 によって表示を要求されたものについて、表示を行う。表示手段 2 0 9 は例えば上述の表示装置 1 0 7 によって実現され、ディスプレイなどであってよい。

## 【 0 0 3 7 】

操作手段 2 1 0 は、ユーザに操作をさせる手段を提供する。操作手段 2 1 0 は例えば上述の入力装置 1 0 6 によって実現され、マウスやキーボード、ペン入力デバイスなどであってよい。

## 【 0 0 3 8 】

再生手段 2 1 1 は、データ格納手段 2 0 4 に格納された映像コンテンツの再生を行う。ユーザから対象となる映像コンテンツ及び再生時間等が指定されると、映像コンテンツを再生する。メディア再生プレーヤなどで実現されればよい。

## 【 0 0 3 9 】

以上これらの機能は、実際には装置の C P U 1 0 1 が実行するプログラムによりコンピュータに実現させるものである。

## 【 0 0 4 0 】

## [情報処理]

上述したように、実施形態に係る会議収録システム 1 は、映像コンテンツを取得、解析、表示（閲覧）、再生を含む一連の処理機能を備え、大きく S 1 0 0 ~ 4 0 0 のステップ（図 1）の流れでその処理を進める。

## 【 0 0 4 1 】

図 5 は、データ格納手段及び解析データ格納手段の D B 構成例を示す図である。以下同図をあわせて参照しながら説明をしていく。

## 【 0 0 4 2 】

（データ入力：S 1 0 0）

実施形態に係る会議収録システム 1 は、データ入力のための周辺機器が接続され、本実施形態において入力されるデータは、映像コンテンツデータ、音声データ、ホワイトボード画像（静止画像）、スライドのスライド画像（静止画像）であるものとする。

## 【 0 0 4 3 】

例えば会議室には会議室全体を見渡すようにビデオカメラが設置される。またもしくは撮影者が随時ビデオカメラを扱って自由に撮影することもできる。ビデオカメラから取得された映像は時間情報を含むストリーミングデータである。また例えば会議室にはホワイトボードが設置され一定間隔又はユーザ操作によりホワイトボードの手書きのスクリーン画像がキャプチャされる。また例えば会議室ではユーザ P C の電子資料データが壁側のスクリーンなど投影されたり、ユーザ P C 間で電子資料データが共有されており、このスライド画像がキャプチャされる。

10

20

30

40

50

## 【 0 0 4 4 】

会議が開始されるとこれらデータが入力され始め、会議の終了とともにデータの入力を終了する。会議中、データはデータ登録手段 2 0 3 によりデータ格納手段 2 0 4 に格納される。なおデータ登録手段 2 0 3 は計時手段 2 0 2 に時刻を問い合わせ、その時刻情報を一緒にデータ格納手段 2 0 4 に登録しておく。

## 【 0 0 4 5 】

図 6 は、映像コンテンツデータ DB 5 1 0 構成例を示す。データ格納手段 2 0 4 は、映像コンテンツデータを格納するにあたり図に示される DB でもってデータを管理する。映像コンテンツデータ DB 5 1 0 は、ID、ファイルの場所、開始時間などから構成される。ID は、DB 上映像を一意に特定するための ID である。ファイルの場所は、その映像が格納されている場所を示すものである。ファイルパスのような記述によって格納場所を特定してもよい。そして例えば「0001.avi」が映像コンテンツデータ（実体）である。開始時間は、その映像の格納が開始された時間が計時手段 2 0 2 によって付与されたものであり、例えば UTC フォーマットで記録されたものである。本図例によると、3 つの ID を持つ映像コンテンツデータが格納（登録）されていることから、3 つの会議分の映像が DB に格納されていることが分かる。

10

## 【 0 0 4 6 】

図 7 は、音声データ DB 5 2 0 構成例を示す。データ格納手段 2 0 4 は、音声データを格納するにあたり図に示される DB でもってデータを管理する。音声データ DB 5 2 0 は、ID、ファイルの場所、開始時間などから構成される。ID は、DB 上音声を一意に特定するための ID である。ファイルの場所は、その音声データが格納されている場所を示すものである。そして例えば「0001.wav」が音声データ（実体）である。開始時間は、その音声の格納が開始された時間が計時手段 2 0 2 によって付与されたものである。本図例によると、3 つの ID を持つ音声データが格納（登録）されていることから、3 つの会議分の音声が DB に格納されていることが分かる。

20

## 【 0 0 4 7 】

図 8 は、ホワイトボード画像 DB 5 3 0 構成例を示す。データ格納手段 2 0 4 は、ホワイトボード画像を格納するにあたり図に示される DB でもってデータを管理する。ホワイトボード画像 DB 5 3 0 は、ID、ファイルの場所、開始時間などから構成される。ID は、DB 上ホワイトボード画像を一意に特定するための ID である。ファイルの場所は、そのホワイトボード画像が格納されている場所を示すものである。そして例えば「w0001.jpg」がホワイトボード画像データ（実体）である。開始時間は、そのホワイトボード画像の格納が開始された時間が計時手段 2 0 2 によって付与されたものである。本図例のホワイトボード画像は、とくに時間情報（2010-04-05 13:10:33～）に注目すれば、図 6 の ID 「m1」及び図 7 の ID 「a1」に対応する会議においてキャプチャされたホワイトボード画像データであることが分かる。

30

## 【 0 0 4 8 】

図 9 は、スライド画像 DB 5 4 0 構成例を示す。データ格納手段 2 0 4 は、スライド画像を格納するにあたり図に示される DB でもってデータを管理する。スライド画像 DB 5 4 0 は、ID、ファイルの場所、開始時間などから構成される。ID は、DB 上スライド画像を一意に特定するための ID である。ファイルの場所は、そのスライド画像が格納されている場所を示すものである。そして例えば「s0001.jpg」がスライド画像データ（実体）である。開始時間は、そのスライド画像の格納が開始された時間が計時手段 2 0 2 によって付与されたものである。本図例のスライド画像は、特に時間情報（2010-04-05 13:10:33～）に注目すれば、図 6 の ID 「m1」及び図 7 の ID 「a1」に対応する会議においてキャプチャされたスライド画像データであることが分かる。

40

## 【 0 0 4 9 】

（データ解析：S 2 0 0）

上述したように入力されたデータはデータ格納手段 2 0 4 に格納されており、会議収録システム 1 はこのデータに対してデータ解析を行う。データ解析は、例えば映像ストーリー

50

ムを分割したり、各シーンに対し重要度を算出したり、音声からテキストを抽出しその話者を特定したり、ホワイトボードやスライドを解析しキーワードを抽出したりまたその重要度の算出などを行う。データ解析結果は解析データ格納手段206に格納される。以下図5をあわせて参照しながら説明する。

【0050】

図10は、映像コンテンツデータから切り出された映像シーンDB511構成例を示す。データ解析手段205は、映像コンテンツデータのストリーミングから映像シーン毎にイメージを切り出し、これを映像シーンDB511に格納する。よって図10の映像IDは図6のIDに対応する。例えば図10のID「si1」は、図6の映像コンテンツデータ510のID「m1」の「0001.avi」から切り出され抽出されたものである。なお映像シーン画像が切り出された時間帯は時間情報として格納される。

10

【0051】

図11は、映像シーンから抽出された場面転換DB512構成例を示す。データ解析手段205は、映像シーンから特に重要と判断される映像シーンを抽出し、これを場面転換DB512に格納する。よって図11の映像シーンIDは図10のIDに対応する。例えば図11のID「sc1」は、図10の映像シーンDB511のID「si1」を重要と判断された結果、抽出されたものである。そしてデータ解析手段205は、各場面転換を示すと判断した映像シーンに対し重要度を計算し、あわせてDBに格納する。重要度は、場面転換について、どのくらいの情報が変化したのかを重要度として計算した結果を示すものであるが、この場面転換重要度の具体的算出方法は、例えば本出願人による特許4414254号などを参考にできる。

20

【0052】

図12は、音声データから取り出された音声認識データDB521構成例を示す。データ解析手段205は、音声データのストリーミングから音声認識処理を行い、これをテキストに変換、また所定文節に区切って音声認識データを作成し、これを音声認識データDB521に格納する。よって図12の音声IDは図7のIDに対応する。例えば図12のID「t1」は、図7の音声データDB520のID「a1」の「0001.wav」から抽出されたものである。時間情報は、音声の開始時間と経過時間から計算され、元の音声データにおける当該音声認識データの開始時間を示すものである。

30

【0053】

図13は、音声認識データから解析された話者DB522構成例を示す。データ解析手段205は、音声データ又は音声認識データから音声認識処理を行い、話者（話者ID）を特定し、これを話者DB522に格納する。よって図13の音声IDは図7のIDに対応する。例えば図13のID「a1」は、図7の音声データDB520のID「a1」の「0001.wav」から話者が特定されたものである。時間情報は、音声の開始時間と経過時間から計算され、元の音声データにおける当該音声認識データの開始時間を示すものである。なお話者を特定する方法としては、例えば予め想定される話者名、話者ID、話者写真画像及びそれぞれの声紋等に対応付けて登録しておき、音声データ又は音声認識データと照合することにより話者を特定できる。

40

【0054】

図14は、音声認識データ、ホワイトボード画像及びスライド画像から解析されたキーワードDB550構成例を示す。キーワードDB550は、ID、コンテンツ、キーワード、重要度、時間情報などから構成される。IDはDB上キーワードを一意に特定するためのIDである。コンテンツは、そのキーワードがどこから取得されたものかの抽出元を示し、音声認識データのDB、ホワイトボード画像のDB、スライド画像のDBなどのIDを示す。キーワードは、入力データ又は入力データから解析された解析データなどの音声認識データ、ホワイトボード画像、スライド画像から抽出されたキーワードあるいはキーワードフレーズを格納する。重要度は、計算によってキーワード重要度が演算されたものを格納したものである。

50

【0055】

データ解析手段205は、音声認識データ、ホワイトボード画像、スライド画像からキーワード抽出処理を行うとともに、抽出されたキーワードの重要度を算出し、これを解析データ格納手段206のDBに格納する。なおキーワードを抽出する方法としては、例えばホワイトボード画像、スライド画像からは文字認識処理(OCR)を行ってテキストを取り出す(ホワイトボード画像OCRDB531、スライド画像OCRDB541に格納)。音声データからは音声認識データとして既にテキストとして取り出されており(音声認識データDB521に格納)、これらテキストを形態素に分解し、各々の形態素に対してTF(単語の出現頻度:Term Frequency) - IDF(逆出現頻度:Inverse Document Frequency)値等の重要度を表す指標を計算し、その指標(重要度)の一定以上高い形態素をキーワードとすることができる。

10

## 【0056】

例えば図中、キーワードのID「k1」は、コンテンツ「s11」(スライド画像)からキーワード「System」が抽出され、その重要度が「0.4」であることを示している。またキーワードのID「k2」は、コンテンツ「w1」(ホワイトボード画像)からキーワード「Architecture」が抽出され、その重要度が「0.2」であることを示している。また同様に、キーワードのID「k3」は、コンテンツ「t1」(音声認識データ)からキーワード「システム構成」が抽出され、その重要度が「0.6」であることを示している。なお音声認識データの「t1」(図12)は、そのコンテンツが「最初の議題は、システム構成についてです。」となっており、本会議の議題を示すキーワード「システム構成」には比較的高い重要度「0.6」が算出されている。

20

## 【0057】

(データの読み出し及び表示:S300)

このステップはユーザが会議の映像コンテンツの閲覧を行う場面であり、例えばユーザは会議収録システム1の表示手段から、収録会議リストの中から1の会議を選択し、選択された会議について重要場面のコンテキストを表示する操作を行う。これを受け会議収録システム1は、選択された会議を対象として、解析データ格納手段206の解析データに基づいて、重要なコンテキストを抽出、表示する。

## 【0058】

図15は、データの読み出し及び表示処理を説明するフローチャートである。本実施形態に係る会議収録システム1は、ユーザにより、1の会議が選択され、重要場面表示の操作を受けてフローチャートの処理を開始する。なおユーザは操作時、重要度判断の対象となるコンテキストを映像シーンにするか、キーワードにするか、それとも映像シーンとキーワードの両方にするか、の指定を行う。また重要場面表示に際し、何コマ(何シーン)を一度に表示させるかのコマ数の指定を行う。ここでは、「映像シーンとキーワードの両方」、「4コマ」と指定されたものとして説明する。

30

## 【0059】

S301:まず表示データ読出制御手段208は、ユーザより選択された会議に基づき、キーワードDB550からこの会議に対応するキーワードを重要度順に並べる。なお解析データ格納手段206によりキーワードDB550は既に重要度順にソート済みであってもよい。

40

## 【0060】

S302:同様に表示データ読出制御手段208は、ユーザより選択された会議に基づき、映像シーンDB511からこの会議に対応する映像シーンを重要度順に並べる。なお解析データ格納手段206により映像シーンDB511は既に重要度順にソート済みであってもよい。

## 【0061】

S303:表示データ読出制御手段208は、所定上位数のキーワード、映像シーンを表示データとして取得する(読み出す)。所定上位数は、指定コマ数により決定される値である。ここでは指定コマ数「4コマ」であるので、所定上位数は4であるので、上位4番目までのキーワード、映像シーンを取得する。取得方法としては、重要度判断の対象と

50

なるコンテンツが「キーワード」(のみ)である場合、重要度が上位4番目までのキーワードを取得する。また重要度判断の対象となるコンテンツが「映像シーン」(のみ)である場合、重要度が上位4番目までの映像シーンを取得する。また、重要度判断の対象となるコンテンツが「映像シーンとキーワードの両方」である場合には、それぞれ2つずつ、つまり重要度が上位2番目まで映像シーンとキーワードを取得してもよいし、どちらかを重視するのであれば重み付けによりいずれかを優先し取得するようにしてもよい。またもしくは映像シーンとキーワードの重要度が正規化(標準化)されているのであれば、映像シーン及びキーワードの両方を含めた中から重要度が上位4番目までの映像シーン及び/又はキーワードを取得するようにしてもよい。

【0062】

S304: また表示データ読出制御手段208は、取得されたキーワード、映像シーンと同時刻のコンテキストを表示データとして取得する。具体的に、重要度に基づきあるキーワードが取得された場合、このキーワードは音声認識データ、ホワイトボード画像、スライド画像のいずれかから抽出されているところ、この抽出元のコンテキストの時間情報を特定し、特定された時間情報と同時刻の他のコンテキストを取得する。

【0063】

具体的に例えばS303にて、重要度「0.6」のキーワードID「k3」が所定上位数内に入ったためこのキーワード「システム構成」が取得されたとする。キーワードDB550を参照すると、このキーワード「システム構成」の抽出元はコンテンツ「t1」となっており、即ち音声認識データのID「t1」(図12)が抽出元である。そして音声認識データのID「t1」において、時間情報は「2010-04-05 13:10:32」である。従って、本キーワードと同時刻のコンテキストである音声認識データID「t1」を取得する。また同時刻の他のコンテキストとして、「2010-04-05 13:10:32」の映像シーン、「2010-04-05 13:10:32」のホワイトボード画像、「2010-04-05 13:10:32」のスライド画像、「2010-04-05 13:10:32」時点での話者をそれぞれ取得する。映像シーンは映像シーンDB511から、ホワイトボード画像はホワイトボード画像DB530(又はホワイトボード画像OCRDB531)から、スライド画像はスライド画像DB540(又はスライド画像OCRDB541)から、話者は話者DB522を読み出すことができる。なお図9のスライド画像DB540について、同時刻のスライド画像がなければ直近のスライド画像をこれに代えて取得する(スライドは切り替わり毎に画像を取得しているため)。

【0064】

図16は、読み出された解析データ例を示す。上述のS303及びS304を経て、図に示されるように5つのコンテキスト(要素)が抽出、読み出されることになる。ここでこれら5つのコンテキストは、会議の開催時間中、同時刻に発生したコンテキストであり、いわば同時刻コンテキストグループといえる。そしてここでは、上述の如く重要場面表示を一度に表示させるかの指定コマ数は「4コマ」であるので、この要領で表示データとして、4つの同時刻コンテキストグループが抽出、読み出されることになる。

【0065】

S305: 表示制御手段207は、表示データ読出制御手段208により読み出された表示データを表示手段209に表示させる。ここで再び図2を参照する。表示の方法は、例えば図2に示されるように同時刻コンテキストグループをコンテキストの種類順に縦に並べる。「13:10:32」の同時刻コンテキストグループ(図16)の場合、グループ内の各コンテキスト要素を上から「映像シーン」、「話者」、「ホワイトボード」、「スライド」、「キーワード」の順に縦に整列させる。同様の要領で、他3つの同時刻コンテキストグループについても縦に整列させる。そして横軸には時系列に縦に整列した同時刻コンテキストグループを並べていく。なおこれらコンテキストは重要度に基づき抽出されているため、その根拠となったコンテキストに対しては強調表示を行うようにする。

【0066】

(シーンの選択及び映像の再生: S400)

このステップでは、会議収録システム1の表示手段上、いくつかの重要なコンテキスト

10

20

30

40

50

が表示されているので、ユーザはこの中から再生しようとする映像シーンを選択し再生できる。再び図2を参照し、ユーザは再生しようとする映像シーン（又は同グループ内のコンテンツ）を選択し「再生」ボタンを押下する。再生手段211は、選択された映像シーンの時間情報が取得し、映像シーンに対応するファイルを特定（図6）し、同ファイルを取得した時間から再生を開始する。これまでの例でいえば、この会議の映像ファイル「0001.avi」が時間ストリーム上「13:10:32」（時間情報）から映像及び音声を含め再生される。

#### 【0067】

##### [総括]

以上本実施形態に係る会議収録システム1によれば、映像コンテンツデータの振り返りにおいて、映像コンテンツデータや、その他の入力データの中から重要度の高いものだけを時間軸に沿って一覧表示される。また表示されたコンテキストのうち重要度の高いものについてはユーザの視認性を高めるように強調表示（色分け含む）などで表示する。また会議場面にて同時刻に平行している他のコンテキストについても同時に表示する。これによりユーザは映像のみならず他の複数の要素から会議場面を思い起こし振り返ることができる。またそのうち、解析されたキーワードを付して表示することで、重要度の高い映像やコンテキスト（話者、ホワイトボード、スライド）とともに、その映像シーン時刻における要点内容までをユーザがキーワードとして一見で把握できるようにした。

#### 【0068】

即ち上述の本実施形態によれば、映像コンテンツなど時間情報のある情報の振り返りにおいて、重要なコンテキストとその要点内容を容易に把握可能なコンテンツ閲覧装置等を提供することが可能となる。

#### 【0069】

各実施形態に基づき本発明の説明を行ってきたが、上記各実施形態にあげたその他の要素との組み合わせなど、ここで示した要件に本発明が限定されるものではない。これら点に関しては、本発明の主旨をそこなわない範囲で変更することが可能であり、その応用形態に応じて適切に定めることができる。また、本発明の構成要素、表現または構成要素の任意の組合せを、方法、装置、システム、コンピュータプログラム、記録媒体、などに適用したのも本発明の態様として有効である。

#### 【符号の説明】

#### 【0070】

- |     |             |    |
|-----|-------------|----|
| 1   | 会議支援システム    |    |
| 101 | CPU         |    |
| 102 | ROM         |    |
| 103 | RAM         |    |
| 104 | 補助記憶装置      |    |
| 105 | 記憶媒体読取装置    |    |
| 106 | 入力装置        |    |
| 107 | 表示装置        |    |
| 108 | 通信装置        | 40 |
| 109 | インターフェース装置  |    |
| 201 | データ入力手段     |    |
| 202 | 計時手段        |    |
| 203 | データ登録手段     |    |
| 204 | データ格納手段     |    |
| 205 | データ解析手段     |    |
| 206 | 解析データ格納手段   |    |
| 207 | 表示制御手段      |    |
| 208 | 表示データ読出制御手段 |    |
| 209 | 操作手段        | 50 |

- 2 1 0 表示手段
- 2 1 1 再生手段
- 5 1 0 映像コンテンツデータ D B
- 5 1 1 映像シーン D B
- 5 1 2 場面転換 D B
- 5 2 0 音声データ D B
- 5 2 1 音声認識データ D B
- 5 2 2 話者 D B
- 5 3 0 ホワイトボード画像 D B
- 5 3 1 ホワイトボード画像 O C R D B
- 5 4 0 スライド画像 D B
- 5 4 1 スライド O C R 画像 D B
- 5 5 0 キーワード D B

【先行技術文献】

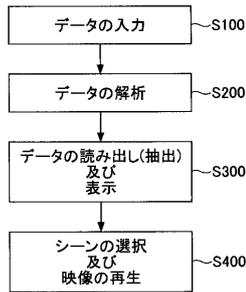
【特許文献】

【0071】

【特許文献1】特許第3185505号

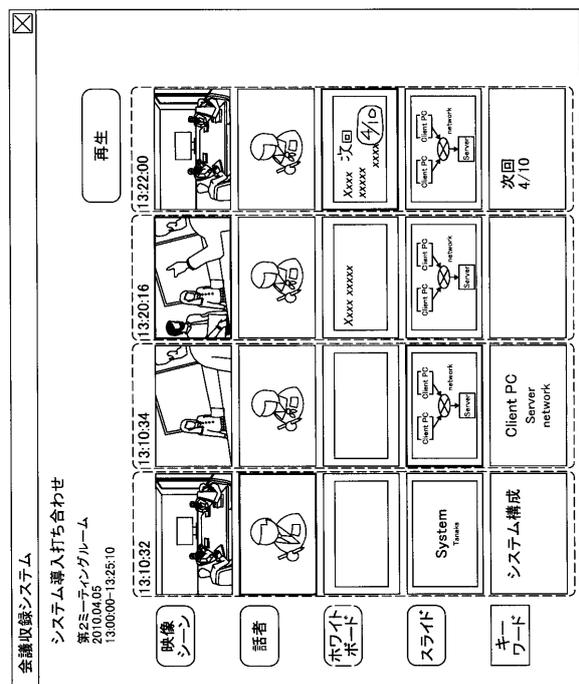
【図1】

実施形態に係る会議収録システムの一連の処理の流れを示す図



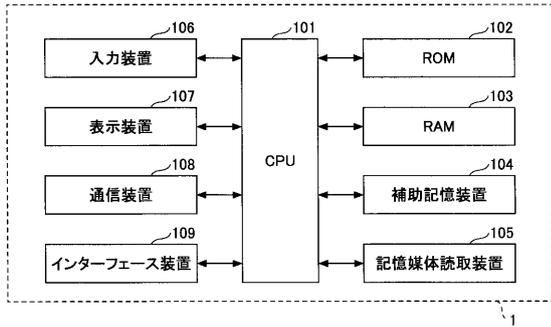
【図2】

本実施形態に係る表示画面例を示す図



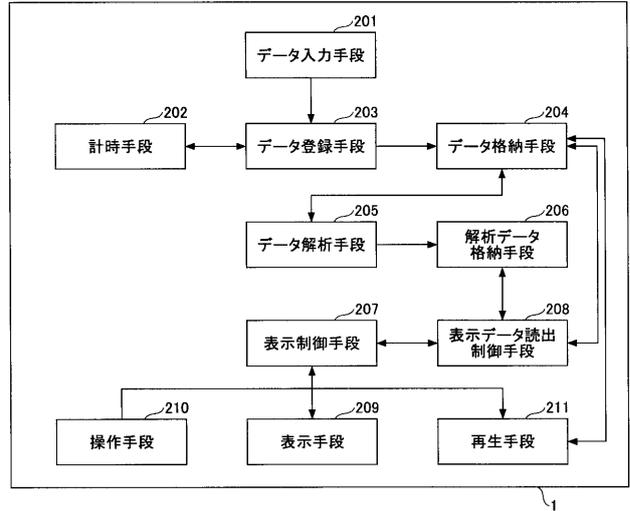
【 図 3 】

会議収録システム1の一実施形態の主要構成を示すハードウェア構成図



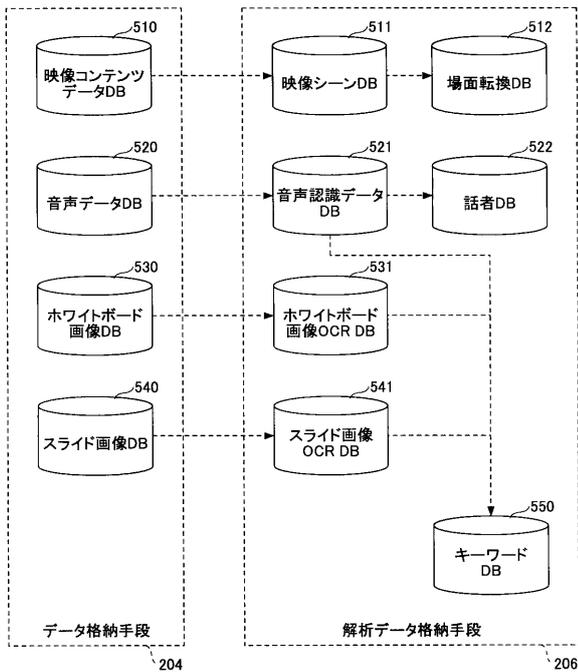
【 図 4 】

本実施形態に係る会議収録システム1の一実施形態の主要機能を示す機能ブロック図



【 図 5 】

データ格納手段及び解析データ格納手段のDB構成例を示す図



【 図 6 】

映像コンテンツデータDB510構成例を示す図

ID	ファイルの場所	開始時間
m1	0001.avi	2010-04-05 13:10:32
m2	0002.avi	2010-04-06 11:40:15
m3	0003.avi	2010-04-07 17:30:06

【 図 7 】

音声データDB520構成例を示す図

ID	ファイルの場所	開始時間
a1	0001.wav	2010-04-05 13:10:32
a2	0002.wav	2010-04-06 11:40:15
a3	0003.wav	2010-04-07 17:30:06

【 図 8 】

ホワイトボード画像DB530構成例を示す図

ID	ファイルの場所	時間情報
w1	w0001.jpg	2010-04-05 13:10:32
w2	w0002.jpg	2010-04-05 13:10:33
w3	w0003.jpg	2010-04-05 13:10:34
.	.	.
.	.	.
.	.	.

【 図 9 】

スライド画像DB540構成例を示す図

ID	ファイルの場所	時間情報
sl1	s0001.jpg	2010-04-05 13:10:32
sl2	s0002.jpg	2010-04-05 13:11:33
sl3	s0003.jpg	2010-04-05 13:14:34
.	.	.
.	.	.
.	.	.

【 図 1 0 】

映像コンテンツデータから切り出された映像シーンDB511構成例を示す図

ID	映像ID	ファイルの場所	時間情報
si1	m1	si0001.jpg	2010-04-05 13:10:32
si2	m1	si0002.jpg	2010-04-05 13:10:33
si3	m1	si0003.jpg	2010-04-05 13:10:34
.	.	.	.
.	.	.	.
.	.	.	.

【 図 1 1 】

映像シーンから抽出された場面転換DB512構成例を示す図

ID	映像シーンID	重要度
sc1	si1	0.1
sc2	si40	0.3
sc3	si200	0.2
.	.	.
.	.	.
.	.	.

【 図 1 2 】

音声データから取り出された音声認識データDB521構成例を示す図

ID	音声ID	コンテンツ	時間情報
t1	a1	本日は、お集まり いただきまして ありがとうございます	2010-04-05 13:10:32
t2	a1	さて早速ですが、 最初の議題から開始 したいと思います。	2010-04-05 13:10:43
t3	a1	最初の議題は、 システム構成に ついてです。	2010-04-05 13:10:50
.	.	.	.
.	.	.	.
.	.	.	.

【 図 1 3 】

音声認識データから解析された話者DB522構成例を示す図

ID	音声ID	話者ID	時間情報
sp1	a1	1	2010-04-05 13:10:32
sp2	a1	2	2010-04-05 13:11:11
sp3	a1	1	2010-04-05 13:12:20
.	.	.	.
.	.	.	.
.	.	.	.

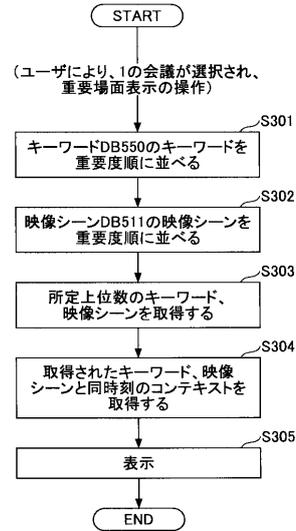
【 図 1 4 】

音声認識データ、ホワイトボード画像及びスライド画像から解析された  
キーワードDB550構成例を示す図

ID	コンテンツ	キーワード	重要度
k1	sl1	System	0.4
k2	w1	Architecture	0.2
k3	t3	システム構成	0.6
k4	w1320	次回 4/10	0.65
.	.	.	.
.	.	.	.
.	.	.	.

【 図 1 5 】

データの読み出し及び表示処理を説明するフローチャート



【 図 1 6 】

読み出された解析データ例を示す図

