(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号

特許第5339303号
(P5339303)

(45)発行日　平成25年11月13日(2013.11.13)　(24)登録日　平成25年8月16日(2013.8.16)

(51)Int.Cl.		F I		
H04N 5/76	(2006.01)	H04N	5/76	B
H04N 5/93	(2006.01)	H04N	5/93	Z
G11B 27/10	(2006.01)	G11B	27/10	A

請求項の数 10　(全 58 頁)

(21)出願番号	特願2010-503907(P2010-503907)
(86)(22)出願日	平成21年3月18日(2009.3.18)
(86)国際出願番号	PCT/JP2009/055315
(87)国際公開番号	WO2009/116582
(87)国際公開日	平成21年9月24日(2009.9.24)
審査請求日	平成24年3月12日(2012.3.12)
(31)優先権主張番号	特願2008-72537(P2008-72537)
(32)優先日	平成20年3月19日(2008.3.19)
(33)優先権主張国	日本国(JP)

(73)特許権者　504173471
国立大学法人北海道大学
北海道札幌市北区北８条西５丁目
(74)代理人　100083806
弁理士　三好　秀和
(74)代理人　100101247
弁理士　高橋　俊一
(72)発明者　長谷山　美紀
北海道札幌市北区北１４条西９丁目　国立大学法人　北海道大学　大学院情報科学研究科内

審査官　木方　庸輔

最終頁に続く

(54)【発明の名称】動画検索装置および動画検索プログラム

(57)【特許請求の範囲】

【請求項１】

動画データのシーンから、クエリ動画データに類似するシーンを検索する動画検索装置であって、

クエリ動画データを含む動画データが記憶された動画データベース記憶装置と、

前記動画データのビジュアル信号をショットに分割して、該ショットに対応するオーディオ信号の特徴量の差分が小さい連続したショットをシーンとして出力するシーン分割部と、

前記シーン分割部によって分割されたそれぞれのシーンについて、前記オーディオ信号のベース音の音高の推移に基づく類似度と、前記ベース音を除く音に基づく類似度を含む、シーン間のオーディオ信号の類似度を算出して、オーディオ信号類似度データを生成するオーディオ信号類似度算出部と、

前記オーディオ信号類似度データに基づいて、前記クエリ動画データのシーンと、前記シーン間の類似度が一定の閾値よりも小さいシーンを検索するオーディオ信号類似度検索部と、

を備える動画検索装置。

【請求項２】

前記オーディオ信号類似度検索部によって検索された各シーンについて該類似度に対応する座標を取得して表示するオーディオ信号類似度表示部

を更に備える請求項１に記載の動画検索装置。

【請求項３】
前記シーン分割部によって分割されたそれぞれのシーンについて、前記ビジュアル信号の特徴量と前記オーディオ信号の特徴量から、シーン間のビデオ信号の類似度を算出して、ビデオ信号類似度データを生成するビデオ信号類似度算出部と、
前記ビデオ信号類似度データに基づいて、前記クエリ動画データのシーンと、前記シーン間の類似度が一定の閾値よりも小さいシーンを検索するビデオ信号類似度検索部と、
を更に備える請求項１に記載の動画検索装置。
【請求項４】
前記ビデオ信号類似度検索部によって検索された各シーンについて該類似度に対応する座標を取得して表示するビデオ信号類似度表示部
を更に備える請求項３に記載の動画検索装置。
【請求項５】
前記オーディオ信号類似度算出部は、さらに、シーン間のオーディオ信号の類似度として、リズムに基づく類似度を算出して、前記オーディオ信号類似度データを生成し、
ビデオ信号類似度とオーディオ信号類似度に対する嗜好の割合である嗜好データを取得し、前記ビデオ信号類似度データおよび前記オーディオ信号類似度データに基づいて、前記ビジュアル信号の特徴量と前記オーディオ信号の特徴量から算出されたシーン間の類似度と、前記オーディオ信号のベース音に基づく類似度と、前記ベース音を除く音に基づく類似度と、前記リズムに基づく類似度とに対する重み係数を決定して、各シーンの各類似度にこの重み係数を乗算して統合された類似度に基づいて、前記シーン間の統合された類似度が一定の閾値よりも小さいシーンを検索する検索部と、
前記検索部によって検索された各シーンについて該統合された類似度に対応する座標を取得して表示する表示部
を更に備えることを特徴とする請求項３に記載の動画検索装置。
【請求項６】
動画データのシーンから、クエリ動画データに類似するシーンごとに検索する動画検索プログラムであって、
コンピュータを、
動画データベース記憶装置に記憶されたクエリ動画データおよび動画データのビジュアル信号をショットに分割して、該ショットに対応するオーディオ信号の特徴量の差分が小さい連続したショットをシーンとして出力するシーン分割手段と、
前記シーン分割手段によって分割されたそれぞれのシーンについて、前記オーディオ信号のベース音の音高の推移に基づく類似度と、前記ベース音を除く音に基づく類似度を含む、シーン間のオーディオ信号の類似度を算出して、オーディオ信号類似度データを生成するオーディオ信号類似度算出手段と、
前記オーディオ信号類似度データに基づいて、前記クエリ動画データのシーンと、前記シーン間の類似度が一定の閾値よりも小さいシーンを検索するオーディオ信号類似度検索手段
として機能させる動画検索プログラム。
【請求項７】
前記オーディオ信号類似度検索手段によって検索された各シーンについて該類似度に対応する座標を取得して表示するオーディオ信号類似度表示手段
として、更に前記コンピュータを機能させる請求項６に記載の動画検索プログラム。
【請求項８】
前記シーン分割手段によって分割されたそれぞれのシーンについて、前記ビジュアル信号の特徴量と前記オーディオ信号の特徴量から、シーン間のビデオ信号の類似度を算出して、ビデオ信号類似度データを生成するビデオ信号類似度算出手段と、
前記ビデオ信号類似度データに基づいて、前記クエリ動画データのシーンと、前記シーン間の類似度が一定の閾値よりも小さいシーンを検索するビデオ信号類似度検索手段
として、更に前記コンピュータを機能させる請求項６に記載の動画検索プログラム。

【請求項９】
前記ビデオ信号類似度検索手段によって検索された各シーンについて該類似度に対応する座標を取得して表示するビデオ信号類似度表示手段
として、更に前記コンピュータを機能させる請求項８に記載の動画検索プログラム。
【請求項１０】
前記オーディオ信号類似度算出手段は、さらに、シーン間のオーディオ信号の類似度として、リズムに基づく類似度を算出して、前記オーディオ信号類似度データを生成し、
ビデオ信号類似度とオーディオ信号類似度に対する嗜好の割合である嗜好データを取得し、前記ビデオ信号類似度データおよび前記オーディオ信号類似度データに基づいて、前記ビジュアル信号の特徴量と前記オーディオ信号の特徴量から算出されたシーン間の類似度と、前記オーディオ信号のベース音に基づく類似度と、前記ベース音を除く音に基づく類似度と、前記リズムに基づく類似度とに対する重み係数を決定して、各シーンの各類似度にこの重み係数を乗算して統合された類似度に基づいて、前記シーン間の統合された類似度が一定の閾値よりも小さいシーンを検索する検索手段と、
前記検索手段によって検索された各シーンについて該統合された類似度に対応する座標を取得して表示する表示手段
として、更に前記コンピュータを機能させる請求項８に記載の動画検索プログラム。
【発明の詳細な説明】
【技術分野】
【０００１】
本発明は、複数の動画データから、クエリ動画データに類似するシーンを検索する動画検索装置および動画検索プログラムに関する。
【背景技術】
【０００２】
近年の記憶媒体の大容量化やインターネットによる映像配信サービスの普及に伴い、ユーザは大量の映像を入手可能となった。しかしながら、ユーザが特定の映像を明示せずに、所望の映像を入手することは一般に困難である。これは、膨大なデータベースにおける映像の入手が、主に映像名や製作元等のキーワードを用いた検索に依存していることに起因する。このため、キーワードによる映像検索だけでなく、映像の構成に注目した検索や同一ジャンルの映像の検索等、映像の内容に基づく様々な検索技術の実現が期待されている。そこで、映像や楽曲間の類似度に着眼した手法が提案されている（例えば、特許文献１および特許文献２参照）。
【０００３】
特許文献１に記載の方法では、各動画データに、複数の単純図形との類似率を求めて記録した被検索用単純図形類似率情報が関連づけられている。一方、画像検索時に、検索画像について複数の単純図形との類似率を求めて記録した検索用類似率情報を作成する。被検索用単純図形類似率情報と、検索用類似率情報とを照合し、複数の単純図形ごとの類似率を集計し平均した類似率が、予め設定した規定類似率以上の場合、その動画データを類似動画として検索する。また、特許文献２に記載の方法では、映像データにおける類似映像区間とそれ以外を区別する類似映像区間情報を生成する。このとき、特許文献２に記載の方法では、ショットの画像の特徴量に基づいて類似パターンに分類する。
【０００４】
一方、感性に基づく単語をメタデータとして映像・楽曲に付加し、単語間の関係に基づいて映像・楽曲の類似度を算出する方法もある（非特許文献１および非特許文献２参照）。
【特許文献１】特開２００７－５８２５８号公報
【特許文献２】特開２００７－２７４２３３号公報
【非特許文献１】L. Lu, D. Liu and H. J. Zhang, “Automatic Mood Detection and Tracking of Music Audio Signals,” IEEE Trans. Audio, Speech and Language Proceeding, vol. 14, no. 1, pp. 5－8, 2006.

【非特許文献２】T. Li and M. Ogihara, “Toward Intelligent Music Information Retrieval, ” IEEE Trans. Multimedia, Vol. 8, No. 3, pp. 564－574, 2006.

【発明の開示】

【０００５】

しかしながら、上記の特許文献１および特許文献２に記載の方法においては、画像の特徴のみに基づいた分類方法である。従って、同様の画像を含むシーンであっても、その画像の有する感性を把握して類似するシーンを取得することは困難である。

【０００６】

また、非特許文献１および非特許文献２に記載の方法では、画像のもつ感性を把握して類似するシーンを検索することはできるが、予め各シーンについてメタデータを付与しなければならない。従って、昨今のデータベースの大容量化に伴い、多量の動画データを分類しなければならない場合には、対応することが困難である。

【０００７】

従って本発明の目的は、動画データのクエリシーンに類似するシーンを検索する動画検索装置および動画検索プログラムを提供することである。

【０００８】

上記課題を解決するために、本発明の第１の特徴は、動画データのシーンから、クエリ動画データに類似するシーンを検索する動画検索装置に関する。即ち本発明の第１の特徴に係る動画検索装置は、クエリ動画データを含む動画データが記憶された動画データベース記憶装置と、動画データのビジュアル信号をショットに分割して、該ショットに対応するオーディオ信号の特徴量の差分が小さい連続したショットをシーンとして出力するシーン分割部と、シーン分割部によって分割されたそれぞれのシーンについて、オーディオ信号のベース音の音高の推移に基づく類似度と、ベース音を除く音に基づく類似度を含む、シーン間のオーディオ信号の類似度を算出して、オーディオ信号類似度データを生成するオーディオ信号類似度算出部と、オーディオ信号類似度データに基づいて、クエリ動画データのシーンと、シーン間の類似度が一定の閾値よりも小さいシーンを検索するオーディオ信号類似度検索部と、を備える。

【０００９】

ここで、オーディオ信号類似度検索部によって検索された各シーンについて該類似度に対応する座標を取得して表示するオーディオ信号類似度表示部を更に備えても良い。

【００１０】

シーン分割部によって分割されたそれぞれのシーンについて、ビジュアル信号の特徴量とオーディオ信号の特徴量から、シーン間のビデオ信号の類似度を算出して、ビデオ信号類似度データを生成するビデオ信号類似度算出部と、ビデオ信号類似度データに基づいて、クエリ動画データのシーンと、シーン間の類似度が一定の閾値よりも小さいシーンを検索するビデオ信号類似度検索部を更に備えても良い。

【００１１】

ビデオ信号類似度検索部によって検索された各シーンについて該類似度に対応する座標を取得して表示するビデオ信号類似度表示部を更に備えても良い。

【００１２】

オーディオ信号類似度算出部は、さらに、シーン間のオーディオ信号の類似度として、リズムに基づく類似度を算出して、オーディオ信号類似度データを生成しても良い。この場合、ビデオ信号類似度とオーディオ信号類似度に対する嗜好の割合である嗜好データを取得し、ビデオ信号類似度データおよびオーディオ信号類似度データに基づいて、ビジュアル信号の特徴量とオーディオ信号の特徴量から算出されたシーン間の類似度と、オーディオ信号のベース音に基づく類似度と、ベース音を除く音に基づく類似度と、リズムに基づく類似度とに対する重み係数を決定して、各シーンの各類似度にこの重み係数を乗算して統合された類似度に基づいて、シーン間の統合された類似度が一定の閾値よりも小さいシーンを検索する検索部と、検索部によって検索された各シーンについて該統合された類似度に対応する座標を取得して表示する表示部を更に備えても良い。

【００１５】
本発明の第２の特徴は、動画データのシーンから、クエリ動画データに類似するシーンを検索する動画検索プログラムに関する。即ち本発明の第３の特徴に係る動画検索プログラムは、コンピュータを、動画データベース記憶装置に記憶されたクエリ動画データおよび動画データのビジュアル信号をショットに分割して、該ショットに対応するオーディオ信号の特徴量の差分が小さい連続したショットをシーンとして出力するシーン分割手段と、シーン分割手段によって分割されたそれぞれのシーンについて、オーディオ信号のベース音の音高の推移に基づく類似度と、ベース音を除く音に基づく類似度を含む、シーン間のオーディオ信号の類似度を算出して、オーディオ信号類似度データを生成するオーディオ信号類似度算出手段と、オーディオ信号類似度データに基づいて、クエリ動画データのシーンと、シーン間の類似度が一定の閾値よりも小さいシーンを検索するオーディオ信号類似度検索手段として機能させる。
【００１６】
オーディオ信号類似度検索手段によって検索された各シーンについて該類似度に対応する座標を取得して表示するオーディオ信号類似度表示手段として、更にコンピュータを機能させても良い。
【００１７】
シーン分割手段によって分割されたそれぞれのシーンについて、ビジュアル信号の特徴量とオーディオ信号の特徴量から、シーン間のビデオ信号の類似度を算出して、ビデオ信号類似度データを生成するビデオ信号類似度算出手段と、ビデオ信号類似度データに基づいて、クエリ動画データのシーンと、シーン間の類似度が一定の閾値よりも小さいシーンを検索するビデオ信号類似度検索手段として、更にコンピュータを機能させても良い。
【００１８】
ビデオ信号類似度検索手段によって検索された各シーンについて該類似度に対応する座標を取得して表示するビデオ信号類似度表示手段として、更にコンピュータを機能させても良い。
【００１９】
オーディオ信号類似度算出手段は、さらに、シーン間のオーディオ信号の類似度として、リズムに基づく類似度を算出して、オーディオ信号類似度データを生成しても良い。この場合、ビデオ信号類似度とオーディオ信号類似度に対する嗜好の割合である嗜好データを取得し、ビデオ信号類似度データおよびオーディオ信号類似度データに基づいて、ビジュアル信号の特徴量とオーディオ信号の特徴量から算出されたシーン間の類似度と、オーディオ信号のベース音に基づく類似度と、ベース音を除く音に基づく類似度と、リズムに基づく類似度とに対する重み係数を決定して、各シーンの各類似度にこの重み係数を乗算して統合された類似度に基づいて、シーン間の統合された類似度が一定の閾値よりも小さいシーンを検索する検索手段と、検索手段によって検索された各シーンについて該統合された類似度に対応する座標を取得して表示する表示手段を更に備えても良い。
【００２９】
本発明によれば、動画データのクエリシーンに類似するシーンを検索する動画検索装置および動画検索プログラムを提供することができる。
【図面の簡単な説明】
【００３０】
【図１】図１は、本発明の最良の実施の形態に係る動画検索装置の機能ブロック図である。
【図２】図２は、本発明の最良の実施の形態に係る動画検索装置が出力する画面例であって、クエリ画像を表示した画面例ある。
【図３】図３は、本発明の最良の実施の形態に係る動画検索装置が出力する画面例であって、類似画像を表示した画面例である。
【図４】図４は、本発明の最良の実施の形態に係る動画検索装置のハードウェア構成図である。

【図５】図５は、本発明の最良の実施の形態に係るシーン分割部によるシーン分割処理を説明するフローチャートである。
【図６】図６は、本発明の最良の実施の形態に係るビデオ信号類似度算出部によるビデオ信号類似度算出処理を説明するフローチャートである。
【図７】図７は、本発明の最良の実施の形態に係るオーディオ信号類似度算出部によるオーディオ信号類似度算出処理を説明するフローチャートである。
【図８】図８は、本発明の最良の実施の形態に係るベース音に基づく類似度算出処理を説明するフローチャートである。
【図９】図９は、本発明の最良の実施の形態に係るベース音以外の他楽器に基づく類似度算出処理を説明するフローチャートである。
【図１０】図１０は、本発明の最良の実施の形態に係るリズムに基づく類似度算出処理を説明するフローチャートである。
【図１１】図１１は、本発明の最良の実施の形態に係るビデオ信号類似度検索処理およびビデオ信号類似度表示処理を説明するフローチャートである。
【図１２】図１２は、本発明の最良の実施の形態に係るオーディオ信号類似度検索処理およびオーディオ信号類似度表示処理を説明するフローチャートである。
【図１３】図１３は、本発明の最良の実施の形態に係る動画検索装置において、オーディオクリップのクラス分類を説明する図である。
【図１４】図１４は、本発明の最良の実施の形態に係る動画検索装置において、オーディオクリップのクラス分類の際に参照される信号を説明するテーブルである。
【図１５】図１５は、本発明の最良の実施の形態に係る動画検索装置において、オーディオクリップの特徴量を算出する処理を説明する図である。
【図１６】図１６は、本発明の最良の実施の形態に係る動画検索装置において、オーディオクリップの特徴量の主成分を出力する処理を説明する図である。
【図１７】図１７は、本発明の最良の実施の形態に係る動画検索装置において、オーディオクリップのクラス分類を詳細に説明する図である。
【図１８】図１８は、本発明の最良の実施の形態に係る動画検索装置において、χ^2検定法による映像のショット分割処理を説明する図である。
【図１９】図１９は、本発明の最良の実施の形態に係る動画検索装置において、ファジィ集合を生成する処理を説明する図である。
【図２０】図２０は、本発明の最良の実施の形態に係る動画検索装置において、ファジィ制御規則を説明する図である。
【図２１】図２１は、本発明の最良の実施の形態に係る動画検索装置において、ファジィ制御規則を説明する図である。
【図２２】図２２は、本発明の最良の実施の形態に係る動画検索装置において、ファジィ制御規則を説明する図である。
【図２３】図２３は、本発明の最良の実施の形態に係る動画検索装置において、ビジュアル信号特徴量算出処理を説明するフローチャートである。
【図２４】図２４は、本発明の最良の実施の形態に係る動画検索装置において、オーディオ信号特徴量算出処理を説明するフローチャートである。
【図２５】図２５は、本発明の最良の実施の形態に係る動画検索装置において、３次元ＤＴＷの格子点を説明する図である。
【図２６】図２６は、本発明の最良の実施の形態に係る動画検索装置において、局所パスを説明する図である。
【図２７】図２７は、本発明の最良の実施の形態に係る動画検索装置において、シーン間の類似度算出処理を説明するフローチャートである。
【図２８】図２８は、一般的なＤＴＷによるパターン間の類似度の算出を説明する図である。
【図２９】図２９は、一般的なＤＴＷによる経路長の算出を説明する図である。
【図３０】図３０は、本発明の最良の実施の形態に係る動画検索装置において、ベース音

に基づく類似度算出処理を説明する図である。
【図３１】図３１は、本発明の最良の実施の形態に係る動画検索装置において、ベース音に基づく類似度算出処理を説明するフローチャートである。
【図３２】図３２は、各音名が有する周波数を説明するテーブルである。
【図３３】図３３は、本発明の最良の実施の形態に係る動画検索装置において、音高の推定処理を説明する図である。
【図３４】図３４は、本発明の最良の実施の形態に係る動画検索装置において、ベース音以外の楽器に基づく類似度算出処理を説明する図である。
【図３５】図３５は、本発明の最良の実施の形態に係る動画検索装置において、他楽器に基づく類似度算出処理を説明するフローチャートである。
【図３６】図３６は、本発明の最良の実施の形態に係る動画検索装置において、２分割フィルタバンクによる低周波・高周波成分の算出処理を説明する図である。
【図３７】図３７は、本発明の最良の実施の形態に係る動画検索装置において、２分割フィルタバンクによって算出された低周波・高周波成分を説明する図である。
【図３８】図３８は、本発明の最良の実施の形態に係る動画検索装置において、全波整流を施す前の信号と、全波整流を施した後の信号と、を説明する図である。
【図３９】図３９は、本発明の最良の実施の形態に係る動画検索装置において、低域通過フィルタによって処理される信号を説明する図である。
【図４０】図４０は、本発明の最良の実施の形態に係る動画検索装置において、ダウンサンプリングを説明する図である。
【図４１】図４１は、本発明の最良の実施の形態に係る動画検索装置において、平均値除去処理を説明する図である。
【図４２】図４２は、Ｓｉｎ波形の自己相関を説明する図である。
【図４３】図４３は、本発明の最良の実施の形態に係る動画検索装置において、自己相関関数の算出処理およびＤＴＷを用いたリズム関数の類似度の算出処理を説明するフローチャートである。
【図４４】図４４は、本発明の最良の実施の形態に係る動画検索装置において、透視変換を説明する図である。
【図４５】図４５は、本発明の変形例に係る動画検索装置の機能ブロック図である。
【図４６】図４６は、本発明の変形例に係る動画検索装置が出力する画面例であって、類似画像を表示した画面例である。
【図４７】図４７は、本発明の変形例に係る動画検索装置の嗜好入力部のインタフェースを説明する図である。
【図４８】図４８は、本発明の変形例に係る表示処理を説明するフローチャートである。
【図４９】図４９は、本発明の実施の形態に係る類似画像の検索シミュレーションにおいて、動画検索装置に入力されるクエリ画像データを説明する図である。
【図５０】図５０は、本発明の実施の形態に係る類似画像の検索シミュレーションにおいて、クエリ画像データと、検索対象の動画データとのシーン毎の類似度を示したグラフである。
【図５１】図５１は、本発明の実施の形態に係る類似画像の検索シミュレーションにおいて、クエリ画像データに類似するシーンとの類似度を示す３次元ＤＴＷのパスを示した図である。
【図５２】図５２は、本発明の実施の形態に係るビデオ信号に基づく類似画像の検索のシミュレーションにおいて、動画検索装置に入力されるクエリ画像データを説明する図である。
【図５３】図５３は、本発明の実施の形態に係るビデオ信号に基づく類似画像の検索のシミュレーションにおいて、動画検索装置に入力される検索対象の画像データを説明する図である。
【図５４】図５４は、本発明の実施の形態に係るビデオ信号に基づく類似画像の検索シミュレーションにおいて、クエリ画像データと、検索対象の動画データとのシーン毎の類似

度を示したグラフである。
【図５５】図５５は、本発明の実施の形態に係るビデオ信号に基づく類似画像の検索シミュレーションにおいて、クエリ画像データに類似するシーンとの類似度を示す３次元ＤＴＷのパスを示した図である。
【図５６】図５６は、本発明の実施の形態に係るオーディオ信号に基づく類似画像の検索のシミュレーションにおいて、動画検索装置に入力されるクエリ画像データを説明する図である。
【図５７】図５７は、本発明の実施の形態に係るオーディオ信号に基づく類似画像の検索のシミュレーションにおいて、動画検索装置に入力される検索対象の画像データを説明する図である。
【図５８】図５８は、本発明の実施の形態に係るオーディオ信号に基づく類似画像の検索シミュレーションにおいて、クエリ画像データと、検索対象の動画データとのシーン毎の類似度を示したグラフである。
【図５９】図５９は、本発明の実施の形態に係るオーディオ信号に基づく類似画像の検索シミュレーションにおいて、クエリ画像データに類似するシーンとの類似度を示す３次元ＤＴＷのパスを示した図である。
【発明を実施するための最良の形態】
【００３１】
次に、図面を参照して、本発明の実施の形態を説明する。以下の図面の記載において、同一又は類似の部分には同一又は類似の符号を付している。
【００３２】
本発明の最良の実施の形態において、「ショット」とは、カメラ切り換えから、次のカメラ切り換えまでの間の連続する画像フレーム列である。ＣＧアニメーションや合成映像についても、カメラを撮影環境の設定に置き換えて、同様の意味で使用される。ここで、ショット間の不連続点を「カット点」と呼ぶ。「シーン」とは、意味を持つ連続したショットの集まりである。「クリップ」とは、ビデオ信号を、所定のクリップ長で分割した信号である。このクリップには、複数のフレームが含まれることが好ましい。「フレーム」とは、動画像データを構成する静止画像データである。
【００３３】
（最良の実施の形態）
図１に示す本発明の最良の実施の形態に係る動画検索装置１は、動画データのシーンから、クエリ動画データに類似するシーンを検索する。本発明の最良の実施の形態に係る動画検索装置１は、動画データベース１１中に存在する動画データをシーンに分類して、クエリ動画データと各シーンとの類似度を算出し、クエリ動画データに類似するシーンの検索を行う。
【００３４】
より具体的には、本発明の最良の実施の形態において、メタデータを用いることなく、映像の構成要素である音響・ビジュアル信号の解析結果を用いて映像間の類似度を算出し、類似映像の検索を行うシステムを説明する。また、それらの検索・分類結果を３次元の空間上に可視化するシステムを説明する。本発明の最良の実施の形態では映像に対して、オーディオ信号およびビジュアル信号を含むビデオ信号に基づいた映像情報の類似度と、オーディオ信号に基づいた音楽情報の類似度の算出の２つの類似度算出機能を持つ。さらに、この機能を用いることで、クエリ映像を与えた場合に自動で類似映像の検索を可能とする。また、クエリ映像が存在しない場合、データベース中の映像の自動分類を行い、注目する映像に対して類似する映像をユーザに呈示することを可能とする。このとき、本発明の最良の実施の形態では、映像間の類似度に基づいて、３次元の空間上に映像を配置することで、空間の距離によって映像の類似性を理解することが可能なユーザインターフェースを実現している。
【００３５】
図１に示す本発明の最良の実施の形態に係る動画検索装置１は、動画データベース１１

から複数の映像を読み込み、シーン分割部２１において、全ての映像に対して、同一の内容を含む区間であるシーンの算出を行う。さらに、分類部２２において、得られる全てのシーン間で類似度の算出を行い、検索部２５でクエリ画像と類似度の高い動画像データを抽出し、表示部２８において、類似したシーンを持つ映像同士が近くなるように３次元空間へ映像を配置する。尚、クエリの映像が与えられた場合は、これを中心に処理が行われる。ここで本発明の最良の実施の形態に係る動画検索装置１の分類部２２において、（１）「映像情報に注目した検索・分類」に基づくビデオ信号類似度算出部２３と、（２）「音楽情報に注目した検索・分類」に基づくオーディオ信号類似度算出部２４の２つに分岐し、それぞれにおいて異なるアルゴリズムを用いて類似度が算出される。

【００３６】

本発明の最良の実施の形態において、動画検索装置１は、図２および図３に示す表示画面Ｐ１０１および表示画面Ｐ１０２を、表示装置に表示する。表示画面Ｐ１０１は、クエリ画像表示部Ａ１０１を備えている。動画検索装置１は、クエリ画像表示部Ａ１０１に表示された動画に類似するシーンを、動画データベース１１から検索して、表示画面Ｐ１０２を表示装置に表示する。表示画面Ｐ１０２には、類似画像表示部Ａ１０２ａおよびＡ１０２ｂを備えている。これらの類似画像表示部Ａ１０２ａおよびＡ１０２ｂには、動画データベース１１から検索された動画データのシーンであって、クエリ画像表示部Ａ１０１に表示されたシーンに類似するシーンが表示されている。

【００３７】

（動画検索装置のハードウェア構成）

図４に示すように、本発明の最良の実施の形態に係る動画検索装置１は、中央処理制御装置１０１、ＲＯＭ（Read Only Memory）１０２、ＲＡＭ（Random Access Memory）１０３及び入出力インタフェース１０９が、バス１１０を介して接続されている。入出力インタフェース１０９には、入力装置１０４、表示装置１０５、通信制御装置１０６、記憶装置１０７及びリムーバブルディスク１０８が接続されている。

【００３８】

中央処理制御装置１０１は、入力装置１０４からの入力信号に基づいてＲＯＭ１０２から動画検索装置１を起動するためのブートプログラムを読み出して実行し、更に記憶装置１０７に記憶されたオペレーティングシステムを読み出す。更に中央処理制御装置１０１は、入力装置１０４や通信制御装置１０６などの入力信号に基づいて、各種装置の制御を行ったり、ＲＡＭ１０３や記憶装置１０７などに記憶されたプログラム及びデータを読み出してＲＡＭ１０３にロードするとともに、ＲＡＭ１０３から読み出されたプログラムのコマンドに基づいて、データの計算又は加工など、後述する一連の処理を実現する処理装置である。

【００３９】

入力装置１０４は、操作者が各種の操作を入力するキーボード、マウスなどの入力デバイスにより構成されており、操作者の操作に基づいて入力信号を作成し、入出力インタフェース１０９及びバス１１０を介して中央処理制御装置１０１に送信される。表示装置１０５は、ＣＲＴ（Cathode Ray Tube）ディスプレイや液晶ディスプレイなどであり、中央処理制御装置１０１からバス１１０及び入出力インタフェース１０９を介して表示装置１０５において表示させる出力信号を受信し、例えば中央処理制御装置１０１の処理結果などを表示する装置である。通信制御装置１０６は、ＬＡＮカードやモデムなどの装置であり、動画検索装置１をインターネットやＬＡＮなどの通信ネットワークに接続する装置である。通信制御装置１０６を介して通信ネットワークと送受信したデータは入力信号又は出力信号として、入出力インタフェース１０９及びバス１１０を介して中央処理制御装置１０１に送受信される。

【００４０】

記憶装置１０７は半導体記憶装置や磁気ディスク装置であって、中央処理制御装置１０１で実行されるプログラムやデータが記憶されている。リムーバブルディスク１０８は、光ディスクやフレキシブルディスクのことであり、ディスクドライブによって読み書きさ

れた信号は、入出力インタフェース１０９及びバス１１０を介して中央処理制御装置１０１に送受信される。

【００４１】

本発明の最良の実施の形態に係る動画検索装置１の記憶装置１０７には、図１に示すように、動画検索プログラムが記憶されるとともに、動画データベース１１、ビデオ信号類似度データ１２およびオーディオ信号類似度データ１３が記憶される。又、動画検索プログラムが動画検索装置１の中央処理制御装置１０１に読み込まれ実行されることによって、シーン分割部２１、分類部２２、検索部２５および表示部２８が、動画検索装置１に実装される。

【００４２】

（動画検索装置の機能ブロック）

動画データベース１１は、複数の動画データが記憶される。この動画データベース１１に記憶される動画データは、本発明の最良の実施の形態に係る動画検索装置１によって分類される対象となる。動画データベース１１に記憶される動画データは、オーディオ信号およびビジュアル信号を含むビデオ信号によって構成されている。

【００４３】

シーン分割部２１は、記憶装置１０７から動画データベース１１を読み出して、動画データのビジュアル信号をショットに分割して、ショットに対応するオーディオ信号の特徴量の差分が小さい連続したショットをシーンとして出力する。より具体的には、シーン分割部２１は、動画データのオーディオ信号から、各クリップの特徴量データを算出して、各クリップの音響の種類を表す各オーディオクラスへの帰属確率を算出する。さらにシーン分割部２１は、動画データのビジュアル信号をショットに分割して、該ショットに対応する複数のクリップの各オーディオクラスへの帰属確率から、各ショットのファジィ推論値を算出する。さらにシーン分割部２１は、隣接するショット間におけるファジィ推論値の差分が小さい連続したショットをシーンとして出力する。

【００４４】

図５を参照して、シーン分割部２１の処理の概要を説明する。まず、動画データベース１１を読み出して、動画データベース１１に記憶された各動画データについて、ステップＳ１０１ないしステップＳ１１０の処理を繰り返す。

ステップＳ１０１において、動画データベース１１に記憶された動画データの一つについて、オーディオ信号を抽出して読み出し、ステップＳ１０２において、オーディオ信号をクリップに分割する。次に、ステップＳ１０２で分割された各クリップについて、ステップＳ１０３ないしステップＳ１０５の処理を繰り返す。

【００４５】

ステップＳ１０３において、クリップの特徴量が算出され、ステップＳ１０４において、ＰＣＡ（主成分分析）によってこの特徴量のパラメータが削減される。次に、ステップＳ１０４において削減された後の特徴量に基づいて、ステップＳ１０５において、ＭＧＤに基づいて、クリップのオーディオクラスの帰属確率が算出される。ここでオーディオクラスは、無音、音声、音楽等のオーディオ信号の種類を表すクラスである。

【００４６】

ステップＳ１０３ないしステップＳ１０５において、オーディオ信号の各クリップについて、オーディオクラスの帰属確率が算出されると、ステップＳ１０６において、ステップＳ１０１で取得したオーディオ信号に対応するビジュアル信号を抽出して読み出し、ステップＳ１０７において、カイ二乗検定法に基づいて、映像データをショットに分割する。このカイ二乗検定法においては、音声信号ではなく、ビジュアル信号の色ヒストグラムが用いられる。ステップＳ１０７において、動画データが複数のショットに分割されると、各ショットについて、ステップＳ１０８およびステップＳ１０９の処理を繰り返す。

【００４７】

ステップＳ１０８において、各ショットに対するオーディオクラスへの帰属確率が算出される。このとき、ショットに対応するクリップについて、ステップＳ１０５で算出され

たオーディオクラスへの帰属確率が取得される。各クリップのオーディオクラスへの帰属確率の平均値が、ショットに対するオーディオクラスへの帰属確率として算出される。さらにステップＳ１０９において、各ショットに対するファジィ推論により、各ショットクラスの出力変数およびメンバシップ関数の値が算出される。

【００４８】

ステップＳ１０７で分割された全てのショットについて、ステップＳ１０８およびステップＳ１０９の処理が実行されると、ステップＳ１１０において、ファジィ推論による各ショットクラスの出力変数およびメンバシップ関数の値に基づいて、各ショットを連結して、動画データをシーンに分割する。

【００４９】

分類部２２は、ビデオ信号類似度算出部２３とオーディオ信号類似度算出部２４を備えている。

【００５０】

ビデオ信号類似度算出部２３は、シーン分割部２１によって分割されたそれぞれのシーンについて、ビジュアル信号の特徴量とオーディオ信号の特徴量から、シーン間のビデオ信号の類似度を算出して、ビデオ信号類似度データ１２を生成する。ここでシーン間の類似度は、あるシーンと他のシーンとのビジュアル信号の類似度である。例えば、動画データベース１１にｎ個のシーンが格納されているとすると、第１のシーンについて、第２のシーンとのビジュアル信号の類似度、第３のシーンとのビジュアル信号の類似度・・・第ｎのシーンとのビジュアル信号の類似度が算出される。より具体的には、ビデオ信号類似度算出部２３は、シーン分割部２１によって分割されたそれぞれのシーンについて、シーンをクリップに分割し、各クリップのビジュアル信号から、各クリップの動画像の所定のフレームの色ヒストグラムに基づいて、ビジュアル信号の特徴量を算出する。さらにビデオ信号類似度算出部２３は、クリップをオーディオ信号のフレームに分割し、各フレームのオーディオ信号が持つエネルギーとスペクトルに基づいて、各オーディオ信号のフレームを音声フレームと背景音フレームに分類して、オーディオ信号の特徴量を算出する。さらにビデオ信号類似度算出部２３は、クリップ単位のビジュアル信号とオーディオ信号の特徴量に基づいて、シーン間の類似度を算出して、ビデオ信号類似度データ１２として、記憶装置１０７に記憶する。

【００５１】

図６を参照して、ビデオ信号類似度算出部２３の処理の概要を説明する。

シーン分割部２１によって分割された各動画データの各シーンについて、ステップＳ２０１ないしステップＳ２０３の処理が繰り返される。まず、ステップＳ２０１において、シーンに対応するビデオ信号がクリップに分割される。つぎに、ステップＳ２０１で分割された各クリップについて、ステップＳ２０２において、ビジュアル信号の特徴量が算出され、ステップＳ２０３において、オーディオ信号の特徴量が算出される。

【００５２】

各動画データの各シーンについて、ビジュアル信号の特徴量およびオーディオ信号の特徴量が算出されると、ステップＳ２０４において、シーン間の類似度が算出される。さらにステップＳ２０５において、ステップＳ２０４においてシーンの類似度を、シーン間の映像情報の類似度であるビデオ信号類似度データ１２として、記憶装置１０７に記憶する。

【００５３】

オーディオ信号類似度算出部２４は、シーン分割部２１によって分割されたそれぞれのシーンについて、オーディオ信号のベース音に基づく類似度と、ベースを除く楽器に基づく類似度と、リズムに基づく類似度を含む、シーン間のオーディオ信号の類似度を算出して、オーディオ信号類似度データ１３を生成する。ここで類似度は、あるシーンと他のシーンとの、ベース音、ベースを除く楽器、リズムのそれぞれに基づく類似度である。例えば、動画データベース１１にｎ個のシーンが格納されているとすると、第１のシーンについて、第２のシーンとのベース音、ベースを除く楽器、リズムのそれぞれに基づく類似度

、第３のシーンとのベース音、ベースを除く楽器、リズムのそれぞれに基づく類似度・・・第ｎのシーンとのベース音、ベースを除く楽器、リズムのそれぞれに基づく類似度が算出される。より具体的には、オーディオ信号類似度算出部２４は、ベース音に基づく類似度を算出する際、オーディオ信号からベース音を取得し、時間および周波数に着目したパワースペクトルを算出して、任意の２シーンについて、ベース音に基づく類似度を算出する。また、オーディオ信号類似度算出部２４は、ベース音を除く楽器に基づく類似度を算出する際、オーディオ信号からベース音より高い周波数域を有する音について、各音名が示す周波数のエネルギーを算出し、任意の２シーンについて、エネルギーの差分の合計を算出して、ベースを除く楽器に基づく類似度を算出する。また、オーディオ信号類似度算出部２４は、リズムに基づく類似度を算出する際、２分割フィルタバンクを用いてオーディオ信号の高周波成分と低周波成分の分割を所定回数繰り返し、高周波成分を含む信号から包絡線を検波して自己相関関数を算出し、この自己相関関数に基づいて、任意の２シーンについてリズムに基づく類似度を算出する。

【００５４】

図７を参照して、オーディオ信号類似度算出部２４の処理の概要を説明する。

シーン分割部２１によって全ての動画データから分割され、得られる全てのシーンのうち、任意の２つのシーンについて、ステップＳ３０１ないしステップＳ３０３の処理が繰り返される。まず、ステップＳ３０１において、シーンに対応するオーディオ信号のベース音に基づく類似度が算出される。つぎに、ステップＳ３０２において、オーディオ信号の、ベース音以外の楽器に基づく類似度が算出される。さらに、ステップＳ３０３において、オーディオ信号のリズムに基づく類似度が算出される。

【００５５】

つぎに、ステップＳ３０４において、ステップＳ３０１ないしステップＳ３０３において算出したベース音、ベースを除く楽器、リズムのそれぞれに基づく類似度が、シーン間の音響情報の類似度であるオーディオ信号類似度データ１３として、記憶装置１０７に記憶される。

【００５６】

次に、図８を参照して、図７のステップＳ３０１におけるベース音に基づく類似度算出処理の概要を説明する。まず、ステップＳ３１１において、所定の帯域通過フィルタを介して、ベース音が抽出される。ここで所定の帯域とは、ベース音に対応する帯域であって、例えば４０Ｈｚないし２５０Ｈｚである。

つぎに、ステップＳ３１２において、時間および周波数に注目して、重み付きパワースペクトルを算出し、ステップＳ３１３において、重み付きパワースペクトルを用いてベースの音高が推定される。さらに、ステップＳ３１４において、ＤＴＷを用いて、ベース音高の類似度が算出される。

【００５７】

図９を参照して、図７のステップＳ３０２におけるベース以外の楽器に基づく類似度算出処理の概要を説明する。まず、ステップＳ３２１において、音名が示す周波数のエネルギーが算出される。ここでは、ベース音より高く、かつ音名を持つ周波数のエネルギーについて、各音名が示す周波数のエネルギーが算出される。

つぎに、ステップＳ３２２において、各音名が示す周波数のエネルギーについて、全周波数域に対するエネルギーの割合が算出される。さらにステップＳ３２３において、ＤＴＷを用いて、音名のエネルギー割合の類似度が算出される。

【００５８】

図１０を参照して、図７のステップＳ３０３におけるリズムに基づく類似度算出処理の概要を説明する。まず、ステップＳ３３１において、２分割フィルタバンクによって、所定回数の分割を繰り返すことにより、低周波成分および高周波成分が算出される。これにより、複数種類の楽器音によるリズムを推定することができる。

さらに、ステップＳ３３２ないしステップＳ３３５の処理によって、包絡線を検波して、各信号の概形が取得される。具体的には、ステップＳ３３２において、ステップＳ３３

１で取得した波形について全波整流が施され、ステップＳ３３３において、低域通過フィルタが施される。さらにステップＳ３３４において、ダウンサンプリングされ、ステップＳ３３５において、平均値が除去される。

包絡線の検波が終了すると、ステップＳ３３６において、自己相関関数が算出され、ステップＳ３３７において、ＤＴＷを用いて、リズム関数の類似度が算出される。

【００５９】

検索部２５は、ビデオ信号類似度検索部２６と、オーディオ信号類似度検索部２７を備える。表示部２８は、ビデオ信号類似度表示部２９と、オーディオ信号類似度表示部３０を備える。

【００６０】

ビデオ信号類似度検索部２６は、ビデオ信号類似度データ１２に基づいて、シーン間の類似度が一定の閾値よりも小さいシーンを検索する。ビデオ信号類似度表示部２９は、ビデオ信号類似度検索部２６によって検索された各シーンについて該類似度に対応する座標を取得して表示する。

【００６１】

図１１を参照して、ビデオ信号類似度検索部２６およびビデオ信号類似度表示部２９の処理を説明する。

図１１（ａ）を参照して、ビデオ信号類似度検索部２６の処理を説明する。まず、記憶装置１０７からビデオ信号類似度データ１２が読み出される。さらに、シーン分割部２１によって分割された各シーンについて、ステップＳ４０１においてクエリ動画シーンとのビジュアル信号の類似度が取得されるとともに、ステップＳ４０２においてクエリ動画シーンとのオーディオ信号の類似度が取得される。

【００６２】

つぎにステップＳ４０３において、ステップＳ４０１およびステップＳ４０２で取得された類似度のうち、所定値以上の類似度のシーンを検索する。ここでは、類似度に基づいて閾値処理する場合について説明するが、類似度が高いものから所定数のシーンが検索されても良い。

【００６３】

図１１（ｂ）を参照して、ビデオ信号類似度表示部２９の処理を説明する。ステップＳ４５１において、ビデオ信号類似度検索部２６によって検索された各シーンについて、三次元空間における座標が算出される。ここで三次元空間における軸は、３次元ＤＴＷによって得られる３つの座標になる。ステップＳ４５２において、ステップＳ４５１で算出された各シーンの座標が透視変換され、各シーンの動画像フレームのサイズが決定される。ステップＳ４５３において、表示装置に表示される。

【００６４】

オーディオ信号類似度検索部２７は、オーディオ信号類似度データ１３に基づいて、オーディオ信号の類似度が一定の閾値よりも小さいシーンを検索する。オーディオ信号類似度表示部３０は、オーディオ信号類似度検索部２７によって検索された各シーンについて類似度に対応する座標を取得して表示する。

【００６５】

図１２を参照して、オーディオ信号類似度検索部２７およびオーディオ信号類似度表示部３０の処理を説明する。

図１２（ａ）を参照して、オーディオ信号類似度検索部27の処理を説明する。まず、記憶装置１０７からオーディオ信号類似度データ１３が読み出される。さらに、シーン分割部２１によって分割された各シーンについて、ステップＳ５０１においてクエリ動画シーンとのベース音に基づく類似度が取得される。ステップＳ５０２においてクエリ動画シーンとの非ベース音に基づく類似度が取得される。ステップＳ５０３においてクエリ動画シーンとのリズムに基づく類似度が取得される。

【００６６】

つぎにステップＳ５０４において、ステップＳ５０１ないしステップＳ５０３で取得さ

れた類似度のうち、所定値以上の類似度のシーンを検索する。ここでは、類似度に基づいて閾値処理する場合について説明するが、類似度が高いものから所定数のシーンが検索されても良い。

【００６７】

図１２（ｂ）を参照して、オーディオ信号類似度表示部３０の処理を説明する。ステップＳ５５１において、オーディオ信号類似度検索部２７によって検索された各シーンについて、三次元空間における座標が算出される。ここで三次元空間における軸は、ベース音に基づく類似度、ベース以外の楽器に基づく類似度およびリズムに基づく類似度である。ステップＳ５５２において、ステップＳ５５１で算出された各シーンの座標が透視変換され、各シーンの動画像フレームのサイズが決定される。ステップＳ５５３において、表示装置に表示される。

以下、図１に示す各ブロックについて詳述する。

【００６８】

（シーン分割部）

次に、図１に示すシーン分割部２１の処理を説明する。

シーン分割部２１は、データベース中に存在する映像間で類似度を算出するために、映像信号をシーン単位に分割する。本発明の最良の実施の形態では、動画データベース１１から得られる映像信号のオーディオ信号と動画像フレームの両方を用いることで、シーンの算出を可能とする。

【００６９】

シーン分割部２１は、まずオーディオ信号をクリップと呼ばれる小区間毎に分け、各々に対して特徴量の算出を行い、さらにＰＣＡ(主成分分析)による特徴量の削減を行う。次に、オーディオ信号の種類を表すオーディオクラス(無音、音声、音楽等)を準備し、各クリップがそれらのクラスに属する確率、つまり帰属確率をＭＧＤにより求める。さらに、本発明の最良の実施の形態では、映像中のビジュアル信号(フレーム)に対し、χ^2検定を用いることで、１台のカメラで連続的に撮影された区間であるショットの分割を行う。また、各ショットに含まれるオーディオ信号のクリップについて、オーディオクラスへの帰属確率の平均を求めることで、ショットとしてのオーディオクラスへの帰属確率が得られる。本発明の最良の実施の形態では、得られる帰属確率から各ショットに対してファジィ推論を行うことで、ショットの種類を表すショットクラスのファジィ推論値を算出する。最後に、隣接する全てのショット間において、ファジィ推論値の差分を求め、その値が小さな連続区間を１つのシーンとして求める。

【００７０】

このように、処理対象であるショットが各ショットクラスに属する度合い(ファジィ推論値)が得られる。オーディオ信号の種類によっては、ユーザの主観評価により、ショットの分類結果が異なる可能性がある。例えば、音楽の付加された音声において、背景に存在する音楽が非常に小さな音量である場合、そのオーディオ信号を「音楽付きの音声」に分類すべきか、それとも主となる「音声」に分類すべきかは、ユーザの要求によって異なる。そこで、ショットに対して、全てのショットクラスタのファジィ推論値を持たせ、最終的にその差分を求めることで、ユーザの主観評価を考慮したシーンの分割が可能となる。

【００７１】

ここで、本発明の最良の実施の形態に係るシーン分割部２１では、処理対象信号をオーディオクラスに分類する。ここで、オーディオ信号には音楽や音声などの単一のオーディオクラスから構成されるものの他に、背景に音楽が存在する環境下での音声(音楽付き音声)や、背景に雑音が存在する環境下での音声(雑音付き音声)等、複数の種類のオーディオクラスから構成されるものも数多く存在し、このようなオーディオ信号では、どのオーディオクラスに分類されるかの境界を定めることが困難である。そこで、本発明の最良の実施の形態ではファジィ推論による推論値を用いることにより、処理対象信号が各オーディオクラスに属する度合いを高精度に算出し、分類を行う。

【００７２】
本発明の最良の実施の形態に係るシーン分割部２１について、具体的なアルゴリズムを説明する。
本発明の最良の実施の形態では、まずＰＣＡとＭＧＤを用いて、オーディオ信号が以下に定義する４種類のオーディオクラスに属する程度（以降、帰属確率）を算出する。
・ 無音(silence: Si)
・ 音声(speech: Sp)
・ 音楽(music: Mu)
・ 雑音(noise: No)
各オーディオクラスへの帰属確率は、図１３に示す「ＣＬＳ＃１」から「ＣＬＳ＃３」の３つの分類処理を施し、それらの分類結果を用いて算出される。ここで、ＣＬＳ＃１からＣＬＳ＃３までの各分類処理は、全て同一の手順であり、処理対象信号および２種類の参照信号に対し、「特徴量の算出」、「ＰＣＡの適用」、及び「ＭＧＤの算出」の３つの処理を行う。ただし、図１４に示すように、参照信号は分類処理の目的に応じてＳｉ、Ｓｐ、Ｍｕ、Ｎｏのいずれか（あるいは複数）のオーディオ信号を含む。以下、各処理について説明する。
【００７３】
まず、オーディオ信号クリップの特徴量算出処理を説明する。この処理は、図５のステップＳ１０３に相当する。
シーン分割部２１は、処理対象であるオーディオ信号、および図１４に示した２種類の参照信号から、以下に示すオーディオ信号のフレーム単位（フレーム長：W_f）の特徴量、およびクリップ単位（クリップ長：W_c，ただし$W_c > W_f$）の特徴量を算出する。
－ フレーム単位の特徴量：
ボリューム、零交差率、ピッチ、周波数中心位置、周波数帯域幅、サブバンドエネルギー比率
－ クリップ単位の特徴量：
非無音率、零比率
さらに、シーン分割部２１は、オーディオ信号のフレーム単位の特徴量のクリップ内での平均値および標準偏差を算出し、それらをクリップ単位の特徴量に加える。
【００７４】
この処理を図１５を参照して説明する。
まず、ステップＳ１１０１において、１クリップのオーディオ信号について、オーディオ信号のフレームに分割する。つぎに、ステップＳ１１０１で分割した各オーディオ信号のフレームについて、ステップＳ１１０２ないしステップＳ１１０７において、ボリューム、零交差率、ピッチ、周波数中心位置、周波数帯域幅、サブバンドエネルギー比率を算出する。つぎに、ステップＳ１１０８において、１クリップに含まれる各オーディオ信号のフレームのボリューム、零交差率、ピッチ、周波数中心位置、周波数帯域幅、サブバンドエネルギー比率の各特徴量に対する平均値と標準偏差を算出する。
一方、ステップＳ１１０９において、１クリップのオーディオ信号について、非無音率を算出し、ステップＳ１１１０において、零比率を算出する。
ステップＳ１１１１において、ステップＳ１１０８ないしステップＳ１１１０において算出した平均値、標準偏差、非無音率および零比率の各特徴量を統合して、クリップにおけるオーディオ信号の特徴量として出力する。
【００７５】
つぎに、ＰＣＡによる特徴量削減処理を説明する。この処理は、図５のステップＳ１０４に相当する。
シーン分割部２１は、処理対象信号のクリップから算出された特徴量、および２種類の参照信号から算出されたクリップ単位の特徴量を正規化し、ＰＣＡを施す。ＰＣＡを施すことで、相関の高い特徴量間の影響を軽減することが可能となる。また、ＰＣＡより得られた主成分のうち、その固有値が１以上であるものを以降の処理で使用することで、計算

量の増加やヒューズの現象を回避することが可能となる。
ここで用いられる参照信号は、分類されるクラスに応じて異なる。例えば、図１３に示す「ＣＬＳ＃１」においては、Ｓｉ＋Ｎｏと、Ｓｐ＋Ｍｕとに分類される。このとき用いられる２種類の参照信号の一つは、無音（Ｓｉ）のみで構成される信号と、雑音（Ｎｏ）のみで構成される信号をと、重ならないように時間軸方向に連結した信号である。もう一つの参照信号は、音声（Ｓｐ）のみで構成される信号と、音楽（Ｍｕ）のみで構成される信号をと、重ならないように時間軸方向に連結した信号である。また、「ＣＬＳ＃２」において用いられる２種類の参照信号は、無音（Ｓｉ）のみで構成される信号と、雑音（Ｎｏ）のみで構成される信号である。同様に、「ＣＬＳ＃３」において用いられる２種類の参照信号は、音声（Ｓｐ）のみで構成される信号と、音楽（Ｍｕ）のみで構成される信号である。

【００７６】
ここで、主成分分析（ＰＣＡ）は複数の変数間の共分散（相関）を少数の合成変数で表わす手法である。共分散行列の固有値問題の解として得ることができる。本発明の最良の実施の形態では、処理対象信号から得られた特徴量に対し主成分分析を施すことで、相関の高い特徴量間の影響を軽減している。また、得られた主成分のうち、その固有値が1以上であるものを選択して用いる事で計算量の増加やヒューズの現象を回避している。

【００７７】
この処理を図１６を参照して説明する。図１６（ａ）は、処理対象信号のクリップの主成分を出力する処理で、図１６（ｂ）は、参照信号１および参照信号２のクリップの主成分を出力する処理である。
図１６（ａ）に示す処理を説明する。まず、ステップＳ１２０１において、図１５を参照して説明した処理に従って算出された処理対象信号のクリップの特徴量が入力される。
つぎに、ステップＳ１２０４において、クリップ単位の特徴量を正規化し、ステップＳ１２０５において、ＰＣＡ（主成分分析）を施す。さらにステップＳ１２０６において、固定値が１以上となる主成分の軸を算出し、処理対象信号のクリップの主成分を出力する。
図１６（ｂ）に示す処理を説明する。まず、ステップＳ１２５１において、参照信号１のクリップから算出される特徴量を入力するとともに、ステップ１２５２において、参照信号２のクリップから算出される特徴量を入力する。
つぎに、ステップＳ１２５３において、参照信号１および参照信号２のそれぞれについて、クリップ単位の特徴量を正規化し、ステップＳ１２５４において、ＰＣＡ（主成分分析）を施す。さらにステップＳ１２５５において、固定値が１以上となる主成分の軸を算出し、参照信号１および参照信号２について、一つの主成分を出力する。
ここで入力される参照信号１および参照信号２は、上述したように、クラスの分類処理によって異なる。後述するＣＬＳ＃１～３ごとに、の各分類処理において用いられる全ての参照信号１および参照信号２について、予め図１６（ｂ）の処理が実行される。

【００７８】
次に、ＭＧＤによるクリップのオーディオクラスへの帰属確率の算出処理を説明する。この処理は、図５のステップＳ１０５に相当する。
ＰＣＡによる特徴量削減処理で得られた主成分を用いて、ＭＧＤを算出する。
ここで、ＭＧＤ（マハラノビス汎距離）は、多変数間の相関に基づき算出される距離である。ＭＧＤでは、処理対象信号と参照信号との特徴ベクトル群との距離をマハラノビス汎距離により算出する。これにより、主成分分析で得られた主成分の分布形状を考慮した距離を算出することが可能となる。

【００７９】
まず、処理対象信号において、ＰＣＡによる特徴量削減処理で得られた主成分を要素とする特徴ベクトルｆ（ｃ）（ｃ＝１，・・・，３；ＣＬＳ＃１～３に対応）と、同様にして算出される２種類の参照信号の特徴ベクトル群との

【数１】

$$\mathrm{MGD}\ d_i^{(c)}\ (i=1,2;\text{参照信号}1,2\text{に対応})$$

を、次式により算出する。

【数２】

$$d_i^{(c)} = \left(f^{(c)} - m_i^{(c)}\right)^{\mathrm{T}} S_i^{(c)^{-1}} \left(f^{(c)} - m_i^{(c)}\right) \qquad \text{（式１－１）}$$

ただし、

【数３】

$$m_i^{(c)}\ \text{および}\ S_i^{(c)}$$

は、それぞれ参照信号ｉから算出された特徴ベクトルの平均ベクトル、および共分散行列を表す。この

【数４】

$$\mathrm{MGD}\ d_i^{(c)}$$

は、固有空間における主成分の分布形状を考慮した距離尺度となる。そこで、この

【数５】

$$\mathbf{MGD}\ d_i^{(c)}$$

を用いて、処理対象信号が参照信号１、２と同一のクラスタに属する帰属度

【数６】

$$D_i^{(c)}$$

を次式で定義する。

【数７】

$$D_i^{(c)} = 1 - \frac{d_i^{(c)}}{d_1^{(c)} + d_2^{(c)}} \qquad \text{（式１－２）}$$

ＣＬＳ＃１～３の各分類処理において、上記３つの処理を行うことで、帰属度

【数８】

$$D_i^{(c)}\ (i=1,2;\ \ c=1,\cdots,3)$$

が得られる。そこで、各オーディオクラス（Si，Sp，Mu，No）への帰属確率

【数９】

$$P_{l_1}\ (l_1=1,\cdots,4\ ;\text{それぞれ Si, Sp, Mu, No に対応})$$

を、以下で定義する。

【数１０】

$$P_1 = D_1^{(1)} D_1^{(2)} \qquad (式１－３)$$

$$P_2 = D_2^{(1)} D_1^{(3)} \qquad (式１－４)$$

$$P_3 = D_2^{(1)} D_2^{(3)} \qquad (式１－５)$$

$$P_4 = D_1^{(1)} D_2^{(2)} \qquad (式１－６)$$

上式は、ＣＬＳ＃１からＣＬＳ＃３の各分類処理において、

【数１１】

$$D_i^{(c)}$$

を、参照信号１、２と同一のクラスタに分類される確率とみなし、それらを積算することで、Ｓｉ、Ｓｐ、Ｍｕ、Ｎｏのオーディオクラスに属する確率を算出することを表す。従って、この帰属確率

【数１２】

$$P_{l_1}\ (l_1=1,\cdots,4)$$

から、処理対象であるオーディオ信号がどのオーディオクラスにどの程度属しているかを知ることが可能となる。

【００８０】

この処理を図１７を参照して説明する。この処理は、処理対象信号の各クリップに対して実行される。

まず、ステップＳ１３０１において、処理対象信号の各クリップの主成分を要素とするベクトルを入力する。ここで入力されるベクトルは、上述した図１６（ａ）によって算出されたデータである。

次に、ＣＬＳ＃１の分類処理として、ステップＳ１３０２ないしステップＳ１３０５の処理を行う。具体的には、ステップＳ１３０２において、処理対象信号と参照信号１との距離を算出し、ステップＳ１３０３において、処理対象信号が参照信号１のクラスタに属する帰属度を算出する。さらに、ステップＳ１３０４において、処理対象信号と参照信号２との距離を算出し、ステップＳ１３０５において、処理対象信号が参照信号２のクラスタに属する帰属度を算出する。

【００８１】

さらに、ＣＬＳ＃２の分類処理として、ステップＳ１３０６ないしステップＳ１３０９の処理を行う。具体的には、ステップＳ１３０６において、処理対象信号と参照信号１との距離を算出し、ステップＳ１３０７において、処理対象信号が参照信号１のクラスタに属する帰属度を算出する。さらに、ステップＳ１３０８において、処理対象信号と参照信号２との距離を算出し、ステップＳ１３０９において、処理対象信号が参照信号２のクラ

スタに属する帰属度を算出する。
ここで、ステップＳ１３１０において、ステップＳ１３０３およびステップＳ１３０７で算出された帰属度に基づいて、オーディオクラスＳｉへの帰属確率$Ｐ_1$が算出される。同様に、ステップＳ１３１１において、ステップＳ１３０３およびステップＳ１３０９で算出された帰属度に基づいて、オーディオクラスＮｏへの帰属確率$Ｐ_4$が算出される。
【００８２】
一方、ＣＬＳ＃３の分類処理として、ステップＳ１３１２ないしステップＳ１３１５の処理を行う。具体的には、ステップＳ１３１２において、処理対象信号と参照信号１との距離を算出し、ステップＳ１３１３において、処理対象信号が参照信号１のクラスタに属する帰属度を算出する。さらに、ステップＳ１３１４において、処理対象信号と参照信号２との距離を算出し、ステップＳ１３１５において、処理対象信号が参照信号２のクラスタに属する帰属度を算出する。
ここで、ステップＳ１３１６において、ステップＳ１３０５およびステップＳ１３１３で算出された帰属度に基づいて、オーディオクラスＳｐへの帰属確率$Ｐ_2$が算出される。同様に、ステップＳ１３１７において、ステップＳ１３０５およびステップＳ１３１５で算出された帰属度に基づいて、オーディオクラスＭｕへの帰属確率$Ｐ_3$が算出される。
【００８３】
次に、χ^2検定法による映像のショット分割処理を説明する。この処理は、図５のステップＳ１０７に相当する。
本発明の最良の実施の形態においては、分割χ^2検定法を用いて、ショットカットを得る。分割χ^2検定法は、まず動画像のフレームを４×４＝１６個の同じ大きさの矩形領域に分割し、各領域ごとに６４色種の色ヒストグラムＨ（ｆ，ｒ，ｂ）を作成する。ただし、ｆはビデオ信号のフレーム番号、ｒは領域番号、ｂはヒストグラムのビン数を表す。隣接する２枚の動画像のフレームの色ヒストグラムから、次式で算出される評価値$Ｅ_r$（ｒ＝１，・・・，１６）を算出する。
【数１３】

$$E_r = \sum_{b=0}^{63} \frac{\{H(f,r,b) - H(f-1,r,b)\}^2}{H(f,r,b)} \qquad （式１－７）$$

さらに、算出された１６個の値$Ｅ_r$（ｒ＝１，・・・，１６）の中で値の小さい８の総和$Ｅ_{sum}$算出し、$Ｅ_{sum}$が予め設定した閾値よりも大きな値を示す時刻に、ショットカットが存在すると判断する。
【００８４】
この処理を図１８を参照して説明する。
まずステップＳ１４０１において、ビジュアル信号のフレームのデータを取得する。次に、ステップＳ１４０２において、ステップＳ１４０１で取得したビジュアル信号のフレームを、４×４＝１６個の矩形領域に分割し、ステップＳ１４０３において、各領域について、６４色種の色ヒストグラムＨ（ｆ，ｒ，ｂ）を作成する。
さらにステップＳ１４０４において、隣接するビジュアル信号のフレーム間で、色ヒストグラムの差分評価$Ｅ_r$を算出する。各矩形領域について算出された差分評価$Ｅ_r$の中で、小さい８つの総和$Ｅ_{sum}$を算出する。
ステップＳ１４０６において、$Ｅ_{sum}$が閾値よりも大きな値を示す時刻で、ショットカットを判定し、ショット区間を出力する。
【００８５】
このように、本発明の最良の実施の形態においては、隣接する区間で大きく色ヒストグラムが変化する時刻をショットカットと判定することにより、ショット区間を出力している。
【００８６】

次に、各ショットに対するオーディオクラスへの帰属確率の算出処理を説明する。この処理は、図５のステップＳ１０８に相当する。

本発明の最良の実施の形態においては、まず単一のショット内における各オーディオクラスへの帰属確率の平均値

【数１４】

$$x_{l_1}\ (l_1=1,\cdots,4;\text{それぞれ Si, Sp, Mu, No に対応})$$

を次式で算出する。

【数１５】

$$x_{l_1}=\frac{1}{N}\sum_{k=0}^{N-1}P_{l_1}(k) \qquad \text{（式１－８）}$$

ただし、Ｎはショット内のクリップの総数、ｋはショット内のクリップ番号、

【数１６】

$$P_{l_1}(k)\ (l_1=1,\cdots,4)$$

はｋ番目のクリップにおける帰属確率

【数１７】

$$P_{l_1}$$

を表す。これら４つの平均値

【数１８】

$$x_{l_1}\ (l_1=1,\cdots,4)$$

の値を観察することで、分類対象であるショットが無音、音声、音楽、雑音のうち、どの種類のオーディオ信号を多く含むかが分かる。

【００８７】

しかしながら、このままでは音楽付き音声や雑音付き音声のクラスが存在せず、音楽付き音声や雑音付き音声が含まれていた場合、分類精度が劣化する危険性がある。ところで、従来手法で算出する帰属確率は、オーディオ信号の各クリップが、各オーディオクラスに属する度合いを表しており、音楽付き音声や雑音付き音声のオーディオ信号を処理対象とした場合、音声のオーディオクラスの帰属確率だけでなく、音楽や雑音のオーディオクラスの帰属確率も高い値を示す。そこで、

【数１９】

$$x_{l_1}$$

に対し、ファジィ推論を行うことで、各ショットを無音、音声、音楽、雑音、音楽付き音声、雑音付き音声の６種類のショットクラスに分類する。

【００８８】

本発明の最良の実施の形態においては、まず処理対象信号を無音、音声、音楽、雑音の4つのオーディオクラスに分類する。しかしながら、この４種類のクラスだけでは、背景

に音楽が存在する環境下での音声(雑音付き音声)や、背景に雑音が存在する環境下での音声(雑音付き音声)等、複数の種類のオーディオ信号が混在する場合に、分類精度が劣化する。そこで、本発明の最良の実施の形態では、上記４つのオーディオクラスに加え、新たに音楽付き音声、雑音付き音声のクラスを含む、６つのオーディオクラスへの分類を行う。これにより、分類精度を向上させ、より高精度に類似シーンを検索することができる。

【００８９】

まず、以下に示す１１段階のファジィ変数を用意する。

NB (Negative Big)
NBM (Negative Big Medium)
NM (Negative Medium)
NSM (Negative Small Medium)
NS (Negative Small)
ZO (Zero)
PS (Positive Small)
PSM (Positive Small Medium)
PM (Positive Medium)
PBM (Positive Big Medium)
PB (Positive Big)

ここで、それぞれのファジィ変数に対し、次式で定義される三角型のメンバシップ関数を定め、図１９に示すように、各変数を割り当てることで、ファジィ集合を生成する。

【数２０】

$$\mu(x_{l_1}) = \max\left(0, \frac{1}{a}(-|x_{l_1} - b| + a)\right) \qquad \text{（式１－９）}$$

ただし、a=0.1, b={0, 0.1, ・ ・ ・ , 0.9, 1.0}とする。（式１－８）で算出した

【数２１】

$$x_{l_1} \ (l_1 = 1, \cdots, 4)$$

を（式１－９）に代入し、各入力変数のメンバシップ関数の値

【数２２】

$$\mu(x_{l_1}) \ (l_1 = 1, \cdots, 4)$$

を算出する。

【００９０】

次に、各ショットに対するファジィ推論処理を説明する。この処理は、図５のステップＳ１０９に相当する。

本発明の最良の実施の形態においては、各ショットに対するオーディオクラスへの帰属確率の算出処理で設定された入力変数、およびメンバシップ関数の値

【数２３】

$$\mu(x_{l_1})$$

に対し、図２０および図２１に示すファジィ制御規則

【数２４】

$$R_{l_2}^{j} \quad (l_2 = 1, \cdots, 6, \text{ それぞれ Si, Sp, Mu, No, SpMu, SpNo に対応；} \ j = 1, 2, \cdots)$$

を適用し、各ショットクラスの出力変数

【数２５】

$$y_{l_2}$$

およびメンバシップ関数の値

【数２６】

$$\mu(y_{l_2})$$

を算出する。

【００９１】

次に、ファジィ推論値を用いたシーン分割処理を説明する。この処理は、図５のステップＳ１１０に相当する。

本発明の最良の実施の形態においては、ファジィ推論処理で算出される各ショットクラスに属する度合

【数２７】

$$\mu_{l_2}$$

を用いて、映像信号のシーン分割を行う。

ここで、ηをショット番号とし、隣接するショット間の距離Ｄ（η_1，η_2）を次式で定義する。

【数２８】

$$D(\eta_1, \eta_2) = \sum_{l_2=1}^{6} |\mu_{l_2}(\eta_1) - \mu_{l_2}(\eta_2)| \qquad \text{（式１－１０）}$$

【００９２】

この距離Ｄ（η_1，η_2）があらかじめ設定した閾値Th_Dよりも高い値を示す場合、ショット間の類似度は低く、ショットの境界にシーンカットが存在すると判断する。逆に、距離Ｄ（η_1，η_2）が閾値Th_Dよりも低い値を示す場合、ショット間の類似度が高く、同一のシーンに属すると判断する。これにより、本発明の最良の実施の形態ではショット間の類似度を考慮したシーン分割が可能となる。

【００９３】

ここで、各ショットに対するオーディオクラスへの帰属確率の算出処理、各ショットに対するファジィ推論処理およびファジィ推論値を用いたシーン分割処理を、図２２参照して説明する。

まずステップＳ１５０１において、各ショットの全クリップにおける帰属確率の平均値を算出する。つぎにステップＳ１５０２において、１１段階のファジィ係数を読み出し、各ショットに対するメンバシップ関数を算出する。ステップＳ１５０１およびステップＳ１５０２の処理は、各ショットに対するオーディオクラスへの帰属確率の算出処理に相当する。

ステップＳ１５０３において、入力変数およびメンバシップ関数の値から、出力およびその出力のメンバシップ関数の値を算出する。このとき、図２０および図２１に示すファジィ制御規則が参照される。ステップＳ１５０３の処理は、各ショットに対するオーディオクラスへの帰属確率の算出処理に相当する。

さらにステップＳ１５０４において、異なるショット間でのメンバシップ関数の距離を算出し、ステップＳ１５０５において、その距離が閾値よりも大きいか判定する。その距離が閾値よりも大きい場合、そのフレーム間で映像信号のシーンカットを判定し、シーン区間を出力する。ステップＳ１５０４およびステップＳ１５０５の処理は、ファジィ推論値を用いたシーン分割処理に相当する。

【００９４】

このように、本発明の最良の実施の形態においては、χ^2検定法によるビジュアル信号のショット分割処理によって分割された各ショットについて、各ショットに属するクリップのオーディオ信号についてオーディオクラスへの帰属確率を算出し、ファジィ推論することにより、ファジィ推論値を用いてシーンを分割することができる。

【００９５】

（ビデオ信号類似度算出部）

次に、図１に示すビデオ信号類似度算出部２３の処理を説明する。

ビデオ信号類似度算出部２３は、映像情報に注目した検索・分類を行うため、シーン分割部２１で算出される各シーンに対して、他のシーンとの類似度を算出する処理について説明を行う。本発明の最良の実施の形態では、動画データベース１１中に存在する映像のシーン間について、ビジュアル(動画像)信号の特徴量とオーディオ信号の特徴量から、それらの類似度を類似度として算出する。本発明の最良の実施の形態では、まず映像中のシーンをクリップに分割し、各々に対してビジュアル信号の特徴量の抽出、およびオーディオ信号の特徴量の抽出を行う。さらに、これらの特徴量に対して３次元のＤＴＷを設定することで、シーン間の類似度の算出を可能とする。

【００９６】

ＤＴＷは，２つの１次元信号に伸縮を施し，信号間の類似度を算出する手法である。このため，信号の伸縮が頻繁に生じる信号間の比較に有効である。

本発明の最良の実施の形態では、従来２次元で定義されているＤＴＷを３次元で再定義し、新たにそれらを用いるためのコストの設定を行っている。このとき、コストをビジュアル信号およびオーディオ信号のそれぞれに設定することにより、２つのシーン間で動画像、音響の一方が異なる場合においても、類似した映像を検索・分類することが可能となる。さらに、ＤＴＷの特徴からシーン間の時間尺が異なる場合や、シーン間でビジュアル信号とオーディオ信号の開始時刻にずれが生じた場合においても、適切にシーン間の類似部分を対応付けることが可能となる。

【００９７】

本発明の最良の実施の形態に係るビデオ信号類似度算出部２３について、具体的なアルゴリズムを説明する。

本発明の最良の実施の形態では、映像に含まれるビジュアル信号（動画像信号）とオーディオ信号（音響信号）の双方に着目してシーン間の類似度を算出する。まず、本発明の最良の実施の形態では、与えられたシーンを短時間のクリップに分割し、シーンをクリップの一次元列として表現する。次に、各クリップからビジュアル信号による特徴量、およびオーディオ信号による特徴量をそれぞれ抽出する。最後に、ＤＴＷを用いてクリップ列間の特徴量の類似部分を対応付けし、得られる最適経路をシーン間の類似度として定義する。ここで本発明の最良の実施の形態では、ＤＴＷを新たに３次元に拡張して用いること

で、ビジュアル信号とオーディオ信号の協調処理によるシーン間の類似度の算出を可能とした。以下、各処理について説明する。

【００９８】

まず、ビデオ信号のクリップへの分割処理を説明する。この処理は、図６のステップＳ２０１に相当する。

本発明の最良の実施の形態では、処理対象であるシーンを、短時間T_c［ｓｅｃ］のクリップに分割する。

【００９９】

次に、ビジュアル信号の特徴量抽出処理を説明する。この処理は、図６のステップＳ２０２に相当する。

本発明の最良の実施の形態では、ビデオ信号のクリップへの分割処理で得られる各クリップからビジュアル信号の特徴量を抽出する。本発明の最良の実施の形態では、ビジュアル信号の特徴として画像の色成分に着目し、各クリップの動画像の所定のフレームからＨＳＶ表色系における色ヒストグラムを算出し特徴量に用いる。ここで、動画像の所定のフレームとは、例えば各クリップの動画像の先頭のフレームである。また、人間の知覚システムにおいて色相がより重要なことに着目し、色相、彩度、明度のヒストグラムのビン数を、例えばそれぞれ１２、２、２とする。よって、クリップ単位から得られるビジュアル信号の特徴量は全部で４８次元になる。本実施例においては、色相、彩度、明度のヒストグラムのビン数が、１２、２、２の場合について説明するが、任意に設定されても良い。

【０１００】

この処理を図２３を参照して説明する。

まず、ステップＳ２１０１において、クリップの動画像の所定のフレームを抽出し、ステップＳ２１０２において、ＲＧＢ表色系からＨＳＶ表色系へ変換する。

つぎにステップＳ２１０３において、例えば、Ｈ軸を１２、Ｓ軸を２、Ｖ軸を２に分割した３次元色ヒストグラムを生成して、この３次元色ヒストグラムを当該クリップのビジュアル信号の特徴量として算出する。

【０１０１】

次に、オーディオ信号の特徴量抽出処理を説明する。この処理は、図６のステップＳ２０３に相当する。

本発明の最良の実施の形態において、ビデオ信号のクリップへの分割処理で得られる各クリップからオーディオ信号の特徴量を抽出する。本発明の最良の実施の形態では、オーディオ信号の特徴量として１０次元の特徴量を用いた。具体的には、クリップに含まれるオーディオ信号を固定長T_f［ｓｅｃ］（$T_f < T_c$）のフレーム毎に解析を行う。

まず、各クリップからオーディオ信号の特徴量を抽出する際に、オーディオ信号に含まれる音声部分の影響を軽減するために、オーディオ信号の各フレームを音声フレームと背景音フレームに分類する。ここで、オーディオ信号における音声部分の特徴は大きな振幅と、大部分がフォルマント周波数と呼ばれる低周波数のパワーを持つことに着目し、短時間のエネルギー（以降、ＳＴＥ）と短時間のスペクトル（以降、ＳＴＳ）を用いてオーディオ信号の各フレームを分類する。

【０１０２】

ここで、オーディオ信号の各フレームから得られるＳＴＥとＳＴＳを次式で定義する。

【数２９】

$$STE(n) = \sqrt{\frac{1}{L}\sum_{m}[x(m)\omega(m - nF_s)]^2}$$

（式２－１）

$$STS(k) = \frac{1}{2\pi L}\left|\sum_{m=0}^{L-1} x(m)e^{-j\frac{2\pi}{L}km}\right|$$

（式２－２）

ここで、ηはオーディオ信号のフレーム番号、Ｆ$_s$はオーディオ信号のフレームの移動幅を表す移動回数、ｘ（ｍ）はオーディオの離散信号、ω（ｍ）はｍが時間枠の中にあれば１を、そうでなければ０を取る。また、ＳＴＳ（ｋ）は周波数が

【数３０】

$$\frac{kf}{L}(k = 0, \ldots, L - 1)$$

のときの短時間のスペクトルであり、ｆは離散サンプリング周波数である。もし、ＳＴＥの値が閾値Ｔｈ$_1$を越えていて、尚かつ４４０－４０００Ｈｚの範囲でのＳＴＳの値が閾値ＴＨ$_2$を越えていれば、そのオーディオ信号のフレームは音声フレームとして、越えていなければ背景音フレームとして分類する。

【０１０３】

これらの分類されたオーディオ信号のフレームを用いて、以下に示すクリップ単位の１０次元の特徴量を算出する。

【数３１】

a) 短時間の平均エネルギー $\overline{STE}$:

$$\overline{STE} = \frac{1}{N}\sum_{n=0}^{N-1} STE(n)$$

（式２－３）

ここで、平均エネルギーとは、クリップ内のオーディオ信号の全フレームが持つエネルギーの平均である。

【数３２】

b) 低 STE 率 $LSTER$:

$$LSTER = \frac{1}{2N_B}\sum_{n=0}^{N_B-1}\left|sgn\left[\overline{STE} - STE(n)\right] + 1\right|$$

（式２－４）

ここで、低エネルギー率（低ＳＴＥ率）とは、クリップ内のエネルギーの平均以下のエネルギーを持つ背景音フレームの割合である。

【数３３】

c) 平均零交差率 $\overline{ZCR}$:

零交差率 $ZCR(n)$ は次式で定義される.

$$ZCR(n) = \frac{1}{2}\sum_{m} |sgn[x(m)] - sgn[x(m-1)]|\omega(m) \qquad (式２－５)$$

ここで，$x(m) \geq 0$ のとき $sgn[x(m)] = 1$，その他では $sgn[x(m)] = -1$ である. $\overline{ZCR}$ は背景音フレームでの ZCR の平均である.

ここで、平均零交差率とは、クリップ内の全背景音フレーム内における隣り合うオーディオ信号の符号が変化する割合の平均である。

【数３４】

d) スペクトルフラックス密度 SF:

$$SF = \frac{1}{(N-1)(K-1)}\sum_{n=1}^{N-1}\sum_{k=1}^{K-1} |log(STS(n,k)) - log(STS(n-1,k))| \qquad (式２－６)$$

ここで，$STS(n,k), (k = 1, ..., K)$ は時刻 n における k 番目のスペクトルである.

ここで、スペクトルフラックス密度とは、クリップ内のオーディオ信号が持つ周波数スペクトルの時間推移の指標である。

ｅ）音声フレーム率ＶＦＲ：

ここで、ＶＦＲはクリップに含まれるオーディオ信号の全フレームにおける音声フレームの割合である。

【数３５】

f) 平均サブバンドエネルギー比率，$\overline{ERSB}1/2/3/4$:

$\overline{ERSB}1/2/3/4$ は 0 – 630Hz，630 – 1720Hz，1720 – 4400Hz，4400 – 11000Hz の各帯域での平均サブバンドエネルギー比率である.

ここで、平均サブバンドエネルギー比率とは、クリップ内のオーディオ信号のオーディオスペクトルに対し全周波数でのパワースペクトルの総和に対しての、０－６３０、６３０－１７２０、１７２０－４４００、４４００－１１０００（Ｈｚ）のそれぞれの範囲におけるパワースペクトルの割合である。

ｇ）ＳＴＥ標準偏差ＥＳＴＤ：

ＳＴＥの標準偏差ＥＳＴＤは、次式で定義される。

【数３６】

$$ESTD = \sqrt{\frac{1}{N}\sum_{n=0}^{N-1}(STE(n) - \overline{STE})^2} \qquad (式２－７)$$

ここで、エネルギー（ＳＴＥ）標準偏差とは、クリップ内のオーディオ信号の全フレームが持つエネルギーの標準偏差である。

【０１０４】

この処理を図２４を参照して説明する。

まずステップＳ２２０１において、オーディオ信号の各クリップについて、短時間のオ

ーディオ信号のフレームへ分割される。つぎにステップＳ２２０２において、オーディオ信号のフレーム内のオーディオ信号が持つエネルギーが算出されるとともに、ステップＳ２２０３において、フレーム内のオーディオ信号が持つスペクトルが算出される。

ステップＳ２２０４において、ステップＳ２２０１で分割されたオーディオ信号の各フレームが、音声フレームと背景音フレームに分類される。この分類されたオーディオ信号のフレームに基づいて、ステップＳ２２０５において、上述したａ）からｇ）の各特徴量が算出される。

【０１０５】

次に、３次元ＤＴＷを用いたシーン間の類似度算出処理を説明する。この処理は、図６のステップＳ２０４に相当する。

本発明の最良の実施の形態では、ビジュアル信号の特徴量抽出処理およびオーディオ信号の特徴量抽出処理で得られたクリップ単位の特徴量を用いて、シーン間の類似度を定義する。一般的に、クリップ列の比較にＤＴＷを用いて類似部分を対応づけ、得られる最適経路をシーン間の類似度として定義している。しかしながら、この場合、ＤＴＷに用いる局所コストをクリップ間の全特徴量の差に基づき決定しているため、シーン間において片方の信号のみが類似している場合や、シーン間においてビジュアル信号とオーディオ信号の開始時刻にズレが発生した場合などに適切な類似度が得られない可能性がある。

【０１０６】

そこで、本発明の最良の実施の形態では、ＤＴＷを３次元に拡張して新たな局所コストと局所パスを設定する事で、これらの問題を解決する。以下、（処理４－１）、（処理４－２）でそれぞれ３次元ＤＴＷで用いられる局所コストと局所パスについて説明する。さらに、（処理４－３）で３次元ＤＴＷにより算出されるシーン間の類似度について説明する。

【０１０７】

（処理４－１）局所コストの設定

本発明の最良の実施の形態では、まず、３次元ＤＴＷの３つの要素として、クエリシーンのクリップτ（$1 \leqq \tau \leqq T_1$）、ターゲットシーンのビジュアル信号のクリップt_x（$1 \leqq t_x \leqq T_2$）、ターゲットシーンのオーディオ信号のクリップt_y（$1 \leqq t_y \leqq T_2$）をそれぞれ用いる。この３つの要素に対し、３次元ＤＴＷ上の各格子点における局所コスト$d(\tau, t_x, t_y)$を以下の３種類で定義する。

【数３７】

$$d(\tau,t_x,t_y)=\begin{cases} |f_{V,\tau}^{query}-f_{V,t_x}^{target}| & \cdots d_v(\tau,t_x,t_y) \\ |f_{A,\tau}^{query}-f_{A,t_y}^{target}| & \cdots d_a(\tau,t_x,t_y) \\ \frac{d_v(\tau,t_x,t_y)+d_a(\tau,t_x,t_y)}{2} & \cdots d_{av}(\tau,t_x,t_y) \end{cases} \quad \text{（式２－８）}$$

ここで、$f_{v,t}$は時刻ｔのクリップに含まれるビジュアル信号から得られる特徴ベクトル、$f_{A,t}$は時刻ｔのクリップに含まれるオーディオ信号から得られる特徴ベクトルであり、各時刻において特徴量の総和が１となるようにそれぞれ正規化されている。

【０１０８】

（処理４－２）局所パスの設定

本発明の最良の実施の形態で用いられる３次元ＤＴＷ上の各格子点は、図２５および図２６に示すように直前の７つの格子点からそれぞれ局所パス＃１～＃７で連結されている。以下に各局所パスが持つ役割を示す。

ａ）局所パス＃１および＃２について

クリップ単位による伸縮を許容するパスである。パス＃１はクエリシーンのクリップの時間軸方向への伸縮を、パス＃２はターゲットシーンのクリップの時間軸方向への伸縮をそれぞれ許容する役割を持つ。

ｂ）局所パス＃３ないし＃５について

類似部分の対応付けを行うパスである。クリップ間において、パス＃３はビジュアル信号を、パス＃４はオーディオ信号を、パス＃５は両方の信号を類似部分としてそれぞれ対応付けを行う役割を持つ。

ｃ）局所パス＃６および＃７について

両信号の同期によるズレを許容するパスである。パス＃６はシーン間におけるビジュアル信号の時間軸方向へのズレを、パス##７はシーン間におけるオーディオ信号の時間軸方向へのズレをそれぞれ許容する役割を持つ。

【０１０９】

（処理４－３）シーン間の類似度の定義

上述した（処理４－１）および（処理４－２）で説明した局所コストと局所パスを用いて、累積コストＳ（τ，t_x，t_y）を直前の７つの格子点からの累積コストと移動コストの和が最小となる格子点を用いて、以下で定義する。

【数３８】

$$S(0,0,0) = \min(d_v(0,0,0), d_a(0,0,0), d_{av}(0,0,0)) \qquad \text{（式２－９）}$$

【数３９】

$$S(\tau,t_x,t_y) = \min\left\{\begin{array}{c} S(\tau-1,t_x,t_y)+d_{av}(\tau,t_x,t_y)+\alpha \\ S(\tau,t_x-1,t_y-1)+d_{av}(\tau,t_x,t_y)+\alpha \\ S(\tau-1,t_x-1,t_y)+d_v(\tau,t_x,t_y)+\beta \\ S(\tau-1,t_x,t_y-1)+d_a(\tau,t_x,t_y)+\beta \\ S(\tau-1,t_x-1,t_y-1)+d_{av}(\tau,t_x,t_y) \\ S(\tau,t_x-1,t_y)+d_v(\tau,t_x,t_y)+\gamma \\ S(\tau,t_x,t_y-1)+d_a(\tau,t_x,t_y)+\gamma \end{array}\right\} \qquad \text{（式２－１０）}$$

ただし、α、β、γはそれぞれ対応する局所パスを用いた場合にかかる移動コストを表す定数である。これにより、最終的なシーン間の類似部分の対応付けと、その対応付けによるシーン間の類似度D_sは次式により定義される。

【数４０】

$$D_S = \min\left(\frac{S(T_1,T_2,t_y)}{T_1+2T_2}, \frac{S(T_1,t_x,T_2)}{T_1+2T_2}\right) \qquad \text{（式２－１１）}$$

【０１１０】

この処理を図２７を参照して説明する。

まず、ステップＳ２３０１において、３次元ＤＴＷを用いたシーン間の特徴量に基づくマッチングを行う。具体的には、上記（式２－１０）における｛｝内の７つの結果のうち、最小のものを選択する。

つぎにステップＳ２３０２において、３次元ＤＴＷに必要な局所コストが設定され、ステップＳ２３０３において、局所パスが設定される。さらにステップＳ２３０４においてα、β、γの各移動コストを設定する。αは、パス＃１およびパス＃２の移動コストであり、βは、パス＃３およびパス＃４の移動コストであり、γは、パス＃６およびパス＃７の移動コストである。

さらにステップＳ２３０５において、マッチングによる最適経路をシーン間の類似度として算出する。

【０１１１】

このように、本発明の最良の実施の形態においては、ビジュアル信号の特徴量とオーデ

ィオ信号の特徴量に基づいて、３次元ＤＴＷを用いてシーン間の類似度を算出する。ここで３次元ＤＴＷを用いることにより、後述する表示部で、３次元座標を基づいてシーンの類似度を可視化することができる。

【０１１２】

（ＤＴＷの概要）

ここで、ＤＴＷの概要について説明する。

本発明の最良の実施の形態における類似度算出処理で用いられるＤＴＷの構成について説明を行う。ＤＴＷは、二つの一次元信号に伸縮を施し、信号間の類似度を算出する手法である。このため、時系列において伸縮の生じる信号等の比較に有効である。特に音楽信号では、演奏速度の変化が頻繁に発生することから、類似度の算出にＤＴＷを用いることは有効と考えられる。以降、類似度算出において、参照する信号を参照パターン、参照パターンとの類似度を求める信号を被参照パターンと呼ぶ。

【０１１３】

まず、ＤＴＷによるパターン間の類似度の算出について説明する。長さＩの一次元の参照パターンに含まれる各要素を順にa_1，a_2，・・・a_Iとし、長さＪの被参照パターンに含まれる各要素を順にb_1，b_2，・・・b_Jと表現する。さらに、各パターンの位置集合を｛１，２，・・・，Ｉ｝，｛１，２，・・・，Ｊ｝で表現すると、パターンの各要素間の対応を決定する伸縮写像ｗ：｛１，２，・・・，Ｉ｝->｛１，２，・・・，Ｊ｝は以下の性質を満たす。

ａ）ｗはパターンの始点、終点を一致させる。

【数４１】

$$w(1) = 1$$
$$w(I) = J \qquad \text{（式２－１２）}$$

ｂ）ｗは単調写像である。

【数４２】

$$\forall i,j \in \{1,2,\ldots,I\} : (i \leq j \Rightarrow w(i) \leq w(j)) \qquad \text{（式２－１３）}$$

【０１１４】

このような写像ｗを用いたとき、パターン間の類似度の算出は図２８における格子点（b_1，a_1）から格子点（b_J，a_I）までの最短経路の探索問題に置換することができる。そこで、ＤＴＷでは、「初期状態の最初の決定が何であろうとも、以後の決定は最初の遷移から生じた状態に関して適切でなければならない」という最適性の原理に基づいて上記の経路探索問題を解く。

【０１１５】

すなわち、全体の経路長を部分の経路長の和で求める。部分の経路長は、経路上の格子点（ｊ，ｉ）におけるコストｄ（ｊ，ｉ）および２つの格子点（ｊ，ｉ）、（ｂ，ａ）間の移動コスト$c_{j,i}$（ｂ，ａ）を用いて算出する。部分の経路長の算出を図２９に示す。ここで、格子点上のコストｄ（ｊ，ｉ）は参照パターンと被参照パターンの間で対応する要素が異なる場合のペナルティである。また、移動コスト$c_{j,i}$（ｂ，ａ）は参照パターンと被参照パターンの間で伸縮が生じた場合、格子点（ｂ，ａ）から格子点（ｊ，ｉ）に移動するペナルティである。

【０１１６】

上記のコストに基づいて部分の経路長を算出し、経路全体のコストが最小となる部分経路を選択する。最後に、選択された部分経路毎のコストの和を算出することで、全体の経路長が得られる。以上より、パターンの部分毎の類似度からパターン全体の類似度を得ることが可能となる。

【０１１７】
本発明の最良の実施の形態においては、ＤＴＷをオーディオ信号に適用することから、オーディオ信号の類似度算出における特徴を考慮し、さらに詳細な類似度の算出法を決定する。
本発明の最良の実施の形態では、音楽の特徴として、同一楽曲の演奏速度が異なる場合にも、楽譜上の音符が欠落することがない点に着眼する。この特徴を換言すると以下の２点で表現可能と考えられる。
ａ）被参照パターンが、参照パターンに伸縮のみを加えたパターンである場合、これらのパターンは同一と見なす。
ｂ）被参照パターンと参照パターンが同一の場合、被参照パターンは参照パターンを欠落することなく含有する。
【０１１８】
上記の特徴を、格子点間の移動による類似度算出に適用すると、参照パターンに含まれる全ての要素について、被参照パターンに含まれる要素との対応を決定することを意味する。これより、伸縮写像ｗは次式に示す傾斜制限を加えることが可能となる。
【数４３】

$$w(i) \leq w(i+1) \leq w(i)+1 (1 \leq i \leq I) \qquad \text{（式２－１４）}$$

本発明の最良の実施の形態では、以上の条件に従ってＤＴＷによる類似度の算出を行う。これより類似度は、（式２－１５）を用いて経路長を漸化的に求めることで算出可能となる。
【数４４】

$$D(j+1,i+1) = d(j+1,i+1) + \min\{(D(j,i) + c_{j+1,i+1}(j,i)), (D(j,i+1) + c_{j+1,i+1}(j,i+1)), (D(j+1,i) + c_{j+1,i+1}(j+1,i))\} \qquad \text{（式２－１５）}$$

【０１１９】
（オーディオ信号類似度算出部）
次に、図１に示すオーディオ信号類似度算出部２４の処理を説明する。
オーディオ信号類似度算出部２４は、シーン分割部２１で算出されるシーンに対して音楽情報に注目した検索・分類を行うため、類似度の算出を行う。本発明の最良の実施の形態では、動画データベース１１からシーン分割部２１で得られる全てのシーン中で、オーディオ信号のベース音に基づく類似度、他楽器に基づく類似度、リズムに基づく類似度を算出する。本発明の最良の実施の形態では、オーディオ信号類似度算出部２４は、オーディオ信号に対して以下の三種類の類似度算出を行う。
・ ベース音に基づく類似度算出
・ 他楽器に基づく類似度算出
・ リズムに基づく類似度算出
【０１２０】
ベース音に基づく類似度算出について、本発明の最良の実施の形態では、オーディオ信号に対して、ベース音を含むと考えられる周波数の信号のみを求めるため、帯域通過フィルタを施す。次に、得られる信号から各時刻におけるスペクトルを求めるため、時間・周波数に注目した重み関数を用いて、重み付きパワースペクトルの算出を行う。さらに、得られる各時刻のパワースペクトルにおいてピークを持つ周波数を求めることで、ベース音高の推定を可能とする。さらに、全ての２シーン間について、そのオーディオ信号のベース音高の推移を求め、これをＤＴＷへ入力することで、二つの信号の類似度の算出を実現する。
【０１２１】
他楽器に基づく類似度算出について、本発明の最良の実施の形態では、オーディオ信号に対して、「ド」、「レ」、「ミ」、「ソ#」等、音名１２要素を示す周波数のエネルギ

ーをパワースペクトルから算出する。さらに、これら１２要素のエネルギーを正規化することで、エネルギーの割合の時間推移を算出する。このようにして得られるエネルギーの割合についてＤＴＷを用いることで、本発明の最良の実施の形態では全ての２シーン間で、オーディオ信号の他楽器に基づく類似度算出が可能となる。

【０１２２】

リズムに基づく類似度算出について、本発明の最良の実施の形態では、まず、オーディオ信号に対して、２分割フィルタバンクを用いることで、異なる周波数を含む信号をそれぞれ算出する。次に、各周波数を含む信号に対して、“信号の各時刻における接線を共有する曲線である”包絡線の検波を行い、信号の概形を得る。尚、この処理は、「全波整流」、「低域通過フィルタの適用」、「ダウンサンプリング」、「平均値除去」を順に施すことで、実現される。さらに、これらの信号をすべて足し合わせて得られる信号に対して、自己相関関数を求め、これをリズム関数として定義する。最後に、全ての２シーン間で、それらのオーディオ信号のリズム関数をDTW へ入力することで、二つの信号の類似度の算出を実現する。

【０１２３】

以上に示す、３つの類似度算出処理を施すことで、本発明の最良の実施の形態では３つの類似度を楽曲間の類似性を表す指標として求めることが可能となる。

【０１２４】

このように本発明の最良の実施の形態では、音楽の構成要素であるメロディーに着眼している。音楽におけるメロディーとは、複数の音源により構成される基本周波数の時間推移である。本発明の最良の実施の形態では、このメロディーの定義に従い、メロディーがベース音と、それ以外の楽器音から構成されると仮定する。さらに、この仮定に基づき、ベース音が示すエネルギーの推移、およびベース以外の楽器が示すエネルギーの推移についてマッチング処理を施すことで類似度を得る。ベース音が示すエネルギーには、ベース音が存在する周波数域のパワースペクトル、その他の楽器音が示すエネルギーには、Ｃ、Ｄ、Ｅ・・・等の音名が示す周波数のエネルギーを用いる。上記のエネルギーを用いると、音楽信号における以下２点の特徴に有効と考えられる。

まず、楽器音は基本周波数の倍音を多く含む(以降、倍音構造)ため、周波数域が高くなるに従い、基本周波数の特定が困難となる点である。次に、楽曲中には発音の際に発生する擦弦音等の雑音が含まれ、音階上に存在しない周波数が楽器音の基本周波数として推定され得る点である。

【０１２５】

本発明の最良の実施の形態は、ベース以外の楽器音のエネルギーとして、各音名が示す周波数のエネルギーを用いるため、上記の倍音構造、雑音の影響を軽減可能とすることができる。また、低周波数域に基本周波数を持つベース音を併せて用いることで、倍音構造の影響をより軽減した類似度算出を可能とすることができる。さらに、類似度の算出にはＤＴＷを用いるため、メロディーの伸縮や欠落が生じた場合にも類似度算出をすることができる。以上により、本発明の最良の実施の形態はメロディーに基づいて楽曲間の類似度を算出することができる。

【０１２６】

さらに、音楽の構成では、メロディーに加えてリズムが重要な要素として知られる。そこで、本発明の最良の実施の形態では、音楽の構成要素として新たにリズムに着眼し、リズムから楽曲間の類似度を算出する。また、類似度算出には、ＤＴＷを用いることで、楽曲の時間軸方向への伸縮を許容し、適切な類似度の算出を可能とする。

【０１２７】

本発明の最良の実施の形態に係るオーディオ信号類似度算出部２４は、映像中の音楽情報、つまりオーディオ信号に対して、「ベース音に基づく類似度」、「他楽器に基づく類似度」、「リズムに基づく類似度」の算出を行う。

まず、本発明の最良の実施の形態においては、音楽のメロディーの推移に着眼し、楽曲の類似度算出を可能とする。本発明の最良の実施の形態では、メロディーがベース音、お

よびベース以外の楽器音から構成されると仮定する。これは、ベース音と他楽器音により同時に発音される音がメロディーの特徴を決定する和音や調の指標となるためである。

【０１２８】

本発明の最良の実施の形態では上記の仮定に基づき、それぞれの楽器音のエネルギーにＤＴＷを適用することで類似度の算出を可能とする。

さらに、本発明の最良の実施の形態においては、楽曲のリズムに基づく新たな類似度を算出する。音楽におけるリズムは、メロディー、コード（和音）と併せて音楽の三要素と呼ばれ、楽曲の細かな構成を決定する重要な要素として知られる。そこで、本発明の最良の実施の形態では、リズムに着眼して楽曲間の類似度を定義する。

【０１２９】

本発明の最良の実施の形態は、音楽信号の自己相関関数に基づいてリズムを表す定量値（以降、リズム関数）を新たに定義し、リズム関数にＤＴＷを適用することで類似度の算出を行う。これにより、本発明の最良の実施の形態は、音楽の構成要素として重要なリズムに基づく類似度の算出を実現可能とする。

以下、「ベース音に基づく類似度」、「他楽器に基づく類似度」、「リズムに基づく類似度」のそれぞれについて、詳述する。

【０１３０】

（ベース音に基づく類似度算出）

オーディオ信号類似度算出部２４において、ベース音に基づく類似度算出処理を説明する。この処理は、図７のステップＳ３０１および図８に相当する。

本発明の最良の実施の形態では、楽曲中のベース音の推移として、ベース音が示す音高の推移を用いる。音高とは、楽譜上に記載される各音符が示す基本周波数とする。したがって、音高の推移はベース音に含まれる主要な周波数におけるエネルギーの推移を意味する。

【０１３１】

ベース音に基づく類似度算出においては、図３０に示すように、まず、帯域通過フィルタによってベース音が抽出される。このときのパワースペクトルを、Ｇ１１に示す。このパワースペクトルから、重み付きパワースペクトルを算出し、Ｇ１２に示すように、それぞれの音階をあてはめる。さらに、Ｇ１３に示すように、音階ごとに、ヒストグラムを算出する。このとき、ヒストグラムで最大値を持つ「Ｂ」が、ベース音の音階として選択される。

図３０においては、パワースペクトルから音階をあてはめ、その後、ベース音の音階を選択する場合について説明したが、この方法には限られない。具体的には、パワースペクトルから、周波数毎のヒストグラムを取得し、最大値の周波数から音階を取得しても良い。

【０１３２】

ベース音に基づく類似度算出処理について、具体的なアルゴリズムを以下に示す。尚、各処理は図８の各ステップに対応する。

【０１３３】

まず、通過帯域フィルタによるベース音の抽出処理を説明する。この処理は、図８のステップＳ３１１に相当する。

この処理では、オーディオ信号に対し、ベース音の周波数域４０－２５０Ｈｚを通過域とする帯域通過フィルタを施し、得られた信号の各時刻でパワースペクトルを算出する。

【０１３４】

つぎに、時間・周波数に注目した重み付きパワースペクトルの算出処理を説明する。この処理は、図８のステップＳ３１２に相当する。

この処理では、通過帯域フィルタによるベース音の抽出処理で得られるパワースペクトルの時間軸方向、および周波数軸方向に、ガウス関数に基づく重みを付加する。ここで、時間軸関数の重みを付加することにより、対象時刻のパワースペクトルが大きく利用される。周波数軸方向の重みを付加することにより、各音階（Ｃ、Ｃ＃、Ｄ、・・・、Ｈ）に

重みを置くことで、音階上の信号が選択される。ここで、ガウス関数による重みとは、ｅｘｐ｛－（ｘ－μ）／（２σ２）｝である（μ＝平均、σ＝標準偏差）。最後に、重み付けされた各時刻のパワースペクトルにおいて最大のエネルギーを与える周波数を音高として推定する。時刻ｔ（０≦ｔ≦Ｔ）、周波数ｆにおいて、パワースペクトルより算出されるエネルギーをＰ（ｔ、ｆ）とし、重み付けされたパワースペクトルを（式３－１）に示すＲ（ｔ、ｆ）で定義する。

【数４５】

$$R(t,f)=\int_0^T P(s,f)\cdot v_t(s)\cdot w(f)ds \qquad \text{（式３－１）}$$

ここで、

【数４６】

・時間軸方向の重み：$v_t(s)$

$$v_t(s)=\begin{cases}\exp\left\{\frac{-(t-s)^2}{2\sigma^2}\right\} & \text{if } t-3\sigma\le s\le t+3\sigma\\ 0 & \text{otherwise}\end{cases}$$

ただし，σは音の持続時間の指標となる定数である． （式３－２）

【数４７】

・周波数軸方向の重み：$w(f)$

$$w(f)=\begin{cases}\exp\left\{\frac{-f^2}{2\sigma_m^2}\right\} & \text{if } \frac{F_{m-1}+F_m}{2}\le f<F_m\\ \exp\left\{\frac{-f^2}{2\sigma_{m+1}^2}\right\} & \text{if } F_m\le f<\frac{F_m+F_{m+1}}{2}\\ 0 & \text{otherwise}\end{cases} \qquad \text{（式３－３）}$$

ただし，mを自然数として，

$$F_m=440\cdot 2^{\frac{m-69}{12}} \qquad \text{（式３－４）}$$

$$\sigma_m=\frac{F_m-F_{m-1}}{6} \qquad \text{（式３－５）}$$

また、（式３－４）で示すＦ$_m$は、MIDI（Musical Instrument Digital Interface）のｍ番目のノートにおける周波数を表す。

（式３－１）に示すＲ（ｔ、ｆ）は、（式３－２）の時間軸方向の重みにより、一定時間持続する基本周波数を音高と推定可能とする。また、（式３－３）に示す周波数軸方向の重みにより、音階上に存在する周波数のみを音高として推定可能とする。

【０１３５】

つぎに、重み付きパワースペクトルを用いたベースの音高推定処理を説明する。この処理は、図８のステップＳ３１３に相当する。

この処理では、Ｒ（ｔ、ｆ）の各時刻ｔにおいて最大値を与える周波数ｆをベースの音

高とし、Ｂ（ｔ）と表す。

【０１３６】

つぎに、ＤＴＷを用いたベース音高の類似度算出処理を説明する。この処理は、図８のステップＳ３１４に相当する。

この処理では、データベース中の全ての二映像間においてオーディオ信号のベース音高を推定し、上述したＤＴＷによる類似度算出を行う。ここで、上述したＤＴＷの説明において、（式２－１５）中で用いる各コストは以下のように設定する。

【数４８】

$$d(j,i)=\begin{cases}\alpha & \text{if } a_i \neq b_j \\ 0 & \text{otherwise}\end{cases} \qquad \text{（式３－６）}$$

$$c_{j,i}(b,a)=\begin{cases}\beta & \text{if } (b,a)=(j-1,i),(j,i-1) \\ 0 & \text{otherwise}\end{cases} \qquad \text{（式３－７）}$$

ただし、α＞βとする。これにより、メロディーの不一致によるコストと比較して、演奏速度の変化等に伴うメロディーのずれに対するコストが小さくなる。以上により得られた類似度をD_bと表す。

【０１３７】

ここで、図３１を参照して、本発明の最良の実施の形態に係るベース音に基づく類似度算出処理を説明する。

まず、動画データベース１１の各シーンについて、ステップＳ３１０１ないしステップＳ３１０９の処理が実行される。

ステップＳ３１０１において、１つのシーンにフーリエ変換をする。ステップＳ３１０２において、４０－２５０Ｈｚを通過域とするフィルタを施す。ステップＳ３１０３において、各時刻について、パワースペクトルＰ（ｓ，ｆ）を算出する。

【０１３８】

一方、ステップＳ３１０４において、時間軸方向の重みを算出するとともに、ステップＳ３１０５において、周波数軸方向の重みを算出する。さらにステップＳ３１０６において、ステップＳ３１０４およびステップＳ３１０５において算出された時間軸方向の重みおよび周波数軸方向の重みに基づいて、重み付きパワースペクトルを算出して、ステップＳ３１０７においてＲ（ｔ，ｆ）を出力する。さらに、各時刻ｔでＲ（ｔ、ｆ）の最大値を与える周波数ｆを求め、Ｂ（ｔ）とする。ステップＳ３１０９において、このＢ（ｔ）をベース音の時間推移として出力する。

【０１３９】

各シーンについて、ステップＳ３１０１ないしステップＳ３１０９の処理が終了すると、ステップＳ３１１０ないしステップＳ３１１２において、任意の２シーンのベース音について、類似度を算出する。

まずステップＳ３１１０において、所定の時刻間において、（式３－６）においてコストｄ（ｉ，ｊ）を決定するために、ベース音の一致不一致を算出する。次に、ステップＳ３１１１において、（式３－６）および（式３－７）に従って、ＤＴＷにおけるコストｄ（ｉ，ｊ）および$C_{i,j}$（ｂ，ａ）を設定する。ステップＳ３１１２において、ＤＴＷによる類似度を算出する。

【０１４０】

（他楽器に基づく類似度算出）

オーディオ信号類似度算出部２４において、他楽器に基づく類似度算出処理を説明する。この処理は、図７のステップＳ３０２および図９に相当する。

一般的な音楽の構成では、主にベース音が楽曲の最低音となるため、その他の楽器音は

ベース音の周波数域より高い周波数を示す。また、ベース音より高い周波数域で、各音名は図３２の周波数を持ち、各周波数の2^k（ｋ＝１，２，・・・）倍の周波数も同一の音名として扱われる。

そこで、本発明の最良の実施の形態では、ベース以外の楽器音が示すエネルギーを、ベース音より高く、かつ音名を持つ周波数のエネルギーとする。さらに、各音名が示す周波数のエネルギーには、図３２の2^k倍の周波数が示すエネルギーの和を用いる。これにより、本発明の最良の実施の形態では、複数の楽器による倍音構造を軽減し、音高の推定が困難な周波数域に存在する楽器音についても類似度算出に用いることを可能とする。

このように、ある音階Ｘ（例えば、Ｃ、Ｃ＃、Ｄ、またはＨ等）について注目するとき、その音は、１オクターブ上、２オクターブ上と、オクターブ単位で同様に存在する。ここで、ある音階の周波数をｆｘと表す場合、図３３に示すように、１オクターブ上、２オクターブ上・・・の各音は、それぞれ、２ｆｘ、４ｆｘ・・・で与えられる。

以下で詳細を説明する。なお、オーディオ信号は信号長Ｔ秒、サンプリングレートf_sとし、時刻ｔ（０≦ｔ≦Ｔ））、周波数ｆに対するエネルギーをパワースペクトルより算出し、Ｐ（ｔ、ｆ）と表す。

【０１４１】

他楽器に基づく類似度算出においては、図３４に示すように、まず、音名が示す周波数のエネルギーが抽出される。具体的には、後述する（式４－１）のエネルギーＰｘ（ｔ）をＧ２１に示す。Ｇ２２に示すように、このエネルギーP_X（ｔ）から、それぞれの音階をあてはめる。さらに、Ｇ２３に示すように、音階ごとに、ヒストグラムを算出する。Ｇ２３においては、各音階について、４オクターブ分のパワースペクトルを加算した結果、具体的には（式４－１）により得られるＰｘ（ｔ）を示している。

図３４に示す処理において、ＣからＨまでの各１２音階について、４オクターブ分の周波数のエネルギーP_C（ｔ）、$P_{C\#}$（ｔ）・・・P_H（ｔ）を算出する。

図３４においては、パワースペクトルから音階をあてはめ、その後、ベース音の音階を選択する場合について説明したが、この方法には限られない。具体的には、パワースペクトルから、周波数毎のヒストグラムを取得し、最大値の周波数から音階を取得しても良い。

【０１４２】

具体的なアルゴリズムを以下に示す。尚、各処理は図９の各ステップに対応する。

【０１４３】

まず、音名が示す周波数のエネルギーの算出処理を説明する。この処理は、図９のステップＳ３２１に相当する。

パワースペクトルから、各音名が示す周波数のエネルギーを算出する。図３２において音名Ｘに対応する周波数をf_Xとして、音名Ｘが示す周波数のエネルギーP_X（ｔ）を次式で定義する。

【数４９】

$$P_X(t) = \sum_{k=1}^{K} P(t, f_X \cdot 2^k) \qquad \text{（式４－１）}$$

ただし、Ｋは

【数５０】

$$\log_2 \frac{f_s}{2f_X}$$

を越えない任意の整数とする。（式４－１）により各音名が示す周波数のエネルギーを定義することで、低周波数域に存在する音の倍音の影響が軽減可能となる。

【０１４４】

次に、エネルギー割合の算出処理を説明する。この処理は、図９のステップＳ３２２に相当する。

音名が示す周波数のエネルギーの算出処理で得られた各音名が示す周波数のエネルギーを全周波数域に対するエネルギーの割合で表現する。これにより、音名毎に時間軸方向での比較が可能となり、推移を得ることが可能となる。音名Ｘが示す周波数のエネルギーの割合ｐｘ（ｔ）は次式で示される。

【数５１】

$$px(t) = \frac{P_X(t)}{\int_0^{\frac{f_s}{2}} P(t,f)df} \qquad \text{（式４－２）}$$

以上を全てのｔ、Ｘについて施し、得られたｐｘ（ｔ）をベース以外の楽器音におけるエネルギーの推移として用いる。

【０１４５】

次に、ＤＴＷを用いた音名エネルギー割合の類似度算出処理を説明する。この処理は、図９のステップＳ３２３に相当する。

データベース中の全ての二映像間においてオーディオ信号のベース以外の楽器音のエネルギーを算出し、それぞれpx_r（ｔ）、px_i（ｔ）と表す。これらを用いて各音名毎にＤＴＷによる類似度算出を行う。したがって、類似度は音名の数である１２だけ得られる。そこで、ベース以外の楽器音の類似度は音名毎に得られた類似度の和により定義する。すなわち、音名Ｘについて得られる類似度をDa_xとすると、ベース以外の楽器による音の類似度Ｄａは次式で表される。

【数５２】

$$D_a = D_{a_C} + D_{a_{C\sharp}} + D_{a_D} + D_{a_{D\sharp}} + D_{a_E} + D_{a_F} + D_{a_{F\sharp}} + D_{a_G} + D_{a_{G\sharp}} + D_{a_A} + D_{a_B} + D_{a_H} \qquad \text{（式４－３）}$$

なお、ＤＴＷによる類似度算出に用いるコストは以下のように設定する。

【数５３】

$$d(j,i) = |px_i(j) - px_r(i)| \qquad \text{（式４－４）}$$

$$c_{j,i}(b,a) = \begin{cases} \gamma & \text{if } (b,a) = (j-1,i),(j,i-1) \\ 0 & \text{otherwise} \end{cases} \qquad \text{（式４－５）}$$

（式４－３）により、全ての音名が示す周波数のエネルギーの推移を用いた類似度算出が可能となる。また、（式４－４）に示すコストを設定することで、エネルギーの大きな周波数に対応する音名が、類似度全体に与える影響を増加する。これにより、メロディーを構成する主要な周波数成分を反映した類似度算出が可能となる。

【０１４６】

ここで、図３５を参照して、本発明の最良の実施の形態に係る他楽器に基づく類似度算出処理を説明する。

まず、動画データベース１１の各シーンについて、ステップＳ３２０１ないしステップＳ３２０６の処理が実行される。

ステップＳ３２０１において、１つのシーンにフーリエ変換をする。ステップＳ３２０

２において、各時刻のパワースペクトルを算出し、ステップＳ３２０３において、音名Ｘが示す周波数エネルギーＰｘ（ｔ）を算出して、ｐｘ（ｔ）を算出する。

一方、ステップＳ３２０４において、全周波数のエネルギーを算出する。さらにステップＳ３２０５において、ステップＳ３２０３で算出された音名が示す周波数のエネルギーＰｘ（ｔ）と、ステップＳ３２０４で算出された全周波数のエネルギーに基づいて、エネルギーの割合ｐｘ（ｔ）を算出する。ステップＳ３２０６において、このエネルギーの割合ｐｘ（ｔ）を、ベース以外の楽器音におけるエネルギーとして出力する。

【０１４７】

各シーンについて、ステップＳ３２０１ないしステップＳ３２０６の処理が終了すると、ステップＳ３２０７ないしステップＳ３２１０において、任意の２シーンのエネルギーの割合について、類似度を算出する。

まずステップＳ３２０７において、ＤＴＷにおけるコストｄ（ｉ，ｊ）および$C_{i,j}$（ｂ，ａ）を設定し、ステップＳ３２０８において、ＤＴＷによって、各音名における２シーン間の類似度を算出する。ステップＳ３２０９において、ステップＳ３２０８において算出された全音名の類似度の和Ｄａを算出する。ステップＳ３２１０において、この和Ｄａを、ベース音以外の楽器による音の類似度として出力する。

【０１４８】

（リズムに基づく類似度算出）

オーディオ信号類似度算出部２４において、リズムに基づく類似度算出処理を説明する。この処理は、図７のステップＳ３０３および図１０に相当する。

楽曲のテンポに代表される細かなリズムは、打楽器を含めた全ての楽器における発音時刻の間隔により定義される。また、大域的なリズムは、連続して発音される楽器音により構成される楽句や楽節等が出現する間隔により決定すると考えられる。したがって、リズムは上記の時間間隔によって与えられるため、一定の区間内では楽曲の時刻に依存しない。そこで、本発明の最良の実施の形態ではオーディオ信号が弱定常性であると仮定し、自己相関関数によりリズム関数を表現する。これにより、本発明の最良の実施の形態は、オーディオ信号を用いて楽曲のリズムを一意に表現し、リズムに基づく類似度の算出を可能とする。

具体的なアルゴリズムを以下に示す。尚、各処理は図１０の各ステップに対応する。

【０１４９】

まず、２分割フィルタバンクによる低周波・高周波成分の算出処理を説明する。この処理は、図１０のステップＳ３３１に相当する。

２分割フィルタバンクによる低周波・高周波成分の算出処理においては、２分割フィルタバンクを用いて、処理対象信号を階層的に高周波、および低周波へＵ回だけ分解し、高周波成分を含む側の信号をx_u（ｎ）（ｕ＝１，・・・・Ｕ；，ｎ＝１，・・・・N_U）と表す。ここで、N_Uはx_uの信号長を示す。このようにして得られた各信号は、それぞれ異なる周波数帯を示すため、含まれる楽器の種類も異なると考えられる。したがって、得られた信号毎のリズムを推定し、結果を統合することで、複数種類の楽器音によるリズムが推定可能となる。

図３６を参照して、２分割フィルタバンクによる低周波・高周波成分の算出処理を説明する。ステップＳ３３０１において、２分割フィルタにより、低周波成分と高周波成分に分ける。次に、ステップＳ３３０１で分割された低周波成分を、ステップＳ３３０２において、さらに低周波成分と高周波成分に分ける。一方、ステップＳ３３０１で分割された高周波成分を、ステップＳ３３０３において、さらに低周波成分と高周波成分に分ける。このように所定回数（Ｕ回）だけ、２分割フィルタ処理を繰り返し、ステップＳ３３０４において、高周波成分を含む側の信号x_u（ｎ）を出力する。図３７に示すように、入力された信号の高周波成分が、２分割フィルタバンクによる低周波・高周波成分の算出処理によって出力されている。

【０１５０】

次に、包絡線の検波処理を説明する。この処理は、図１０のステップＳ３３２ないしス

テップＳ３３５に相当する。以下の１）ないし４）は、それぞれ図１０のステップＳ３３２ないしステップＳ３３５である。

２分割フィルタバンクによる低周波・高周波成分の算出処理で得られた信号x_u（ｎ）から包絡線を検波する。包絡線は、信号の各時刻における接線を共有する曲線であり、信号の概形を得ることを可能とする。したがって、包絡線検波により、楽器の発音に伴って音量が増加する時刻が推定可能となる。以下に包絡線を検波する処理の詳細を示す。

【０１５１】

１）全波整流

（式５－１）に示す全波整流を施し、信号y_{1u}（ｎ）（ｕ＝１，・・・・，Ｕ；，ｎ＝１，・・・・，N_U）を得る。

【数５４】

$$y_{1_u}(n) = |x_u(n)| \qquad \text{（式５－１）}$$

全波整流を施すことにより、図３８（ａ）に示す波形から、図３８（ｂ）に示す波形を得ることができる。

【０１５２】

２）低域通過フィルタの適用

１）全波整流で得られた信号y_{1u}（ｎ）に対し、（式５－２）に示す単純な低域通過フィルタを施し、信号y_{2u}（ｎ）（ｕ＝１，・・・・，Ｕ；，ｎ＝１，・・・・，N_u）を得る。

【数５５】

$$y_{2_u}(n) = (1-\alpha)y_{1_u}(n) + \alpha y_{2_u}(n-1) \qquad \text{（式５－２）}$$

ただし、αは遮断周波数を定める定数である。

低域通過フィルタを通すことにより、低周波数の信号から、図３９（ａ）に示す信号が出力される。具体的には、ローパスフィルタを通しても信号は変化せず、ハイパスフィルタを通すことにより、小刻みな波の信号が出力される。また、低域通過フィルタを通すことにより、高周波数の信号から、図３９（ｂ）に示す信号が出力される。具体的には、ハイパスフィルタを通しても信号は変化せず、ローパスフィルタを通すことにより、なだらかな波の信号が出力される。

【０１５３】

３）ダウンサンプリング

２）低域通過フィルタの適用で得られた信号y_{2u}（ｎ）に対し、（式５－３）に示すダウンサンプリングを施し、信号

【数５６】

$$y_{3_u}(n)\ \left(u=1,\ldots,U;\ n=1,\ldots,\frac{N_u}{s}\right)$$

を得る。

【数５７】

$$y_{3_u}(n) = y_{2_u}(sn) \qquad \text{（式５－３）}$$

ただし、ｓはサンプリング間隔を定める定数である。

ダウンサンプリング処理をすることにより、図４０（ａ）に示す信号から間引きされ、図４０（ｂ）に示す信号が出力される。

【０１５４】
４）平均値除去
３）ダウンサンプリングで得られた信号y_{3u}（ｎ）に（式５－４）を施し、信号の平均が０となる信号y_u（ｎ）（ｕ＝１，・・・・，Ｕ；，ｎ＝１，・・・・，N_u）を得る。
【数５８】

$$y_u(n) = y_{3_u}(n) - E[y_{3_u}(n)] \qquad (式５－４)$$

ただし、Ｅ［y_{3u}（ｎ）］は信号y_{3u}（ｎ）の平均値を示す。
平均値除去処理をすることにより、図４１（ａ）に示す信号から、図４１（ｂ）に示す信号が出力される。
【０１５５】
次に、自己相関関数の算出処理を説明する。この処理は、図１０のステップＳ３３６に相当する。
包絡線の検波処理で得られた信号y_u（ｎ）を2^{u-1}倍のサンプリングレートにアップサンプリングし、信号長を等しくした後、すべてを加算する。これにより得られた信号をｙ（ｎ）（ｎ＝１，・・・・，N_1）とする。ただし、N_1は信号長を表す。さらに、ｙ（ｎ）を用いて、自己相関関数ｚ（ｍ）（ｍ＝０，・・・，N_1-1）を次式により算出する。
【数５９】

$$z(m) = \frac{1}{N_1}\sum_{n}^{N_1} y(n)y(n-m) \qquad (式５－５)$$

自己相関について、図４２を参照して説明する。自己相関関数とは、信号とそれ自身をｍだけ移動（シフト）した信号との相関を表しており、ｍ＝０のときに最大となる関数である。ここで、信号に繰り返しが存在する場合、その倍数位置（ｍ）においてｍ＝０の場合と同様に高い値を持つことが知られており、そのピークを検出することにより、繰り返しを見つけることが可能となる。
自己相関を用いることにより、信号に含まれる繰り返しパターンを探し、ノイズに含まれる周期的な信号を抽出することが容易となる。
このように、本発明の最良の実施の形態においては、様々なオーディオ信号の特徴を、自己相関関数から抽出されるファクターによって表すことができる。
【０１５６】
次に、ＤＴＷを用いたリズム関数の類似度の算出処理を説明する。この処理は、図１０のステップＳ３３７に相当する。
本発明の最良の実施の形態では、時刻ｔから一定時間の信号を用いて算出される上記の自己相関関数を時刻ｔにおけるリズム関数とし、楽曲間の類似度算出に利用する。リズム関数は、複数の周波数域において音量が増加する時刻の周期を表現するため、複数の楽器音によるリズムを含む。このため、本発明の最良の実施の形態では、局所的なリズムから大域的なリズムを含む複数のリズムを用いて楽曲の類似度算出を可能とする。
次に、得られたリズム関数を用いて楽曲の類似度を算出する。そこで、まずリズムの類似度について考察する。楽曲におけるリズムは、演奏者や編曲者によって変動する。このため、同一の楽曲であっても、楽曲の全体、または一部が異なる速度で演奏される場合が存在する。このため、リズムに基づいて楽曲間の類似度を定義するには、リズムの変動を許容する必要がある。そこで、本発明の最良の実施の形態では、リズムに基づく類似度の

算出に、メロディーに基づく類似度と同様にＤＴＷを利用する。これにより、本発明の最良の実施の形態では、演奏者や編曲者によってリズムが変更された楽曲を変更前の楽曲と同一と判断可能とする。また、楽曲自体が異なる場合にも、類似するリズムを示す楽曲を類似楽曲として判断可能とする。

【０１５７】

図４３を参照して、自己相関関数の算出処理およびＤＴＷを用いたリズム関数の類似度の算出処理を説明する。

ステップＳ３４０１において、包絡線が入力されると、処理対象のシーンの楽曲と参照楽曲について、ステップＳ３４０２ないしステップＳ３４０４の処理が繰り返される。

まず、ステップＳ３４０２において、対象シーンのオーディオ信号に基づいて出力された包絡線をアップサンプリングする。ステップＳ３４０３において、y_u（ｎ）をｕに対して全て加算し、ｙ（ｎ）を取得し、ステップＳ３４０４において、ｙ（ｎ）の自己相関関数Ｚ（ｍ）を算出する。

一方、参照楽曲における自己関数Ｚ（ｍ）が算出される。ステップＳ３４０５において、処理対象シーンの楽曲における自己関数Ｚ（ｍ）をリズム関数として、参照楽曲における自己関数Ｚ（ｍ）との類似度を、ＤＴＷを適用して算出して、ステップＳ３４０６において、類似度を出力する。

【０１５８】

表示部２８は、ビデオ信号類似度表示部２９と、オーディオ信号類似度表示部３０を備える。

表示部２８は、検索部２５による検索結果を表示するとともに、映像の再生、検索、および、検索・分類結果の可視化を行うユーザインターフェースである。表示部２８のユーザインターフェースは、下記の各機能を有していることが好ましい。

・映像の再生

動画データベース１１に記憶されたの映像データを任意の位置に配置し再生する。このとき、再生中の映像の現在のフレーム位置より後方にあるフレームの画像を、３次元空間上で、映像の後方に配置し表示する。

それぞれの画像を配置する位置を常に更新することで、画像が奥から手前に向かって流れているような視覚効果を得ることができる。

・シーン単位の頭出し

シーン分割部２１によって分割されたシーンを単位とした頭出しを行う。ユーザの操作により再生中のシーンの前後シーンの開始位置へ動画像のフレーム位置を移動する。

・検索結果の表示

映像の再生中に検索操作を行うことで、検索部２５によって類似シーン検索を行い、検索結果を表示する。検索部２５による類似シーンの検索は、分類部によって求められた類似度に基づいて行う。表示部２８は、クエリシーンとの類似度が一定の閾値よりも小さいシーンを、動画データベース１１から抽出して、検索結果として表示する。

表示する際はクエリシーンの表示位置を原点とした３次元空間で表示する。このとき検索結果の各シーンについて、類似度と対応した座標をそれぞれのシーンに与える。それらを、図４４に示す透視変換を行うことにより、検索結果の各シーンの表示位置および大きさを決定する。

ただし、分類部２２のビデオ信号類似度算出部２３において映像情報に注目した分類のアルゴリズムを用いた場合、３次元空間上の軸は、３次元DTWによって得られる３つの座標となる。また、分類部２２のオーディオ信号類似度算出部２４において音楽情報に注目した分類のアルゴリズムを用いた場合、３次元空間上の軸はそれぞれ、ベース音に基づく類似度、他の楽器に基づく類似度、リズムに基づく類似度となる。

これにより、検索結果の中でクエリシーンとより類似したシーンがクエリシーンの近くに表示される。また、表示された検索結果の映像に対しても同様に、その映像を選択することによって、その時刻に再生中のシーンをクエリとした類似シーン検索を行うことができる。

このように本発明では、映像情報に注目した分類および音楽情報に注目した分類のそれぞれについて、表示装置に表示する座標を変更させることにより、さらに分類パラメータを重み付けした分類結果を取得することができる。例えば、音楽情報に着目した分類について、リズムに基づく類似度が高い座標には、リズムの類似度が高く、ベース音や他の楽器に基づく類似度が低いシーンが表示される。

【０１５９】

（効果）

このような本発明の最良の実施の形態に係る動画検索装置１によれば、映像の構成要素であるオーディオ信号およびビデオ信号を用いて映像間の類似度を算出し、それらの分類結果を３次元の空間上に可視化することができる。本発明の最良の実施の形態では、映像に対して楽曲に基づいた類似度の算出、および音響・ビジュアル信号の双方に基づいた類似度の算出の２つの類似度算出機能を持ち、映像の異なる要素に注目することで、ユーザの好みに応じた検索モードを実現することができる。さらに、この機能を用いることで、クエリ映像を与えた場合に自動で類似映像の検索をすることができる。また、クエリ映像が存在しない場合、データベース中の映像の自動分類を行い、注目する映像に対して類似する映像をユーザに呈示することができる。

【０１６０】

さらに、本発明の最良の実施の形態では映像間の類似度に基づいて、３次元の空間上に映像を配置することで、空間の距離によって映像の類似性を理解することが可能なユーザインターフェースを実現することができる。具体的に、映像情報に注目した検索・分類のアルゴリズムを用いた場合、３次元空間上の軸は３次元ＤＴＷによって得られる３つの座標とし、音楽情報に注目した検索・分類のアルゴリズムを用いた場合、ベース音に基づく類似度、他の楽器に基づく類似度、リズムに基づく類似度とした。これにより、ユーザは３次元空間上において、映像および音楽のどの部分が似ているかを主観的に評価することができる。

【０１６１】

（変形例）

図４５に示す本発明の変形例に係る動画検索装置１ａは、図１に示す本発明の最良の実施の形態に係る動画検索装置１と比べて、検索部２５ａおよび表示部２８ａが異なる。本発明の最良の実施の形態に係る検索部２５では、ビデオ信号類似度検索部２６が、ビデオ信号類似度データ１２に基づいてクエリ動画データに類似する動画データを検索するとともに、オーディオ信号類似度検索部２７が、オーディオ信号類似度データ１３に基づいてクエリ動画データに類似する動画データを検索する。さらに、本発明の最良の実施の形態に係る表示部２８では、ビデオ信号類似度表示部２９が、ビデオ信号類似度検索部２６による検索結果を画面に表示するとともに、オーディオ信号類似度表示部３０が、オーディオ信号類似度検索部２７による検索結果を画面に表示する。

一方、本発明の変形例においては、検索部２５ａが、ビデオ信号類似度データ１２およびオーディオ類似度データ１３に基づいてクエリ動画データに類似する動画データを検索し、表示部２８ａが、検索結果を画面に表示する。具体的には、検索部２５ａは、ユーザから嗜好データが入力されると、その嗜好データに従って、各シーンに対するビデオ信号類似度データ１２およびオーディオ類似度データ１３の類似度の割合を決定して、その割合に基づいた検索結果を取得する。表示部２８ａはさらに、検索部２５ａによって取得された検索結果を、画面に表示する。

これにより、本発明の変形例においては、一つの操作で、複数のパラメータを考慮して算出された分類結果を出力することができる。

【０１６２】

検索部２５ａは、ユーザによる入力装置等の操作によって、ビデオ信号類似度とオーディオ信号類似度に対する嗜好の割合である嗜好データを取得する。さらに検索部２５ａは、ビデオ信号類似度データ１２およびオーディオ信号類似度データ１３に基づいて、ビジュアル信号の特徴量とオーディオ信号の特徴量から算出されたシーン間の類似度と、オー

ディオ信号のベース音に基づく類似度と、ベースを除く楽器に基づく類似度と、リズムに基づく類似度とに対する重み係数を決定する。さらに検索部２５ａは、各シーンの各類似度にこの重み係数を乗算して統合された類似度に基づいて、シーン間の統合された類似度が一定の閾値よりも小さいシーンを検索する。

表示部２８ａは、検索部２５ａによって検索された各シーンについて該統合された類似度に対応する座標を取得して表示する。

【０１６３】

ここで、表示部２８ａにおいて各検索結果に与えられる３次元座標は、以下のように決定される。Ｘ座標は音楽情報に注目した類似度算出部において算出されたシーン間の類似度に対応する。Ｙ座標は映像情報に注目した類似度算出部において算出されたシーン間の類似度に対応する。Ｚ座標は嗜好パラメータを基に求められた最終的なシーン間の類似度に対応する。ただし、これらの座標は全ての検索結果が画面内に表示され、かつ、検索結果同士が重なり合わないように調整される。

【０１６４】

嗜好データを取得する際、例えば、検索部２５ａは、図４６に示す表示画面Ｐ２０１を、表示装置に表示する。表示画面Ｐ２０１は、嗜好入力部Ａ２０１を備えている。嗜好入力部Ａ２０１は、分類部２２のビデオ信号類似度算出部２３およびオーディオ信号類似度算出部２４によって算出されたビデオ信号類似度データ１２およびオーディオ信号類似度データ１３について、各類似度データをどのような重みで表示するかを決定するための嗜好パラメータの入力を受け付ける。嗜好入力部Ａ２０１は、例えば、マウスによってクリックされた座標に基づいて重みが算出される。

【０１６５】

嗜好入力部Ａ２０１は、例えば、図４７に示すような軸を有している。図４７においては、軸Ｐｘおよび軸Ｐｙで分割される４つの領域を有する。右側には、ビデオ信号類似度データ１２に関連する類似度が関連づけられており、右上のセルには、音響による類似度が、右下のセルには、動画像による類似度が、関連づけられている。一方、左側には、オーディオ信号類似度データ１３に関連する類似度が関連づけられており、左上のセルには、リズムによる類似度が、左下のセルには、他楽器およびベースによる類似度が関連づけられている。

嗜好入力部Ａ２０１のいずれかに、ユーザがマウスでクリックすると、クリック点のＰｘの座標に基づいて、検索部２５ａは、ビデオ信号類似度算出部２３によって算出されたビデオ信号類似度データ１２と、オーディオ信号類似度算出部２４によって算出されたオーディオ信号類似度データ１３のそれぞれを重み付けする。さらに、検索部２５ａは、クリック点のＰｙの座標に基づいて、各類似度データについて、各パラメータの重み付けを決定する。具体的には、検索部２５ａは、ビデオ信号類似度データ１２の音響による類似度と、動画像による類似度の各重みを決定するとともに、オーディオ信号類似度データ１３のリズムによる類似度と、他楽器およびベースによる類似度の各重みを決定する。

【０１６６】

ここで、図４８を参照して、本発明の変形例に係る検索部２５ａおよび表示部２８ａの処理を説明する。

図４８（ａ）を参照して、検索部２５ａによる処理を説明する。まず、記憶装置１０７からビデオ信号類似度データ１２およびオーディオ信号類似度データ１３が読み出される。さらに、シーン分割部２１によって分割された各シーンについて、ビデオ信号類似度データ１２から、ステップＳ６０１においてクエリ動画シーンとのビジュアル信号の類似度が取得されるとともに、ステップＳ６０２においてクエリ動画シーンとのオーディオ信号の類似度が取得される。さらに、シーン分割部２１によって分割された各シーンについて、オーディオ信号類似度データ１３から、ステップＳ６０３において、クエリ動画シーンとのベース音に基づく類似度が取得される。ステップＳ６０４においてクエリ動画シーンとの非ベース音に基づく類似度が取得される。ステップＳ６０５においてクエリ動画シーンとのリズムに基づく類似度が取得される。

【０１６７】
つぎに、ステップＳ６０６において、嗜好入力部Ａ２０１における座標から、嗜好パラメータを取得し、ステップＳ６０７において、嗜好パラメータに基づいて、重み係数を算出する。つぎにステップＳ６０８において、ステップＳ６０１およびステップＳ６０５で取得された類似度のうち、所定値以上の類似度のシーンを検索する。ここでは、類似度に基づいて閾値処理する場合について説明するが、類似度が高いものから所定数のシーンが検索されても良い。
【０１６８】
図４８（ｂ）を参照して、表示部２８ａの処理を説明する。ステップＳ６５１において、ステップ検索部２５ａによって検索された各シーンについて、三次元空間における座標が算出される。ステップＳ６５２において、ステップＳ６５１で算出された各シーンの座標が透視変換され、各シーンの動画像のフレームの大きさが決定される。ステップＳ６５３において、表示装置に表示される。
【０１６９】
このように、本発明の変形例に係る検索部２５ａにおいては、類似シーン検索を行う際に、映像情報に注目したビデオ信号類似度算出部２３において算出されたシーン間の類似度と、音楽情報に注目したオーディオ信号類似度算出部２４において算出されたシーン間の類似度のうち、どの要素を重視して検索を行うかをユーザが指定することができる。
ユーザが指定するのは図４７に示されるような二次元の嗜好パラメータであり、この嗜好パラメータをもとに、それぞれの類似度に対する重み係数が決定される。そして重み係数を掛けた類似度の総和を最終的なシーン間の類似度とし、これに基づいて類似シーンの検索を行う。
ここで、ユーザが指定する嗜好パラメータP_x，P_yと最終的なシーン間の類似度Ｄの関係は以下の式で示される。
【数６０】

$$D = W_{sv}D_{sv} + W_{sa}D_{sa} + W_bD_b + W_aD_a + W_rD_r$$
$$W_{sv} = P_xP_y$$
$$W_{sa} = P_x(1-P_y)$$
$$W_b = (1-P_x)(1-P_y)$$
$$W_a = \frac{(1-P_x)P_y}{2}$$
$$W_r = \frac{(1-P_x)P_y}{2}$$

ただし、D_{sv}，D_{sa}は映像情報に注目した類似度算出部において算出されるシーン間の類似度である。D_{sv}はビジュアル信号に基づく類似度、D_{sa}はオーディオ信号に基づく類似度である。また、D_b，D_a，D_γは音楽情報に注目した類似度算出部において算出されるシーン間の類似度でありD_bはベース音に基づく類似度、D_aは他楽器に基づく類似度、D_γはリズムに基づく類似度を示す。
【０１７０】
このような変形例に係る動画検索装置１ａによれば、複数のパラメータを複合して嗜好パラメータを生成し、その嗜好パラメータに合致するシーンを表示することができる。従って、ユーザに直感的に分かりやすい動画検索装置を提供することができる。
【０１７１】

（効果）
図４９ないし図５９を参照して、本発明の実施の形態に係る動画検索装置によるシミュレーション結果を説明する。このシミュレーションにおいては、動画データベース１１に、クエリシーンを含む動画データと、このクエリシーンに類似するシーンを含む約１０分間の動画データとを記憶している。本シミュレーションにおいては、このクエリシーンに類似するシーンを含む動画データを検索対象の動画データとし、この動画データに含まれる複数のシーンから、クエリシーンに類似するシーンを検索できるかをシミュレーションする。

【０１７２】
図４９ないし図５１は、分類部２２および検索部２５によるシミュレーション結果を示している。
図４９は、クエリシーンの動画データを示している。上段の画像は、動画データのビジュアル信号により構成された一定時間ごとのフレーム画像である。下段の画像は、動画データのオーディオ信号の波形である。

【０１７３】
図５０は、実験対象の動画データの各シーンについて、クエリシーンとの類似度を示した図である。図５０においては、横軸は、検索対象の動画データの開始位置からの時間で、縦軸は、クエリシーンとの類似度である。図５０において類似度がプロットされている位置が、検索対象の動画データのシーンの開始位置である。図５０において、類似度が約「１．０」になっているシーンが、クエリシーンと類似しているシーンである。実際に本シミュレーションにおいては、図４９に示したシーンと同じシーンが、類似度が高いシーンとして検索された。
図５１に示す図は、３次元ＤＴＷによって得られる３つの座標を示したものである。図５１に示したパス＃５は、上述したとおり、ビジュアル信号とオーディオ信号との両方の信号を類似部分としてそれぞれ対応付けを行う役割を持つパスである。
図５０に示す結果により、高精度にシーン間の類似度が算出されていることを確認することができる。また、図５１により、実施の形態で用いた３次元ＤＴＷにより、適切にシーン間の類似度の対応付けがなされていることを確認することができる。

【０１７４】
図５２ないし図５５は、ビデオ信号類似度算出部２３およびビデオ信号類似度検索部２６によるシミュレーション結果を示している。
図５２は、クエリシーンの動画データを示している。上段の画像は、動画データのビジュアル信号により構成された一定時間ごとのフレーム画像である。下段の画像は、動画データのオーディオ信号の波形である。一方、図５３は、検索対象の動画データに含まれるシーンを示している。図５２に示したクエリシーンのフレームＦ１３ないしＦ１７は、図５３に示した検索対象のシーンのフレームＦ２１ないしＦ２５と類似している。図５２に示したオーディオ信号と、図５３に示したオーディオ信号は、明らかに異なる。

【０１７５】
図５４は、実験対象の動画データの各シーンについて、クエリシーンとの類似度を示した図である。図５４においては、横軸は、検索対象の動画データの開始位置からの時間で、縦軸は、クエリシーンとの類似度である。図５４において類似度がプロットされている位置が、検索対象の動画データのシーンの開始位置である。図５４において、類似度が約「０．８」になっているシーンが、クエリシーンと類似しているシーンである。実際に本シミュレーションにおいては、類似度が約「０．８」であるシーンは、図５３に示したシーンである。このシーンが類似度が高いシーンとして検索された。
図５５に示す図は、３次元ＤＴＷによって得られる３つの座標を示したものである。図５５に示したパス＃１は、上述したとおり、クエリシーンのクリップの時間軸方向への伸縮を許容する役割を持つパスである。また、図５５に示したパス＃３は、ビジュアル信号を類似部分として対応付けを行う役割を持つ。
図５５に示す結果により、時間軸方向にずれたビジュアル信号についても、高精度にシ

ーン間の類似度が算出されていることを確認することができる。また、図５５により、実施の形態で用いた３次元ＤＴＷにより、適切にシーン間の類似度の対応付けがなされていることを確認することができる。

【０１７６】

図５６ないし図５９は、オーディオ信号類似度算出部２４およびオーディオ信号類似度検索部２７によるシミュレーション結果を示している。

図５６は、クエリシーンの動画データを示している。上段の画像は、動画データのビジュアル信号により構成された一定時間ごとのフレーム画像である。下段の画像は、動画データのオーディオ信号の波形である。一方、図５７は、検索対象の動画データに含まれるシーンを示している。図５６に示したクエリシーンのビジュアル信号により構成されたフレーム画像は、図５７に示した検索対象シーンのビジュアル信号により構成されたフレーム画像とは明らかに異なる。一方、図５６に示したクエリデータのオーディオ信号と、図５７に示した検索対象シーンのオーディオ信号は類似している。

【０１７７】

図５８は、実験対象の動画データの各シーンについて、クエリシーンとの類似度を示した図である。図５８においては、横軸は、検索対象の動画データの開始位置からの時間で、縦軸は、クエリシーンとの類似度である。図５８において類似度がプロットされている位置が、検索対象の動画データのシーンの開始位置である。図５８において、類似度が約「０．８」になっているシーンが、クエリシーンと類似しているシーンである。実際に本シミュレーションにおいては、類似度が約「０．８」であるシーンは、図５７に示したシーンである。このシーンが類似度が高いシーンとして検索された。

図５９に示す図は、３次元ＤＴＷによって得られる３つの座標を示したものである。図５９に示したパス＃４は、オーディオ信号を類似部分として対応付けを行う役割を持つ。

図５９に示す結果により、高精度にシーン間の類似度が算出されていることを確認することができる。また、図５９により、実施の形態で用いた３次元ＤＴＷにより、適切にシーン間の類似度の対応付けがなされていることを確認することができる。

【０１７８】

このように、本発明の実施の形態に係る動画検索装置によれば、動画データのビデオ信号を用いて、ビデオ信号が類似する画像を高精度に検索することができる。これにより、毎週、毎日放送される番組などにおいて、繰り返し同じ動画像で始まる特定のコーナーを、ビデオ信号を用いて高精度に検索することができる。また、タイトルに日付が入っている場合や音響に変化があるなどの場合でも、全体として類似している限り、類似度の高い画像として検索することができる。また、異なる番組においても、動画像や音響が類似するシーンを容易に検索することができる。

【０１７９】

また、本発明の実施の形態に係る動画検索装置によれば、動画データのオーディオ信号を用いて、オーディオ信号が類似する画像を高精度に検索することができる。また、本発明の実施の形態においては、ベース音およびメロディの動きに基づいて、楽曲の類似度を算出しているので、曲のテンポの変化や変調にかかわらず、類似する楽曲を検索することができる。

【０１８０】

（その他の実施の形態）

上記のように、本発明の最良の実施の形態および変形例によって記載したが、この開示の一部をなす論述及び図面はこの発明を限定するものであると理解すべきではない。この開示から当業者には様々な代替実施の形態、実施例及び運用技術が明らかとなる。

例えば、本発明の最良の実施の形態に記載した動画検索装置は、図１に示すように一つのハードウェア上に構成されても良いし、その機能や処理数に応じて複数のハードウェア上に構成されても良い。又、既存の情報システム上に実現されても良い。

【０１８１】

また、本発明の最良の実施の形態においては、動画検索装置１が、分類部２２、検索部

２５および表示部２８を備え、分類部２２が、ビデオ信号類似度算出部２３およびオーディオ信号類似度算出部２４を備える場合について説明している。ここで、本発明の最良の実施の形態においては、動画検索装置１が、ビデオ信号とオーディオ信号との両方に基づいて、類似度を算出、検索および表示する。具体的には、検索部２５が、ビデオ信号類似度検索部２６およびオーディオ信号類似度検索部２７を備え、分類部２２が、ビデオ信号類似度算出部２３およびオーディオ信号類似度算出部２４を備え、表示部２８が、ビデオ信号類似度表示部２９およびオーディオ信号類似度表示部３０を備える。

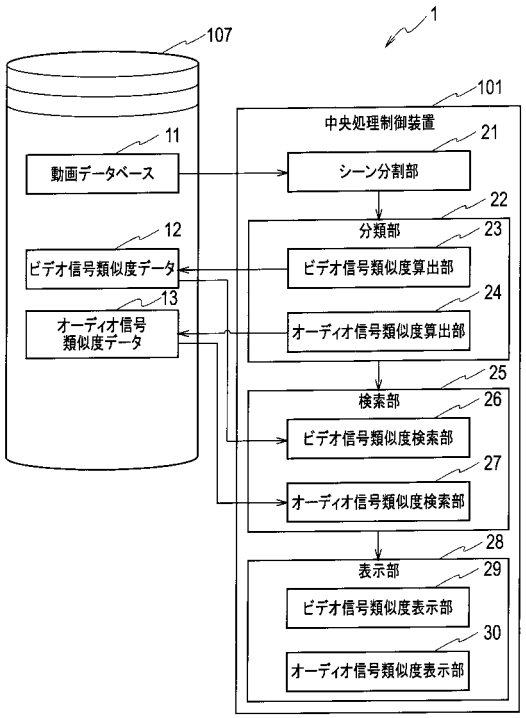
一方、ビデオ信号のみに基づいて類似度を算出、検索および表示する実施態様も考えられる。具体的には、分類部２２はビデオ信号類似度算出部２３を備え、検索部２５はビデオ信号類似度検索部２６を備え、表示部２８はビデオ信号類似度表示部２９を備える。

同様に、オーディオ信号のみに基づいて類似度を算出、検索および表示する実施態様も考えられる。具体的には、分類部２２はオーディオ信号類似度算出部２４を備え、検索部２５はオーディオ信号類似度検索部２７を備え、表示部２８はオーディオ信号類似度算出部３０を備える。

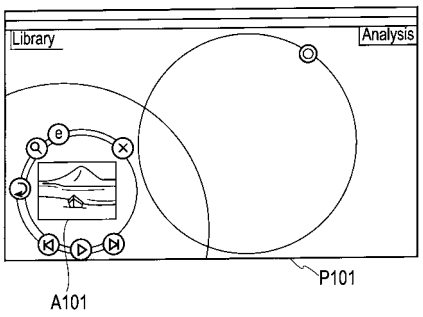
【０１８２】

本発明はここでは記載していない様々な実施の形態等を含むことは勿論である。従って、本発明の技術的範囲は上記の説明から妥当な特許請求の範囲に係る発明特定事項によってのみ定められるものである。

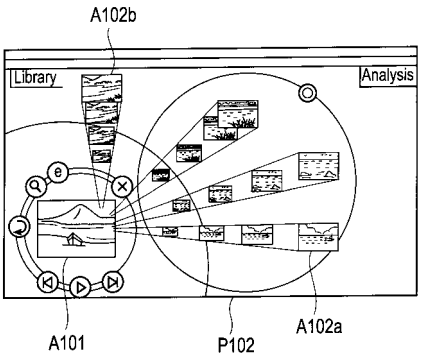
【図１】



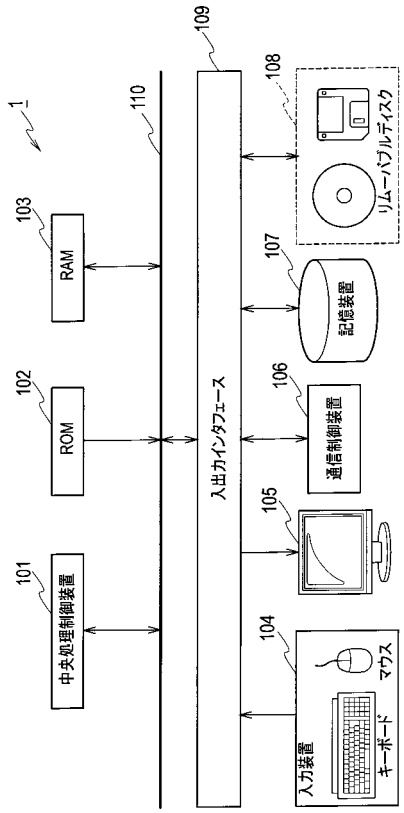
【図２】



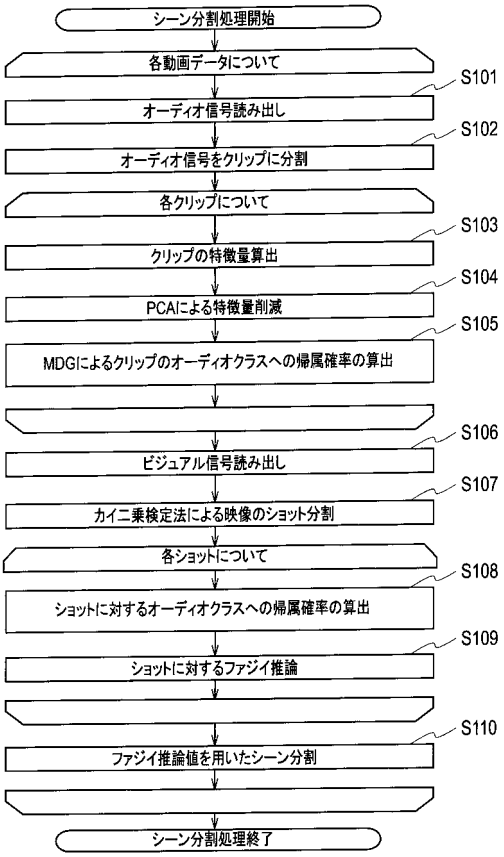
【図３】



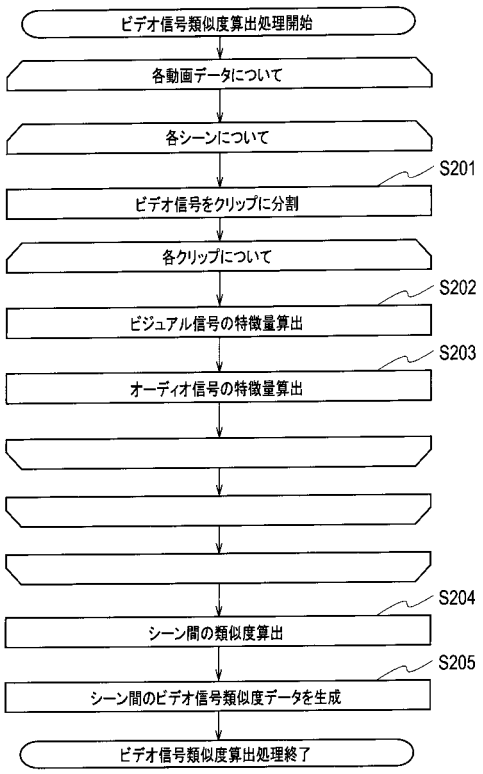
【図４】


1
101
中央処理制御装置
102
ROM
103
RAM
104
入力装置
マウス
キーボード
105
106
通信制御装置
107
記憶装置
108
リムーバブルディスク
109
入出力インタフェース
110

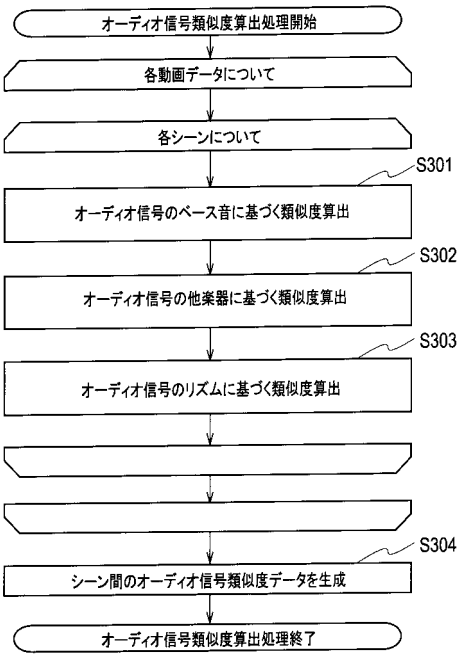
【図５】


シーン分割処理開始
各動画データについて
S101
オーディオ信号読み出し
S102
オーディオ信号をクリップに分割
各クリップについて
S103
クリップの特徴量算出
S104
PCAによる特徴量削減
S105
MDGによるクリップのオーディオクラスへの帰属確率の算出
S106
ビジュアル信号読み出し
S107
カイニ乗検定法による映像のショット分割
各ショットについて
S108
ショットに対するオーディオクラスへの帰属確率の算出
S109
ショットに対するファジイ推論
S110
ファジイ推論値を用いたシーン分割
シーン分割処理終了

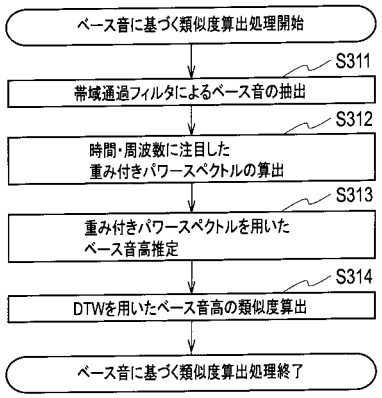
【図６】


ビデオ信号類似度算出処理開始
各動画データについて
各シーンについて
S201
ビデオ信号をクリップに分割
各クリップについて
S202
ビジュアル信号の特徴量算出
S203
オーディオ信号の特徴量算出
S204
シーン間の類似度算出
S205
シーン間のビデオ信号類似度データを生成
ビデオ信号類似度算出処理終了

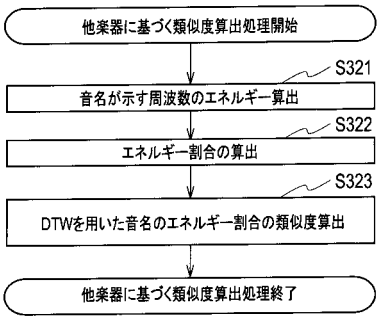
【図７】


オーディオ信号類似度算出処理開始
各動画データについて
各シーンについて
S301
オーディオ信号のベース音に基づく類似度算出
S302
オーディオ信号の他楽器に基づく類似度算出
S303
オーディオ信号のリズムに基づく類似度算出
S304
シーン間のオーディオ信号類似度データを生成
オーディオ信号類似度算出処理終了

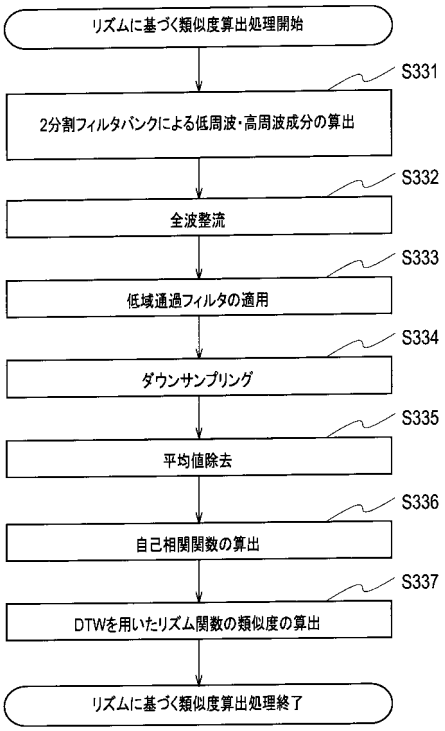
【図８】



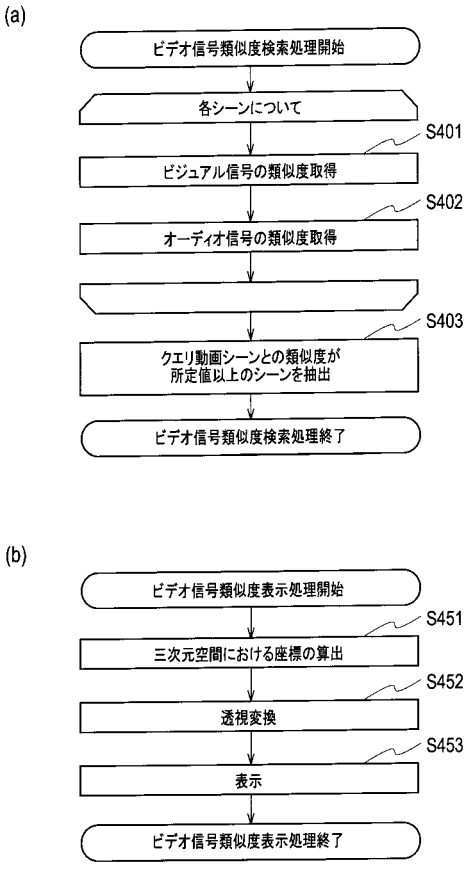
【図９】



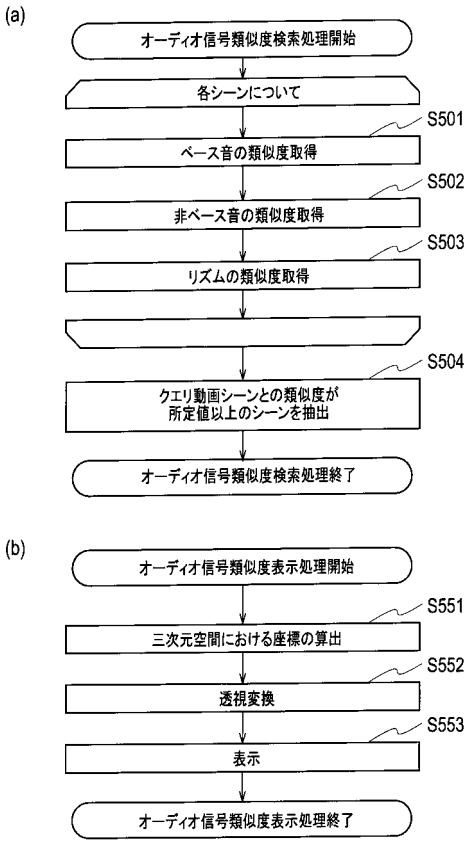
【図１０】



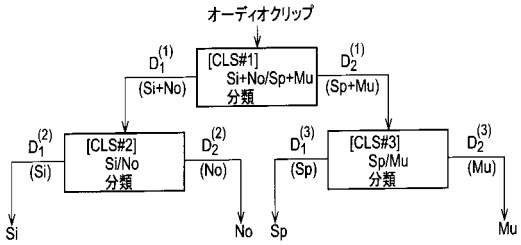
【図１１】



【図１２】



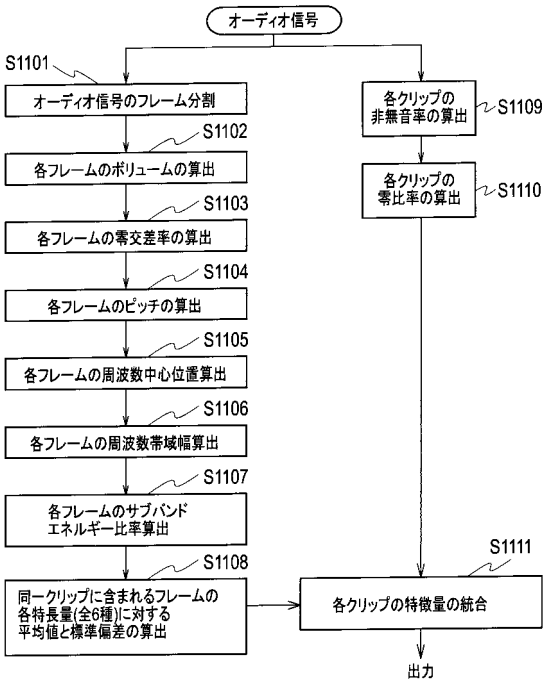
【図１３】



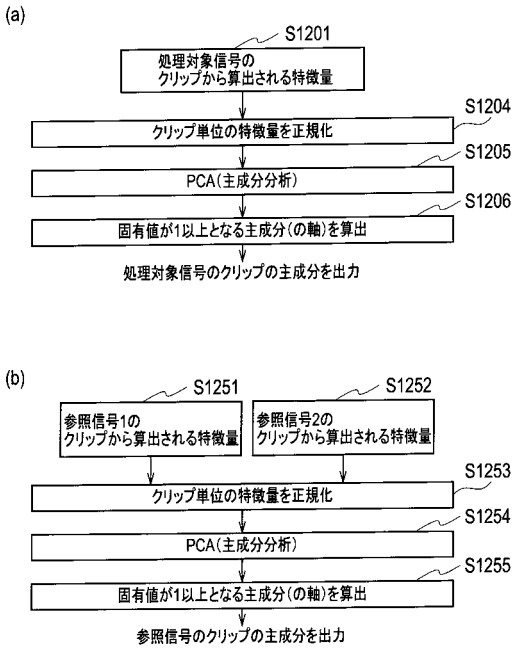
【図１４】

	参照信号1	参照信号2
CLS#1	Si,No	Sp,Mu
CLS#2	Si	No
CLS#3	Sp	Mu

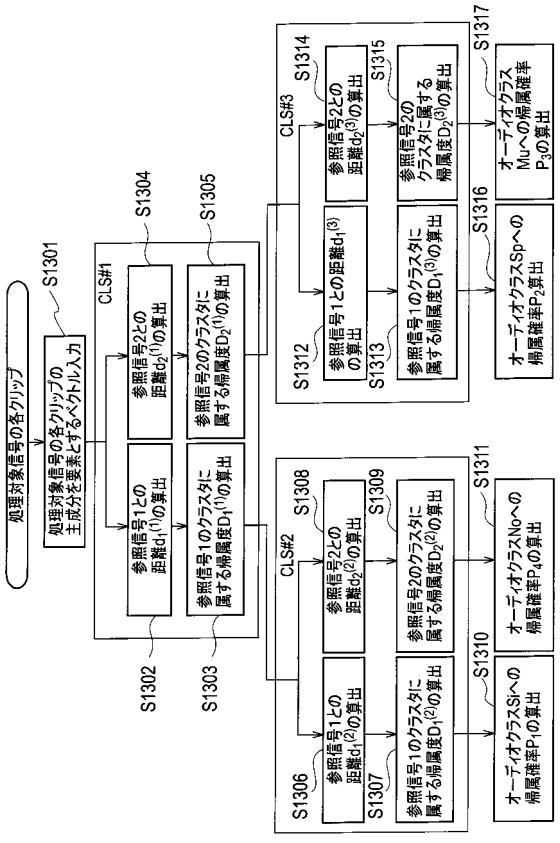
【図１５】



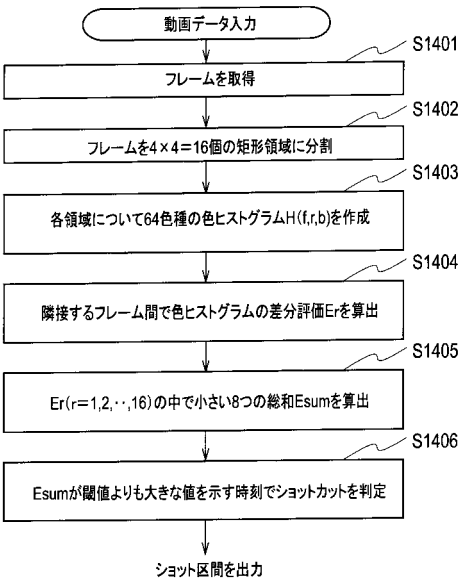
【図１６】



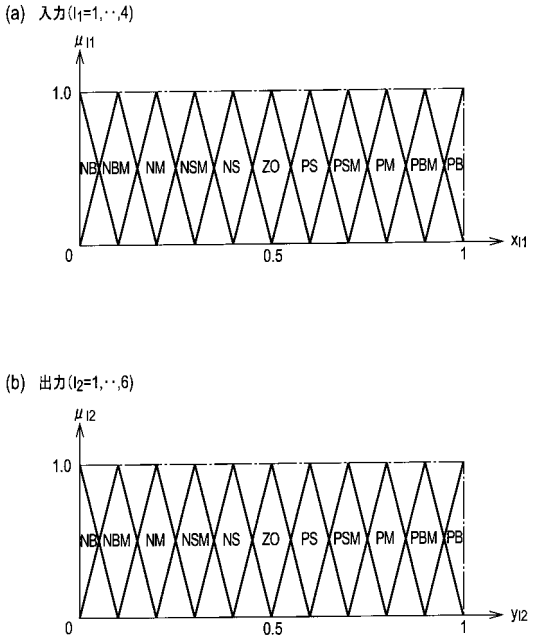
【図１７】



【図１８】



【図１９】



【図２０】

ファジィ制御規則1(Si,Sp,Mu,No)
(I_1=1,･･4;それぞれSi,Sp,Mu,Noに対応)

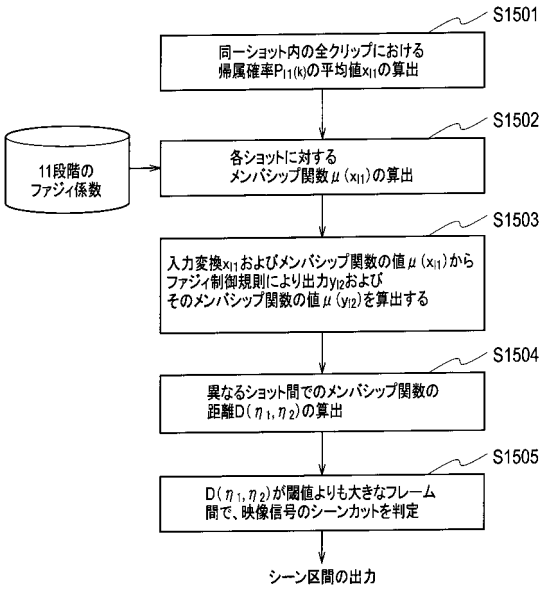
R^1_{I1} :	IF "x_{I1} is NB"	THEN	"y_{I1} is NB"
R^2_{I1} :	IF "x_{I1} is NBM"	THEN	"y_{I1} is NBM"
R^3_{I1} :	IF "x_{I1} is NM"	THEN	"y_{I1} is NM"
R^4_{I1} :	IF "x_{I1} is NSM"	THEN	"y_{I1} is NSM"
R^5_{I1} :	IF "x_{I1} is NS"	THEN	"y_{I1} is NS"
R^6_{I1} :	IF "x_{I1} is ZO"	THEN	"y_{I1} is ZO""
R^7_{I1} :	IF "x_{I1} is PS"	THEN	"y_{I1} is PS"
R^8_{I1} :	IF "x_{I1} is PSM"	THEN	"y_{I1} is PSM"
R^9_{I1} :	IF "x_{I1} is PM"	THEN	"y_{I1} is PM"
R^{10}_{I1} :	IF "x_{I1} is PBM"	THEN	"y_{I1} is PBM"
R^{11}_{I1} :	IF "x_{I1} is PB"	THEN	"y_{I1} is PB"

【図２１】

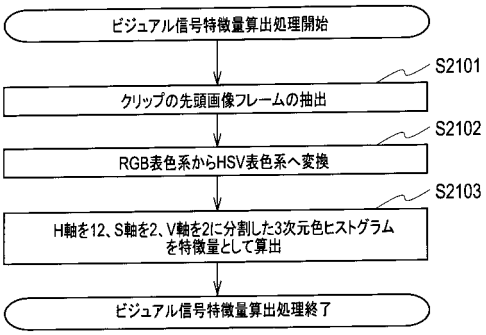
ファジィ制御規則2(SpMu,SpNo)

R^1_5 :	IF "x_1 is NB"and "x_3 is NBM"	and"x_4 is NBM"	THEN	"y_5 is NB"
R^2_5 :	IF "x_1 is NB"and "x_2 is PS"	and"x_3 is NSM"	THEN	"y_5 is PM"
R^3_5 :	IF "x_1 is NB"and "x_2 is PM"	and"x_3 is NBM"	THEN	"y_5 is PBM"
R^4_5 :	IF "x_1 is NB"and "x_2 is PSM"	and"x_3 is NM"	THEN	"y_5 is PB"
R^1_6 :	IF "x_1 is NB"and "x_3 is NBM"	and"x_4 is NBM"	THEN	"y_6 is NB"
R^2_6 :	IF "x_1 is NB"and "x_2 is PS"	and"x_4 is NSM"	THEN	"y_6 is PM"
R^3_6 :	IF "x_1 is NB"and "x_2 is PM"	and"x_4 is NBM"	THEN	"y_6 is PBM"
R^4_6 :	IF "x_1 is NB"and "x_2 is PSM"	and"x_4 is NM"	THEN	"y_6 is PB"

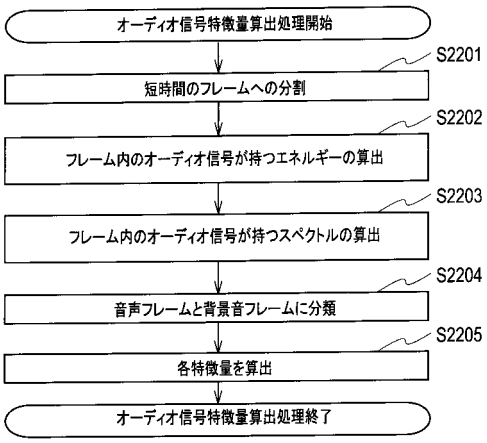
【図２２】



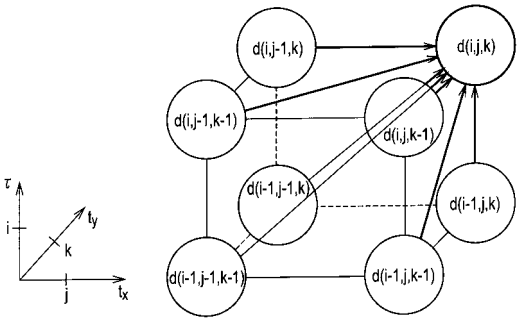
【図２３】



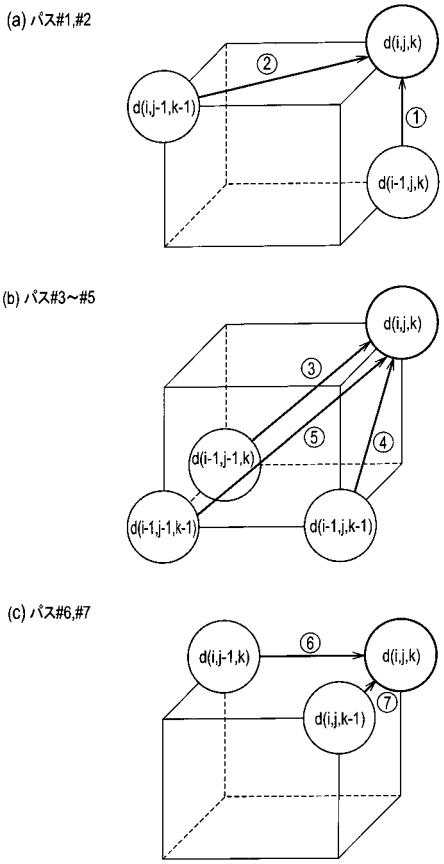
【図２４】



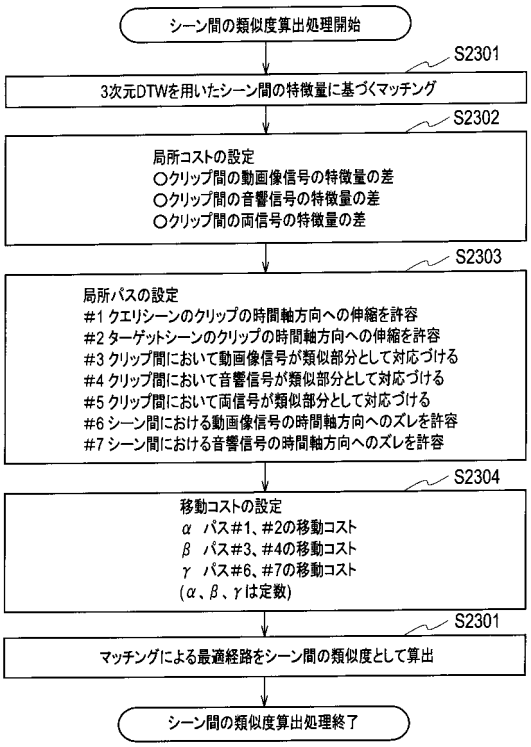
【図２５】



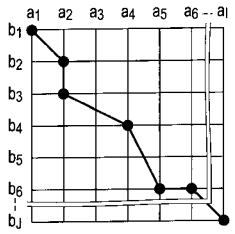
【図２６】



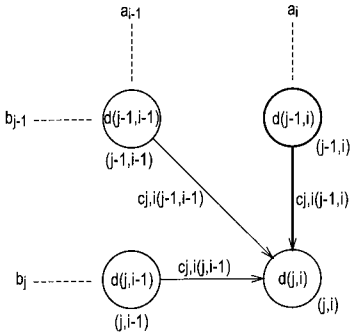
【図２７】



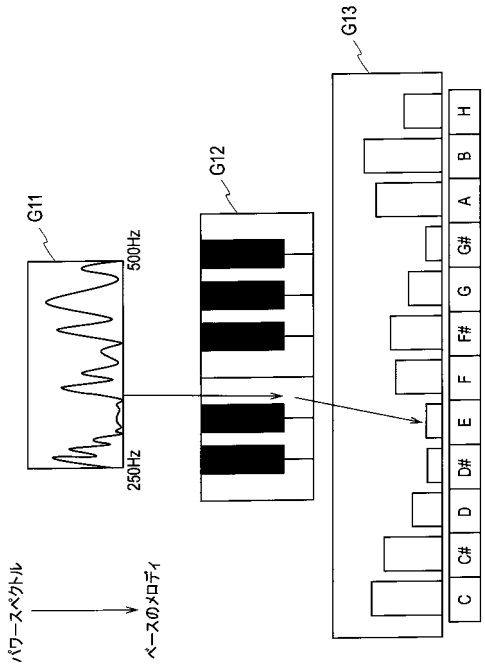
【図２８】



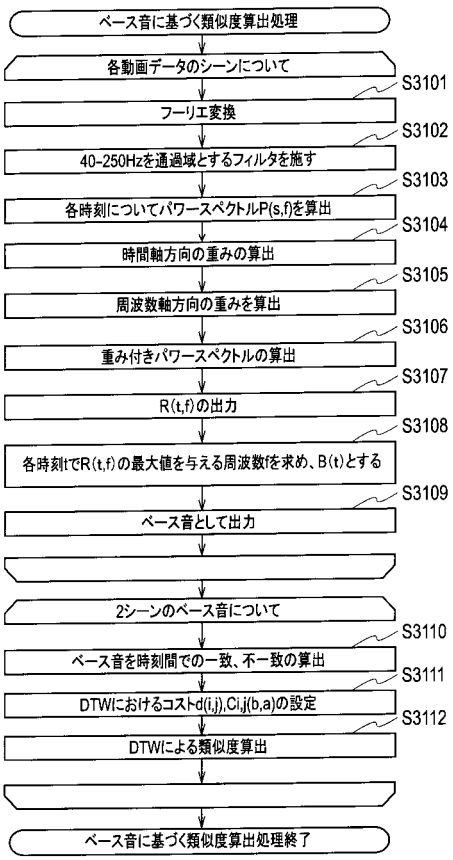
【図２９】



【図３０】



【図３１】

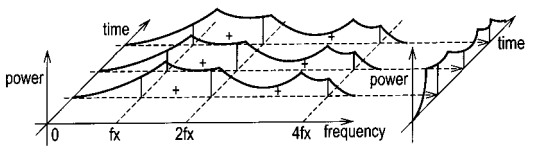


【図３２】

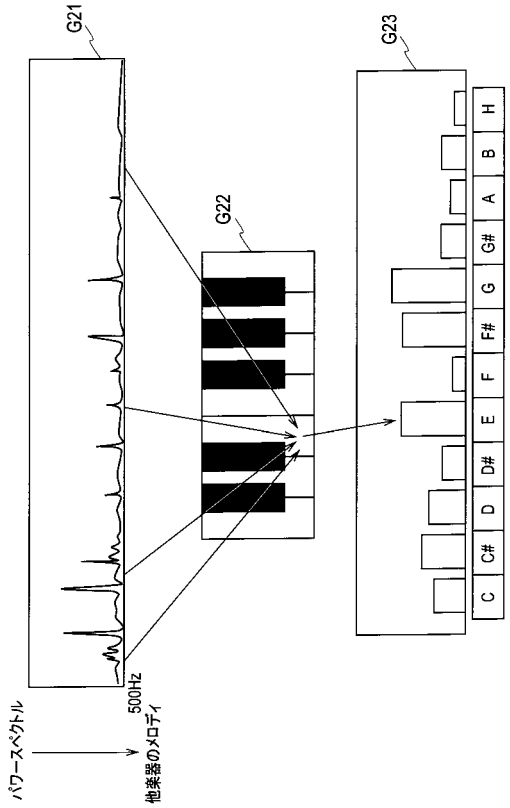
各音名の基本周波数

音名	周波数[Hz]	音名	周波数[Hz]
C	262.8	Fis	371.7
Cis	278.4	G	393.8
D	295.0	Gis	417.2
Dis	312.5	A	442.0
E	331.1	B	468.3
F	350.8	H	496.1

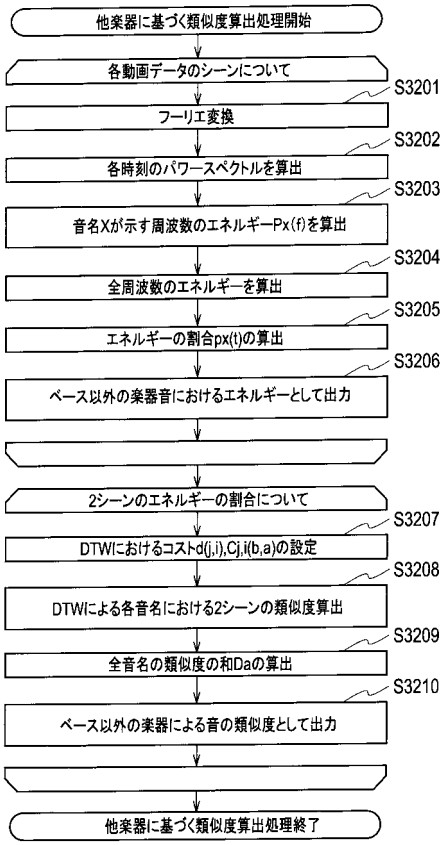
【図３３】



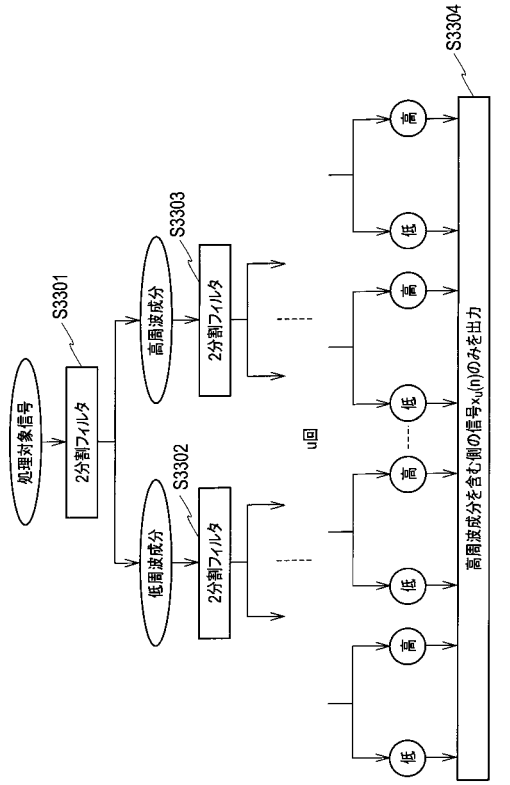
【図３４】



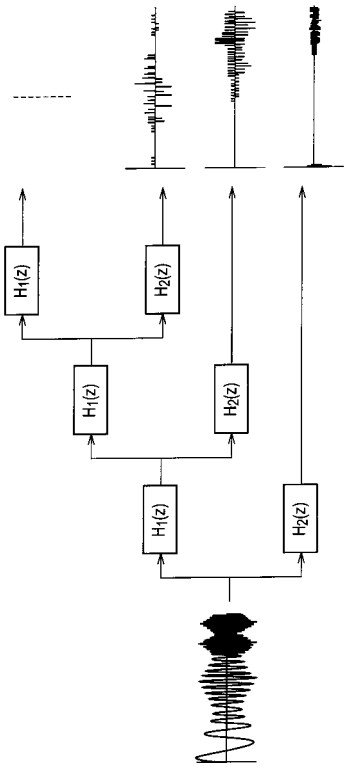
【図３５】



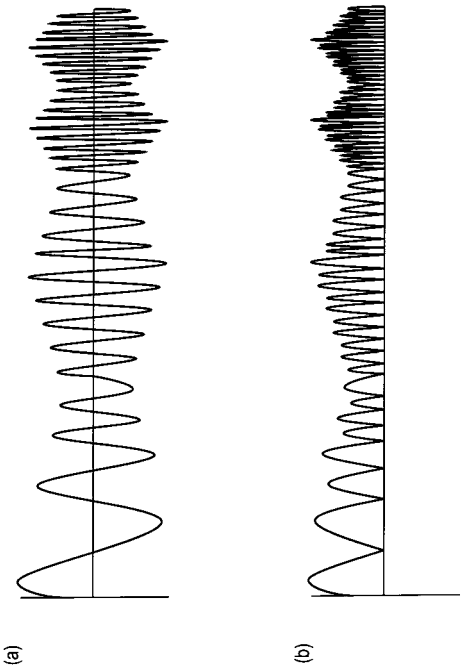
【図３６】



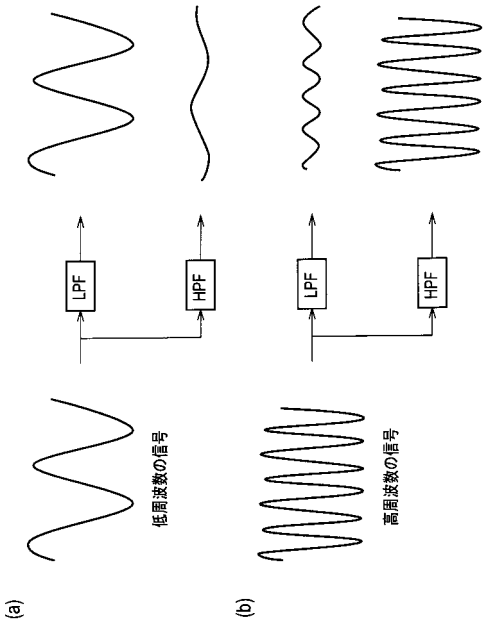
【図３７】



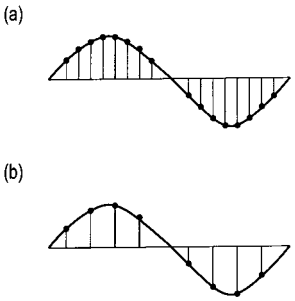
【図３８】



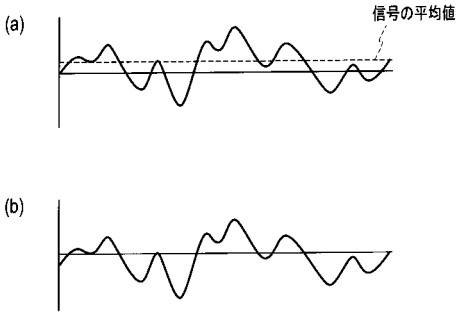
【図３９】



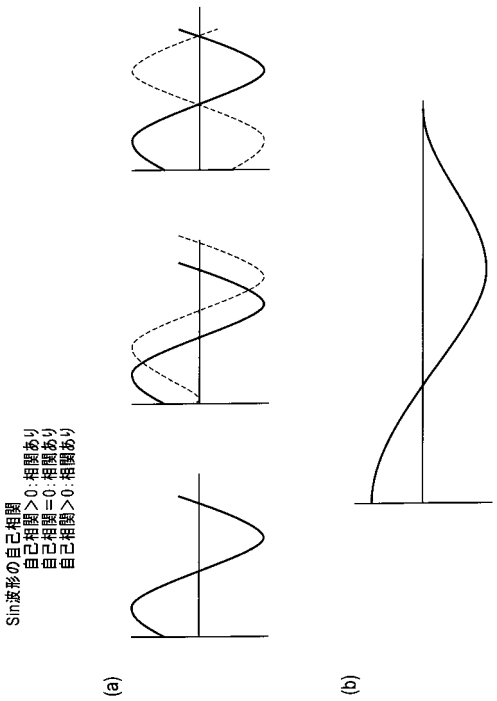
【図４０】



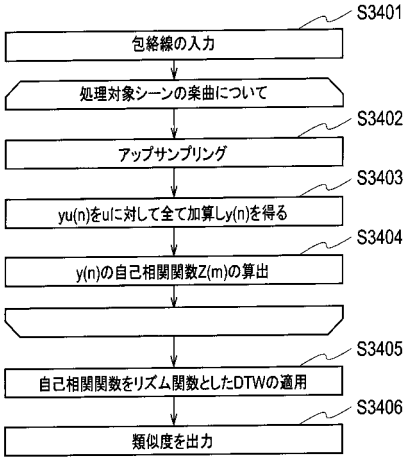
【図４１】



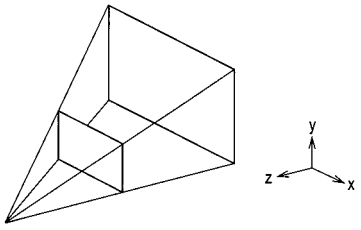
【図４２】



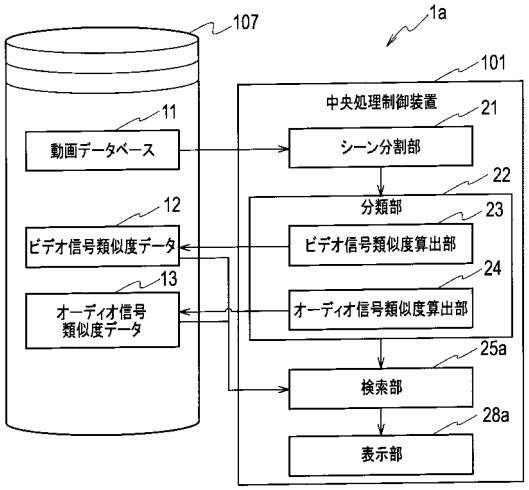
【図４３】



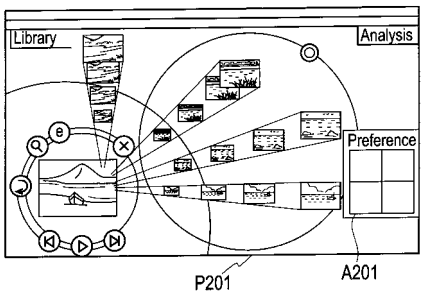
【図４４】



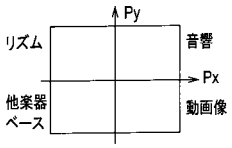
【図４５】



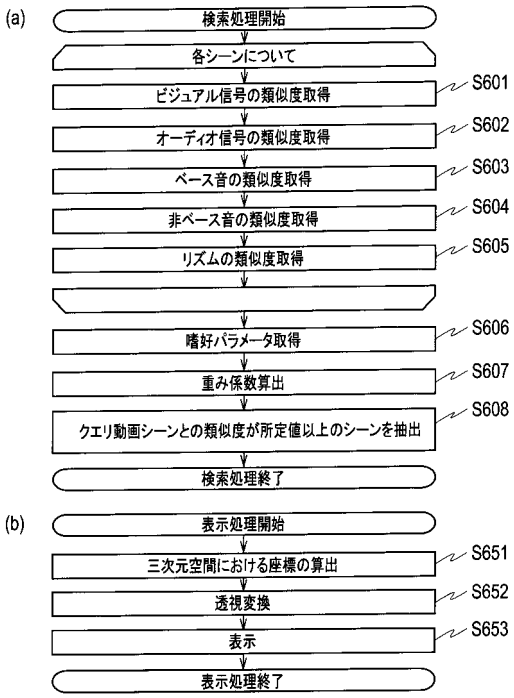
【図４６】



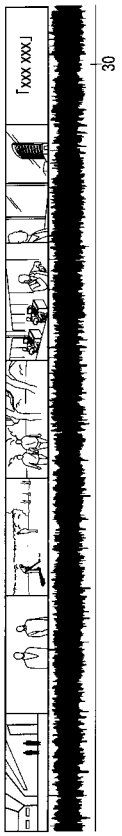
【図４７】



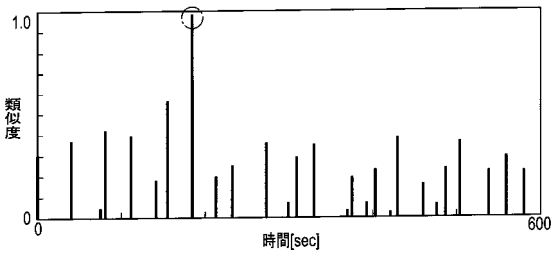
【図４８】



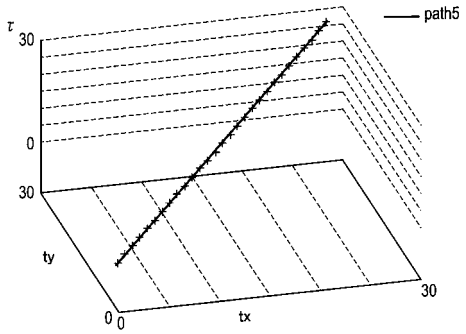
【図４９】



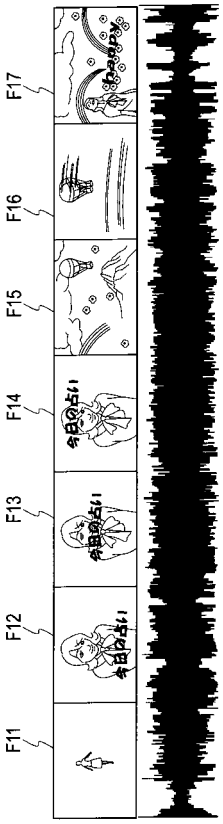
【図５０】



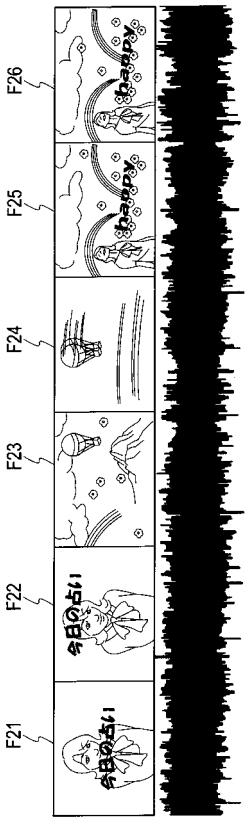
【図５１】



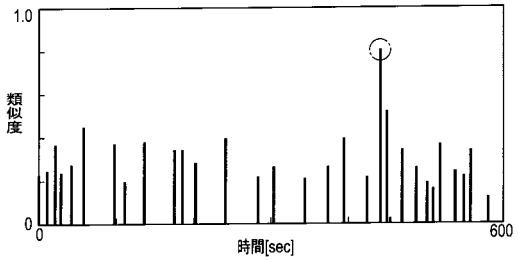
【図５２】



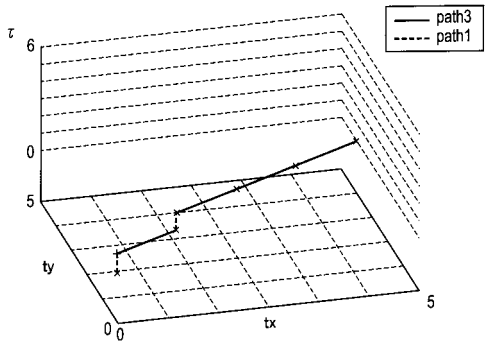
【図５３】



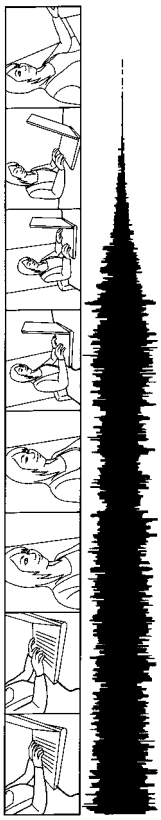
【図５４】



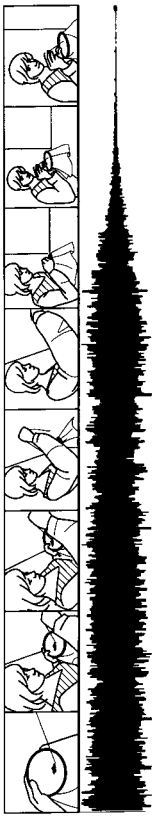
【図５５】



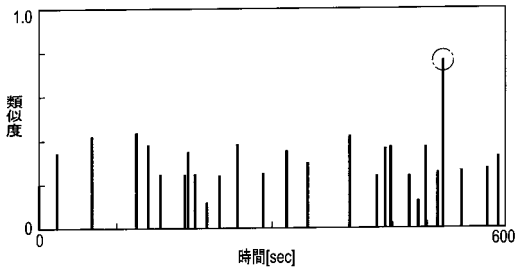
【図５６】



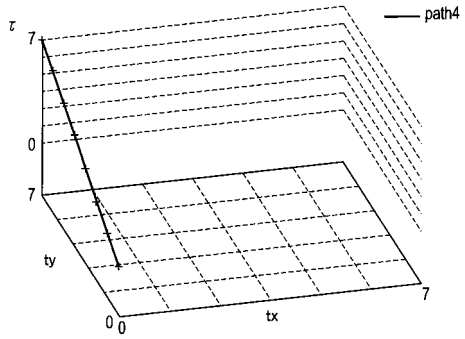
【図５７】



【図５８】



【図５９】



フロントページの続き

(56)参考文献 特開２００６－０１４０８４（ＪＰ，Ａ）
特開２００８－００５１６７（ＪＰ，Ａ）
特開２００５－２５２８５９（ＪＰ，Ａ）

(58)調査した分野(Int.Cl.，ＤＢ名)
Ｈ０４Ｎ　　５／７６
Ｈ０４Ｎ　　５／９３