

(19) 日本国特許庁(JP)

再公表特許(A1)

(11) 国際公開番号

W02009/116582

発行日 平成23年7月21日 (2011. 7. 21)

(43) 国際公開日 平成21年9月24日 (2009. 9. 24)

(51) Int.Cl.	F I	テーマコード (参考)
HO4N 5/76 (2006.01)	HO4N 5/76	B 5C052
HO4N 5/93 (2006.01)	HO4N 5/93	Z 5C053
G11B 27/10 (2006.01)	G11B 27/10	A 5D077

審査請求 未請求 予備審査請求 未請求 (全 69 頁)

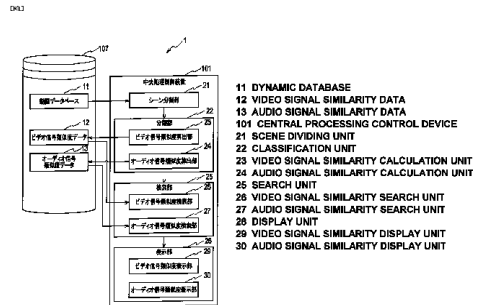
出願番号 特願2010-503907 (P2010-503907)	(71) 出願人 504173471 国立大学法人北海道大学 北海道札幌市北区北8条西5丁目
(21) 国際出願番号 PCT/JP2009/055315	
(22) 国際出願日 平成21年3月18日 (2009. 3. 18)	
(31) 優先権主張番号 特願2008-72537 (P2008-72537)	(74) 代理人 100083806 弁理士 三好 秀和
(32) 優先日 平成20年3月19日 (2008. 3. 19)	(74) 代理人 100101247 弁理士 高橋 俊一
(33) 優先権主張国 日本国 (JP)	(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
	Fターム(参考) 5C052 AC08 DD04 5C053 GB09 HA29 JA01 5D077 AA21 CA20 HC25

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) 【要約】

動画データが記憶された動画データベース11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21と、シーン分割部11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を生成するビデオ信号類似度算出部23と、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検索部26と、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ビデオ信号類似度表示部29を備える。



【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各シーンと、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、を備える動画検索装置。

【請求項 2】

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 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各シーンと、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 3 に記載の動画検索装置。

【請求項 5】

前記シーン分割部は、該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とともに、該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し、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

請求項 1 に記載の動画検索装置。

【請求項 6】

前記ビデオ信号類似度算出部は、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、該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、前記ビジュアル信号の特徴量を算出するとともに、前記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、前記オーディオ信号の特徴量を算出し、前記クリップ単位の前記ビジュアル信号と前記オーディオ信号の特徴量に基づいて、前記シーン間の類似度を算出する

請求項 1 に記載の動画検索装置。

【請求項 7】

請求項 1 に記載の動画検索装置。

10

20

30

40

50

前記オーディオ信号類似度算出部は、前記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し、前記オーディオ信号から前記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、前記エネルギーの差分の合計を算出して、前記ベースを除く楽器に基づく類似度を算出し、2分割フィルタバンクを用いて前記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請求項3に記載の動画検索装置。

【請求項8】

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を生成するオーディオ信号類似度算出部と、

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、前記ベースを除く楽器に基づく類似度と、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

を備える動画検索装置。

【請求項9】

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各シーンと、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機能させる動画検索プログラム。

【請求項10】

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9に記載の動画検索プログラム。

10

20

30

40

50

【請求項 1 1】

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各シーンと、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、更に前記コンピュータを機能させる請求項 9 に記載の動画検索プログラム。

【請求項 1 2】

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 1 1 に記載の動画検索プログラム。

【請求項 1 3】

前記シーン分割手段は、該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とともに、該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し、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

請求項 9 に記載の動画検索プログラム。

【請求項 1 4】

前記ビデオ信号類似度算出手段は、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、該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、前記ビジュアル信号の特徴量を算出するとともに、前記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、前記オーディオ信号の特徴量を算出し、前記クリップ単位の前記ビジュアル信号と前記オーディオ信号の特徴量に基づいて、前記シーン間の類似度を算出する

請求項 9 に記載の動画検索プログラム。

【請求項 1 5】

前記オーディオ信号類似度算出手段は、前記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し、前記オーディオ信号から前記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、前記エネルギーの差分の合計を算出して、前記ベースを除く楽器に基づく類似度を算出し、2 分割フィルタバンクを用いて前記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

請求項 1 1 に記載の動画検索プログラム。

【請求項 1 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、

コンピュータに、

動画データベース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、前記ビジュアル信

10

20

30

40

50

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を生成するオーディオ信号類似度算出手段と、

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、前記ベースを除く楽器に基づく類似度と、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

として機能させる動画検索プログラム。

【請求項 17】

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各シーンと、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 18】

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 17 に記載の動画検索装置。

【請求項 19】

前記オーディオ信号類似度算出部は、前記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し、前記オーディオ信号から前記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、前記エネルギーの差分の合計を算出して、前記ベースを除く楽器に基づく類似度を算出し、2 分割フィルタバンクを用いて前記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請求項 17 に記載の動画検索装置。

【請求項 20】

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

10

20

30

40

50

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各シーンと、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。

10

【請求項 2 1】

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 2 0 に記載の動画検索プログラム。

【請求項 2 2】

前記オーディオ信号類似度算出手段は、前記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し、前記オーディオ信号から前記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、前記エネルギーの差分の合計を算出して、前記ベースを除く楽器に基づく類似度を算出し、2 分割フィルタバンクを用いて前記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

20

請求項 2 0 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。

30

【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）。

40

【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】

50

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. Ogiwara, "Toward Intelligent Music Information Retrieval," IEEE Trans. Multimedia, Vol. 8, No. 3, pp. 564 - 574, 2006.

10

【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。

20

【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、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各シーンと、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と、を備える。

30

【0009】

ここで、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。

【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、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各シーンと、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、を更に備えても良い。このとき、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。

40

【0011】

シーン分割部は、該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とともに、該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

50

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し、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。

【0012】

ビデオ信号類似度算出部は、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、該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とともに、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し、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する。

10

【0013】

オーディオ信号類似度算出部は、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し、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し、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。

20

【0014】

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装置に関する。即ち本発明の第2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、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、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を生成するオーディオ信号類似度算出部と、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備える。

30

40

【0015】

本発明の第3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、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各シーンと、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

50

するビデオ信号類似度検索手段として機能させる。

【0016】

ここで、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。

【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、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各シーンと、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、更にコンピュータを機能させても良い。

10

【0018】

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。

【0019】

シーン分割手段は、該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とともに、該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し、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。

20

【0020】

ビデオ信号類似度算出手段は、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、該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とともに、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し、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する。

30

【0021】

オーディオ信号類似度算出手段は、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し、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し、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。

40

【0022】

本発明の第4の特徴は、動画データのシーンを、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、コンピュータに、動画データベース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、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、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

50

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を生成するオーディオ信号類似度算出手段と、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として機能させる。

10

【0023】

本発明の第5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5の特徴にかかる動画検索装置は、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各シーンと、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。

20

【0024】

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よい。

【0025】

オーディオ信号類似度算出部は、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し、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し、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してもよい。

30

【0026】

本発明の第6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。本発明の第6の特徴にかか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各シーンと、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。

40

【0027】

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よい。

50

【 0 0 2 8 】

オーディオ信号類似度算出手段は、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し、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し、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してもよい。

【 0 0 2 9 】

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 】

【 0 0 3 0 】

【 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である。

【 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 1 0 】 図 1 0 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 1 1 】 図 1 1 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 1 2 】 図 1 2 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 1 3 】 図 1 3 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 1 4 】 図 1 4 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 1 5 】 図 1 5 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 1 6 】 図 1 6 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 1 7 】 図 1 7 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 1 8 】 図 1 8 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。

10

20

30

40

50

【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。

10

【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。

20

【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。

30

【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。

40

【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。

50

【図 4 5】図 4 5 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 4 6】図 4 6 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 4 7】図 4 7 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 4 8】図 4 8 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 4 9】図 4 9 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 5 0】図 5 0 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 5 1】図 5 1 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 D T W のパスを示した図である。

【図 5 2】図 5 2 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 5 3】図 5 3 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 5 4】図 5 4 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 5 5】図 5 5 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 D T W のパスを示した図である。

【図 5 6】図 5 6 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 5 7】図 5 7 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 5 8】図 5 8 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 5 9】図 5 9 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 D T W 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。

10

20

30

40

50

【 0 0 3 3 】

(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 1 1 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。

【 0 0 3 4 】

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。

【 0 0 3 5 】

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 1 1 から複数の映像を読み込み、シーン分割部 2 1 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 2 2 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 2 5 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 2 8 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 2 2 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 2 3 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 2 4 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。

【 0 0 3 6 】

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 P 1 0 1 および表示画面 P 1 0 2 を、表示装置に表示する。表示画面 P 1 0 1 は、クエリ画像表示部 A 1 0 1 を備えている。動画検索装置 1 は、クエリ画像表示部 A 1 0 1 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 1 1 から検索して、表示画面 P 1 0 2 を表示装置に表示する。表示画面 P 1 0 2 には、類似画像表示部 A 1 0 2 a および A 1 0 2 b を備えている。これらの類似画像表示部 A 1 0 2 a および A 1 0 2 b には、動画データベース 1 1 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 A 1 0 1 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。

【 0 0 3 7 】

(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 1 0 1、ROM (Read Only Memory) 1 0 2、RAM (Random Access Memory) 1 0 3 及び入出力インタフェース 1 0 9 が、バス 1 1 0 を介して接続されている。入出力インタフェース 1 0 9 には、入力装置 1 0 4、表示装置 1 0 5、通信制御装置 1 0 6、記憶装置 1 0 7 及びリムーバブルディスク 1 0 8 が接続されている。

【 0 0 3 8 】

中央処理制御装置 1 0 1 は、入力装置 1 0 4 からの入力信号に基づいて ROM 1 0 2 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

10

20

30

40

50

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。

10

【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。

20

【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。

30

【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。

40

【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

50

S 1 0 1 ないしステップ S 1 1 0 の処理を繰り返す。

ステップ S 1 0 1 において、動画データベース 1 1 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 S 1 0 2 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 S 1 0 2 で分割された各クリップについて、ステップ S 1 0 3 ないしステップ S 1 0 5 の処理を繰り返す。

【 0 0 4 5 】

ステップ S 1 0 3 において、クリップの特徴量が算出され、ステップ S 1 0 4 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 S 1 0 4 において削減された後の特徴量に基づいて、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。

10

【 0 0 4 6 】

ステップ S 1 0 3 ないしステップ S 1 0 5 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 S 1 0 6 において、ステップ S 1 0 1 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 S 1 0 7 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 S 1 0 7 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 S 1 0 8 およびステップ S 1 0 9 の処理を繰り返す。

【 0 0 4 7 】

20

ステップ S 1 0 8 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 S 1 0 5 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 S 1 0 9 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。

【 0 0 4 8 】

ステップ S 1 0 7 で分割された全てのショットについて、ステップ S 1 0 8 およびステップ S 1 0 9 の処理が実行されると、ステップ S 1 1 0 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。

30

【 0 0 4 9 】

分類部 2 2 は、ビデオ信号類似度算出部 2 3 とオーディオ信号類似度算出部 2 4 を備えている。

【 0 0 5 0 】

ビデオ信号類似度算出部 2 3 は、シーン分割部 2 1 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 1 2 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 1 1 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 2 3 は、シーン分割部 2 1 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 2 3 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 2 3 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 1 2 として、

40

50

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。

10

【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。

20

30

【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。

40

【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

50

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S331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】

10

20

30

40

50

図 1 1 (b) を参照して、ビデオ信号類似度表示部 2 9 の処理を説明する。ステップ S 4 5 1 において、ビデオ信号類似度検索部 2 6 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 D T W によって得られる 3 つの座標になる。ステップ S 4 5 2 において、ステップ S 4 5 1 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 S 4 5 3 において、表示装置に表示される。

【 0 0 6 4 】

オーディオ信号類似度検索部 2 7 は、オーディオ信号類似度データ 1 3 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 3 0 は、オーディオ信号類似度検索部 2 7 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。

10

【 0 0 6 5 】

図 1 2 を参照して、オーディオ信号類似度検索部 2 7 およびオーディオ信号類似度表示部 3 0 の処理を説明する。

図 1 2 (a) を参照して、オーディオ信号類似度検索部 2 7 の処理を説明する。まず、記憶装置 1 0 7 からオーディオ信号類似度データ 1 3 が読み出される。さらに、シーン分割部 2 1 によって分割された各シーンについて、ステップ S 5 0 1 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 S 5 0 2 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 S 5 0 1 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。

20

【 0 0 6 6 】

つぎにステップ S 5 0 4 において、ステップ S 5 0 1 ないしステップ S 5 0 3 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。

【 0 0 6 7 】

図 1 2 (b) を参照して、オーディオ信号類似度表示部 3 0 の処理を説明する。ステップ S 5 5 1 において、オーディオ信号類似度検索部 2 7 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 S 5 5 2 において、ステップ S 5 5 1 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 S 5 5 3 において、表示装置に表示される。

30

以下、図 1 に示す各ブロックについて詳述する。

【 0 0 6 8 】

(シーン分割部)

次に、図 1 に示すシーン分割部 2 1 の処理を説明する。

シーン分割部 2 1 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 1 1 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。

40

【 0 0 6 9 】

シーン分割部 2 1 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 P C A (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 M G D 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、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

50

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。

10

【0071】

ここで、本発明の最良の実施の形態に係るシーン分割2-1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(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。

20

【0072】

本発明の最良の実施の形態に係るシーン分割部2-1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)

30

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。

40

シーン分割部2-1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2-1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

50

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。

10

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。

20

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、 $S_i + N_o$ と、 $S_p + M_u$ 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(S_i)のみで構成される信号と、雑音(N_o)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(S_p)のみで構成される信号と、音楽(M_u)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(S_i)のみで構成される信号と、雑音(N_o)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(S_p)のみで構成される信号と、音楽(M_u)のみで構成される信号である。

30

【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。

40

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

50

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 1 2 5 2 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 S 1 2 5 3 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 S 1 2 5 4 において、PCA（主成分分析）を施す。さらにステップ S 1 2 5 5 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 1 6 (b) の処理が実行される。

【 0 0 7 8 】

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 S 1 0 5 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。

【 0 0 7 9 】

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 $f^{(c)}$ ($c = 1, \dots, 3$; 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 】

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

を、次式により算出する。

【 数 2 】

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

ただし、

【 数 3 】

$m_i^{(c)}$ 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 】

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

10

20

30

40

【数 5】

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

を用いて、処理対象信号が参照信号 1、2 と同一のクラスに属する帰属度

【数 6】

$$D_i^{(c)}$$

10

を次式で定義する。

【数 7】

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

(式 1 - 2)

20

C L S # 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】

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

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】

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

30

を、以下で定義する。

【数 10】

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

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

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

40

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

上式は、C L S # 1 から C L S # 3 の各分類処理において、

【数 11】

$$D_i^{(c)}$$

を、参照信号 1、2 と同一のクラスに分類される確率とみなし、それらを積算すること

50

で、 S_i 、 S_p 、 M_u 、 N_o 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 1 2】

$$P_{l_i} (l_i = 1, \dots, 4)$$

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ら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 S 1301 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 S 1302 ないしステップ S 1305 の処理を行う。具体的には、ステップ S 1302 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 S 1303 において、処理対象信号が参照信号 1 のクラスに属する帰属度を算出する。さらに、ステップ S 1304 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 S 1305 において、処理対象信号が参照信号 2 のクラス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 S 1306 ないしステップ S 1309 の処理を行う。具体的には、ステップ S 1306 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 S 1307 において、処理対象信号が参照信号 1 のクラスに属する帰属度を算出する。さらに、ステップ S 1308 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 S 1309 において、処理対象信号が参照信号 2 のクラスに属する帰属度を算出する。

ここで、ステップ S 1310 において、ステップ S 1303 およびステップ S 1307 で算出された帰属度に基づいて、オーディオクラス S_i への帰属確率 P_1 が算出される。同様に、ステップ S 1311 において、ステップ S 1303 およびステップ S 1309 で算出された帰属度に基づいて、オーディオクラス N_o 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 S 1312 ないしステップ S 1315 の処理を行う。具体的には、ステップ S 1312 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 S 1313 において、処理対象信号が参照信号 1 のクラスに属する帰属度を算出する。さらに、ステップ S 1314 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 S 1315 において、処理対象信号が参照信号 2 のクラスに属する帰属度を算出する。

ここで、ステップ S 1316 において、ステップ S 1305 およびステップ S 1313 で算出された帰属度に基づいて、オーディオクラス S_p への帰属確率 P_2 が算出される。同様に、ステップ S 1317 において、ステップ S 1305 およびステップ S 1315 で算出された帰属度に基づいて、オーディオクラス M_u への帰属確率 P_3 が算出される。

【0083】

次に、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 S 107 に相当する。

本発明の最良の実施の形態においては、分割²検定法を用いて、ショットカットを得る。分割²検定法は、まず動画像のフレームを $4 \times 4 = 16$ 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。

10

20

30

40

50

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 ($r = 1, \dots, 16$) を算出する。

【数 1 3】

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

さらに、算出された 16 個の値 E_r ($r = 1, \dots, 16$) の中で値の小さい 8 の総和 E_{sum} を算出し、 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。

10

【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、 $4 \times 4 = 16$ 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。

20

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

30

【数 1 4】

$$x_{l_1} \quad (l_1 = 1, \dots, 4; \text{それぞれ } S_i, S_p, \mu, N_o \text{ に対応})$$

を次式で算出する。

【数 1 5】

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

40

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 1 6】

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

は k 番目のクリップにおける帰属確率

【数 17】

$$R_{l_1}$$

を表す。これら 4 つの平均値

【数 18】

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

10

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、

20

【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。

30

【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)

40

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。

50

【数 2 0】

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

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

【数 2 1】

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

10

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 2 2】

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

を算出する。

【0 0 9 0】

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 S 1 0 9 に相当する。

20

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 2 3】

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

に対し、図 2 0 および図 2 1 に示すファジィ制御規則

【数 2 4】

30

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

を適用し、各ショットクラスの出力変数

【数 2 5】

$$y_{l_2}$$

40

およびメンバシップ関数の値

【数 2 6】

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

を算出する。

50

【 0 0 9 1 】

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 S 1 1 0 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 2 7】

$$\mu_{l_2}$$

10

を用いて、映像信号のシーン分割を行う。

ここで、 l_1 をショット番号とし、隣接するショット間の距離 $D(l_1, l_2)$ を次式で定義する。

【数 2 8】

$$D(\eta_1, \eta_2) = \sum_{l_2=1}^6 |\mu_{l_2}(\eta_1) - \mu_{l_2}(\eta_2)| \quad (\text{式 1 - 1 0})$$

20

【 0 0 9 2 】

この距離 $D(l_1, l_2)$ 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 $D(l_1, l_2)$ 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。

【 0 0 9 3 】

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 2 2 参照して説明する。

30

まずステップ S 1 5 0 1 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 S 1 5 0 2 において、1 1 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 S 1 5 0 1 およびステップ S 1 5 0 2 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 S 1 5 0 3 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 2 0 および図 2 1 に示すファジィ制御規則が参照される。ステップ S 1 5 0 3 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 S 1 5 0 4 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 S 1 5 0 5 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 S 1 5 0 4 およびステップ S 1 5 0 5 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。

40

【 0 0 9 4 】

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。

【 0 0 9 5 】

50

(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。

10

【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に対応付けることが可能となる。

20

【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1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に対応付けし、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。

30

【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。

40

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】

50

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】

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

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

ここで、 n 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、 $\omega(m)$ 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】

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

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 - 4000 Hz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。

10

20

30

40

【数 3 1】

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) 10

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 3 2】

b) 低 STE 率 $LSTER$:

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

(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 3 3】

c) 平均零交差率 \overline{ZCR} :

零交差率 $ZCR(n)$ は次式で定義される。

$$ZCR(n) = \frac{1}{2} \sum_m |\text{sgn}[x(m)] - \text{sgn}[x(m-1)]| \omega(m)$$

(式 2 - 5)

ここで、 $x(m) \geq 0$ のとき $\text{sgn}[x(m)] = 1$ 、その他では $\text{sgn}[x(m)] = -1$ 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 3 4】

d) スペクトルフラックス密度 SF :

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

ここで、 $STS(n,k)$, ($k = 1, \dots, K$) は時刻 n における k 番目のスペクトルである。 (式 2 - 6)

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, $\overline{ERSB1/2/3/4}$:

$\overline{ERSB1/2/3/4}$ 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】

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

(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を用いて類似部分に対応づけ、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。

10

20

30

40

50

【 0 1 0 7 】

(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 (1 T₁)、ターゲットシーンのビジュアル信号のクリップ t_x (1 t_x T₂)、ターゲットシーンのオーディオ信号のクリップ t_y (1 t_y T₂) 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 (, t_x , t_y) を以下の3種類で定義する。

【数 3 7】

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

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。

【 0 1 0 8 】

(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 2 5 および図 2 6 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。

【 0 1 0 9 】

(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 (, 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 3 8】

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

【数 3 9】

$$S(\tau, t_x, t_y) = \min \left\{ \begin{array}{l} S(\tau - 1, t_x, t_y) + d_{av}(\tau, t_x, t_y) + \alpha \\ S(\tau, t_x - 1, t_y - 1) + d_{av}(\tau, t_x, t_y) + \alpha \\ S(\tau - 1, t_x - 1, t_y) + d_v(\tau, t_x, t_y) + \beta \\ S(\tau - 1, t_x, t_y - 1) + d_a(\tau, t_x, t_y) + \beta \\ S(\tau - 1, t_x - 1, t_y - 1) + d_{av}(\tau, t_x, t_y) \\ S(\tau, t_x - 1, t_y) + d_v(\tau, t_x, t_y) + \gamma \\ S(\tau, t_x, t_y - 1) + d_a(\tau, t_x, t_y) + \gamma \end{array} \right. \quad (式 2 - 1 0)$$

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 4 0】

$$D_S = \min \left(\frac{S(T_1, T_2, t_y)}{T_1 + 2T_2}, \frac{S(T_1, t_x, T_2)}{T_1 + 2T_2} \right) \quad (式 2 - 1 1)$$

【0 1 1 0】

この処理を図 2 7 を参照して説明する。

まず、ステップ S 2 3 0 1 において、3次元 D T W 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 - 1 0) 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 S 2 3 0 2 において、3次元 D T W に必要な局所コストが設定され、ステップ S 2 3 0 3 において、局所パスが設定される。さらにステップ S 2 3 0 4 において、 α 、 β 、 γ の各移動コストする。 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 S 2 3 0 5 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。

【0 1 1 1】

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 D T W を用いてシーン間の類似度を算出する。ここで 3次元 D T W を用いることにより、後述する表示部で、3次元座標に基づいてシーンの類似度を可視化することができる。

【0 1 1 2】

(D T W の概要)

ここで、D T W 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 D T W の構成について説明を行う。D T W 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、類似度より求められる類似度の算出に D T W 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。

【0 1 1 3】

まず、D T W 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, \dots, a_I とし、長さ J の被参照パター

10

20

30

40

50

ンに含まれる各要素を順に b_1, b_2, \dots, b_J と表現する。さらに、各パターンの位置集合を $\{1, 2, \dots, I\}, \{1, 2, \dots, J\}$ 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 $w: \{1, 2, \dots, I\} \rightarrow \{1, 2, \dots, J\}$ 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 4 1】

$$\begin{aligned} w(1) &= 1 \\ w(I) &= J \end{aligned} \quad (\text{式 2-12})$$

10

b) w は単調写像である。

【数 4 2】

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

【0 1 1 4】

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 2 8 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。

20

【0 1 1 5】

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 2 9 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。

30

【0 1 1 6】

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。

【0 1 1 7】

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。

40

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。

【0 1 1 8】

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。

50

【数 4 3】

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

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 4 4】

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

10

【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。

20

- ・ 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。

30

【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。

40

【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。

50

【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。

10

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。

20

【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。

30

【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。

40

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。

50

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、ベース音を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。

10

【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。

20

【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 - 250 Hz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、 $e^{-((x - \mu) / \sigma)^2}$ 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 ($0 \leq t \leq T$)、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。

30

40

【数45】

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

ここで、

【数 4 6】

• 時間軸方向の重み： $v_t(s)$

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

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 4 7】

10

• 周波数軸方向の重み： $w(f)$

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

(式 3 - 3)

20

ただし、 m を自然数として、

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

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

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。

30

(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。

【0 1 3 5】

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 S 3 1 3 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。

【0 1 3 6】

40

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 S 3 1 4 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 4 8】

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

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

10

ただし、 $\alpha > \beta$ 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 3 1 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 1 1 の各シーンについて、ステップ S 3 1 0 1 ないしステップ S 3 1 0 9 の処理が実行される。

ステップ S 3 1 0 1 において、1 つのシーンにフーリエ変換をする。ステップ S 3 1 0 2 において、4 0 - 2 5 0 Hz を通過域とするフィルタを施す。ステップ S 3 1 0 3 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。

20

【0138】

一方、ステップ S 3 1 0 4 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 S 3 1 0 5 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 S 3 1 0 6 において、ステップ S 3 1 0 4 およびステップ S 3 1 0 5 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 S 3 1 0 7 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 S 3 1 0 9 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 S 3 1 0 1 ないしステップ S 3 1 0 9 の処理が終了すると、ステップ S 3 1 1 0 ないしステップ S 3 1 1 2 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。

30

まずステップ S 3 1 1 0 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 S 3 1 1 1 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 S 3 1 1 2 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)

40

オーディオ信号類似度算出部 2 4 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 S 3 0 2 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 3 2 の周波数を持ち、各周波数の 2^k ($k = 1, 2, \dots$) 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 3 2 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

50

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 f_x と表す場合、図 3 3 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、 $2 f_x$ 、 $4 f_x$ ・・・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 ($0 \leq t \leq T$)、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 3 4 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 $P_x(t)$ を G 2 1 に示す。G 2 2 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、G 2 3 に示すように、音階ごとに、ヒストグラムを算出する。G 2 3 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 $P_x(t)$ を示している。

図 3 4 に示す処理において、 C から H までの各 1 2 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 3 4 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 S 3 2 1 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 3 2 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 4 9】

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

ただし、 K は

【数 5 0】

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

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 S 3 2 2 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 $p_x(t)$ は次式で示される。

【数 5 1】

$$p_x(t) = \frac{P_x(t)}{\int_0^{\frac{1}{2}} P(t, f) df} \quad (\text{式 4-2})$$

以上を全ての t、X について施し、得られた $p_x(t)$ 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。

10

【0 1 4 5】

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 S 3 2 3 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 $p_{x_r}(t)$ 、 $p_{x_i}(t)$ 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 D_{a_x} とすると、ベース以外の楽器による音の類似度 D_a は次式で表される。

20

【数 5 2】

$$D_a = D_{a_C} + D_{a_{Cis}} + D_{a_D} + D_{a_{Dis}} + D_{a_E} + D_{a_F} + D_{a_{Fis}} + D_{a_G} + D_{a_{Gis}} + D_{a_A} + D_{a_B} + D_{a_H} \quad (\text{式 4-3})$$

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 5 3】

$$d(j, i) = |p_{X_i}(j) - p_{X_r}(i)| \quad (\text{式 4-4})$$

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

30

(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。

【0 1 4 6】

ここで、図 3 5 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。

40

まず、動画データベース 1 1 の各シーンについて、ステップ S 3 2 0 1 ないしステップ S 3 2 0 6 の処理が実行される。

ステップ S 3 2 0 1 において、1 つのシーンにフーリエ変換をする。ステップ S 3 2 0 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 S 3 2 0 3 において、音名 X が示す周波数エネルギー $P_x(t)$ を算出して、 $p_x(t)$ を算出する。

一方、ステップ S 3 2 0 4 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 S 3 2 0 5 において、ステップ S 3 2 0 3 で算出された音名が示す周波数のエネルギー $P_x(t)$ と、ステップ S 3 2 0 4 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 $p_x(t)$ を算出する。ステップ S 3 2 0 6 において、このエネルギーの割

50

合 $p \times (t)$ 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 S 3 2 0 1 ないしステップ S 3 2 0 6 の処理が終了すると、ステップ S 3 2 0 7 ないしステップ S 3 2 1 0 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 S 3 2 0 7 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 S 3 2 0 8 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 S 3 2 0 9 において、ステップ S 3 2 0 8 において算出された全音名の類似度の和 D_a を算出する。ステップ S 3 2 1 0 において、この和 D_a 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。

10

【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 2 4 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 S 3 0 3 および図 1 0 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。

20

具体的なアルゴリズムを以下に示す。尚、各処理は図 1 0 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 1 0 のステップ S 3 3 1 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 $x_u(n)$ ($u = 1, \dots, U; n = 1, \dots, N_u$) 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。

30

図 3 6 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 S 3 3 0 1 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 S 3 3 0 1 で分割された低周波成分を、ステップ S 3 3 0 2 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 S 3 3 0 1 で分割された高周波成分を、ステップ S 3 3 0 3 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 S 3 3 0 4 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 3 7 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。

40

【0150】

次に、包絡線の検波処理を説明する。この処理は、図 1 0 のステップ S 3 3 2 ないしステップ S 3 3 5 に相当する。以下の 1) ないし 4) は、それぞれ図 1 0 のステップ S 3 3 2 ないしステップ S 3 3 5 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。

50

【 0 1 5 1 】

1) 全波整流

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

【数 5 4】

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

全波整流を施すことにより、図 3 8 (a) に示す波形から、図 3 8 (b) に示す波形を得ることができる。

10

【 0 1 5 2 】

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 $y_{2u}(n)$ ($u = 1, \dots, U; n = 1, \dots, N_u$) を得る。

【数 5 5】

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

20

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 3 9 (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 3 9 (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。

【 0 1 5 3 】

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

30

【数 5 6】

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

を得る。

【数 5 7】

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

40

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 4 0 (a) に示す信号から間引きされ、図 4 0 (b) に示す信号が出力される。

【 0 1 5 4 】

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 $y_u(n)$ ($u = 1, \dots, U; n = 1, \dots, N_u$) を得る。

【数 5 8】

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

ただし、 $E[y_{3_u}(n)]$ は信号 $y_{3_u}(n)$ の平均値を示す。

平均値除去処理をすることにより、図 4 1 (a) に示す信号から、図 4 1 (b) に示す信号が出力される。

【0 1 5 5】

次に、自己相関関数の算出処理を説明する。この処理は、図 1 0 のステップ S 3 3 6 に相当する。 10

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 $y(n)$ ($n = 1, \dots, N_1$) 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 $z(m)$ ($m = 0, \dots, N_1 - 1$) を次式により算出する。

【数 5 9】

$$z(m) = \frac{1}{N_1} \sum_n^{N_1} y(n)y(n-m) \quad (\text{式 5 - 5}) \quad 20$$

自己相関について、図 4 2 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。 30

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。

【0 1 5 6】

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 1 0 のステップ S 3 3 7 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。 40

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。

【0 1 5 7】

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。

10

20

30

40

50

【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。

10

【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D T W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。

20

【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、z1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、ビデオ信号類似度検索部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検索部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。

30

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。

40

【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。さらに表示部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

50

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。

10

【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。

20

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データ13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。

30

【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。

40

【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

50

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】

$$D = W_{sv} D_{sv} + W_{sa} D_{sa} + W_b D_b + W_a D_a + W_r D_r$$

$$W_{sv} = P_x P_y$$

$$W_{sa} = P_x (1 - P_y)$$

$$W_b = (1 - P_x)(1 - P_y)$$

$$W_a = \frac{(1 - P_x) P_y}{2}$$

$$W_r = \frac{(1 - P_x) P_y}{2}$$

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_r 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_r 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1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53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53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53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53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52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54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54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54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54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54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

10

20

30

40

50

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。

10

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 54 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 54 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 54 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。

20

【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。

30

【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。

40

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

50

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。

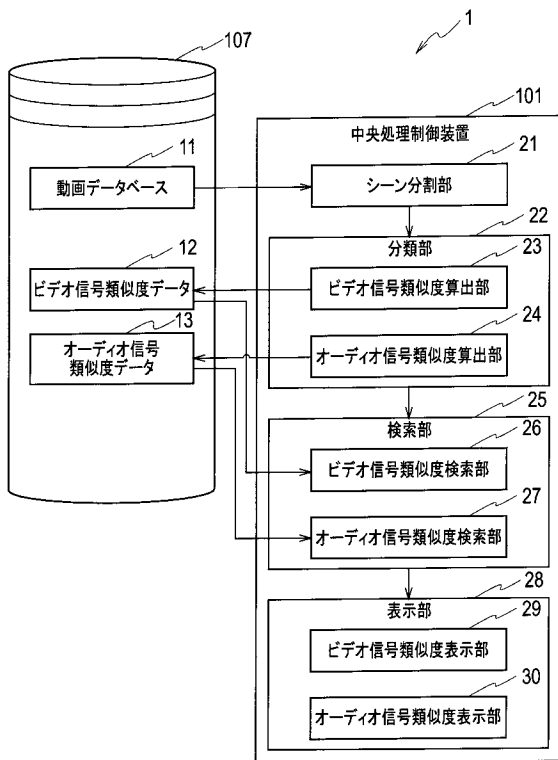
一方、ビデオ信号のみに基づいて類似度を算出、検索および表示する実施態様も考えられる。具体的には、分類部 22 はビデオ信号類似度算出部 23 を備え、検索部 25 はビデオ信号類似度検索部 26 を備え、表示部 28 はビデオ信号類似度表示部 29 を備える。

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表示部 30 を備える。

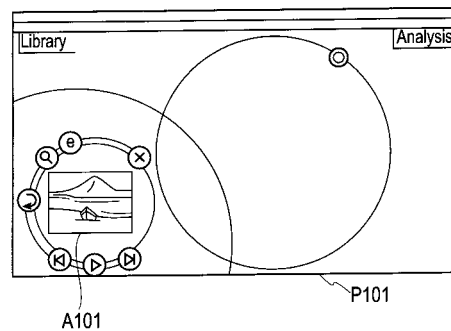
【0182】

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。

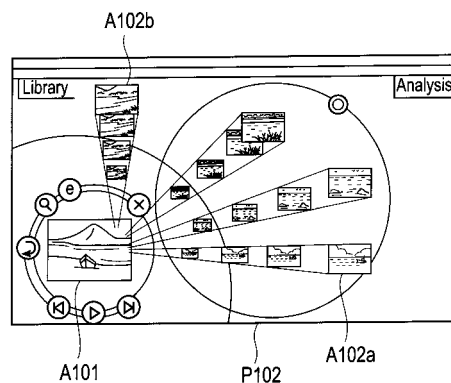
【図 1】



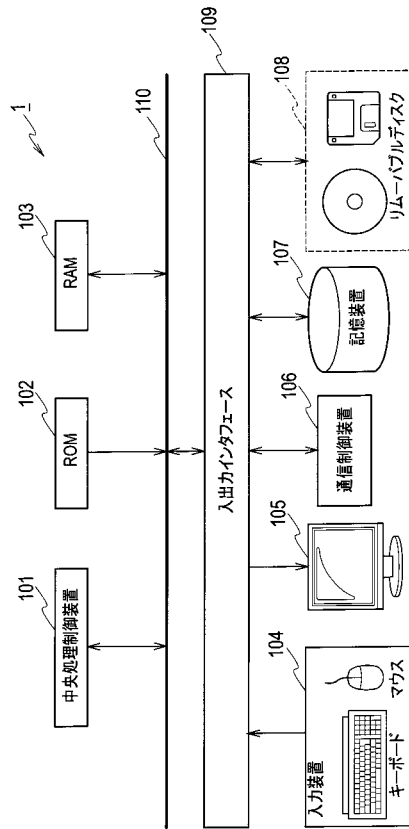
【図 2】



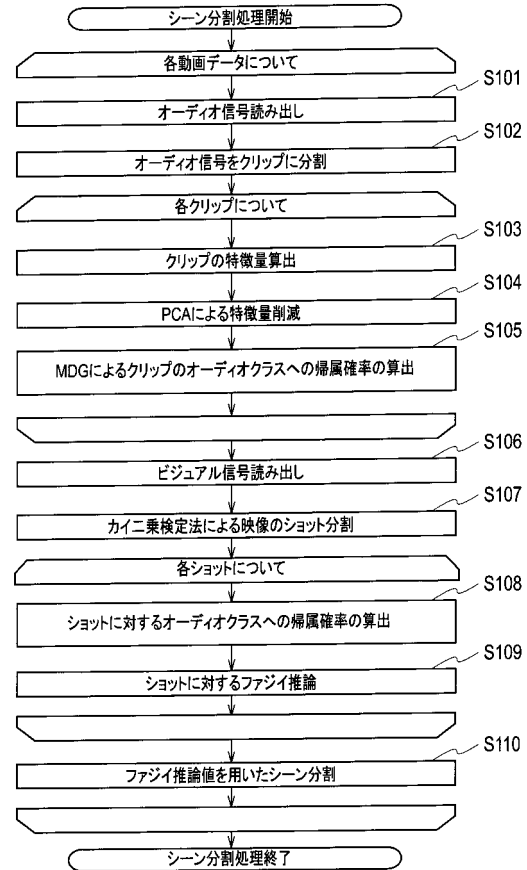
【図 3】



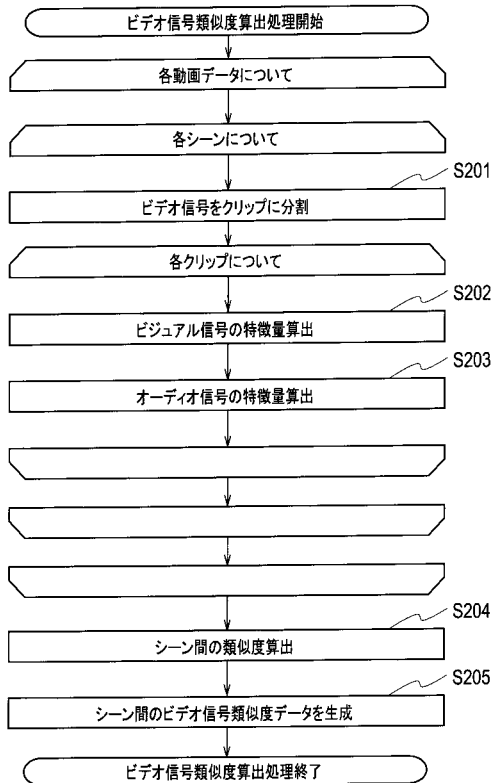
【 図 4 】



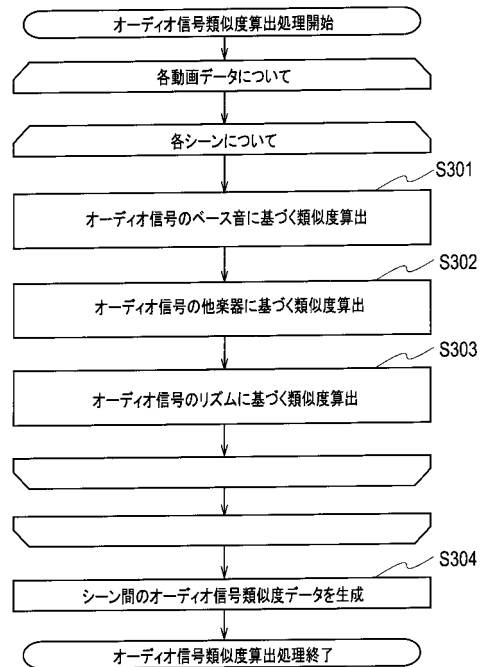
【 図 5 】



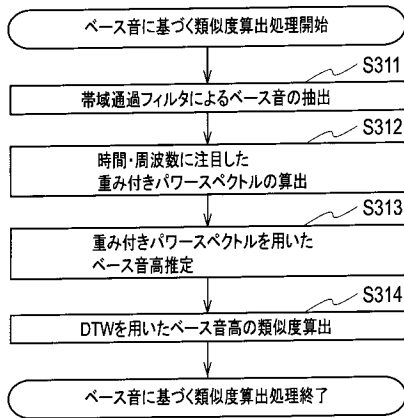
【 図 6 】



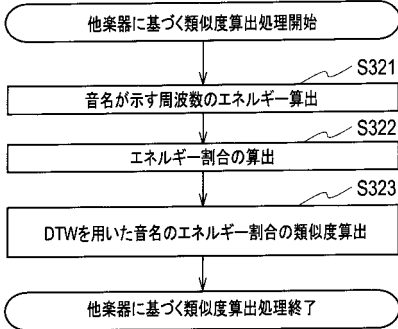
【 図 7 】



【 図 8 】

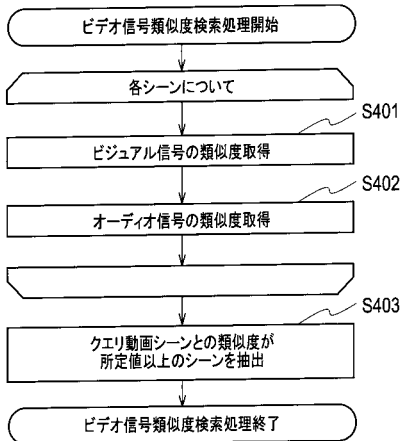


【 図 9 】

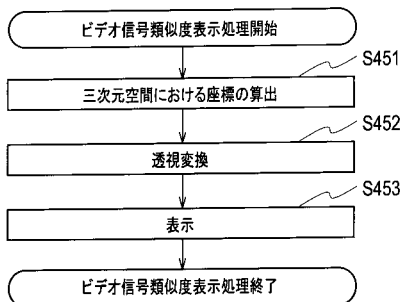


【 図 1 1 】

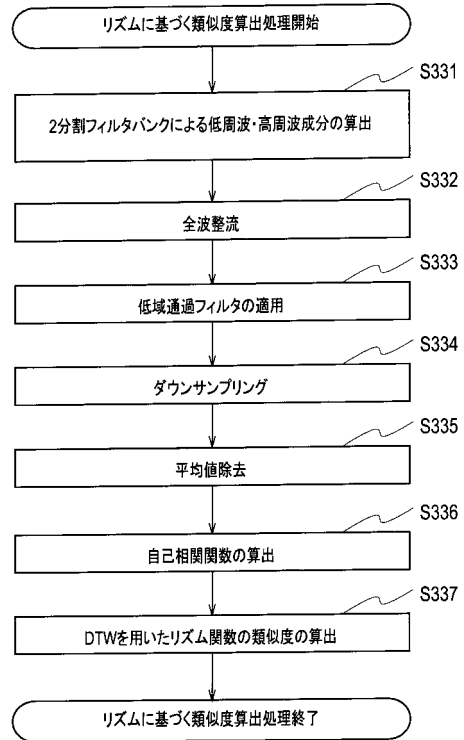
(a)



(b)

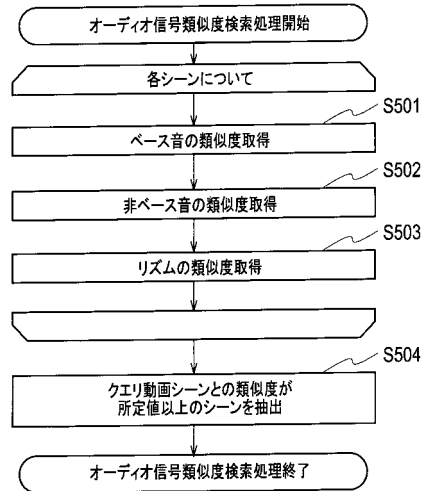


【 図 1 0 】

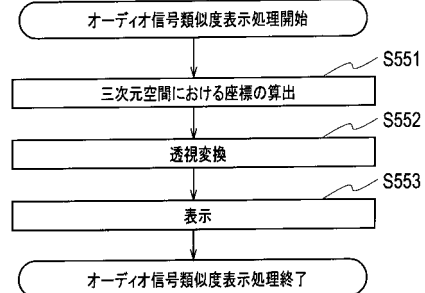


【 図 1 2 】

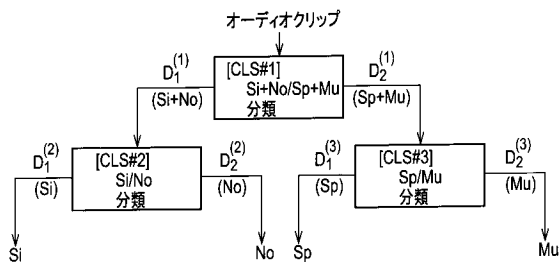
(a)



(b)



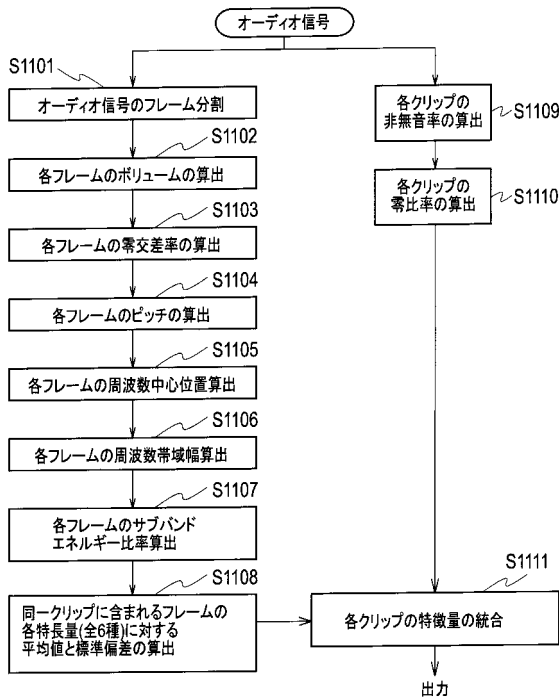
【 図 1 3 】



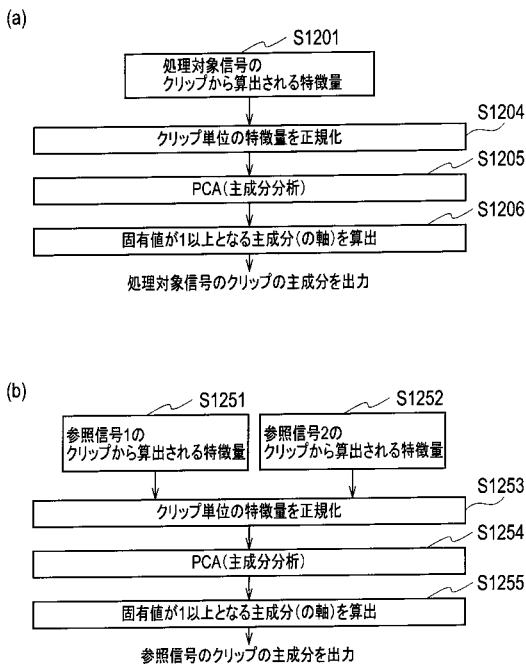
【 図 1 4 】

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

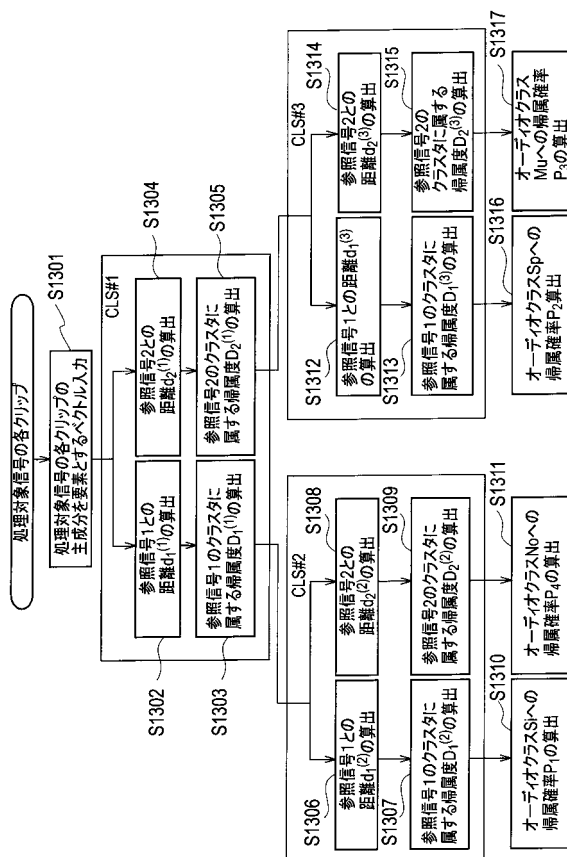
【 図 1 5 】



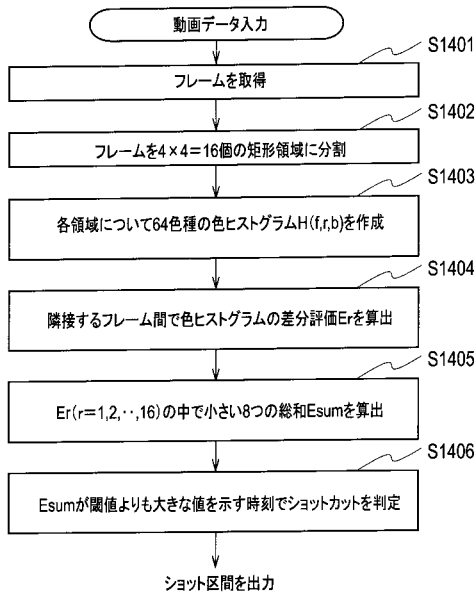
【 図 1 6 】



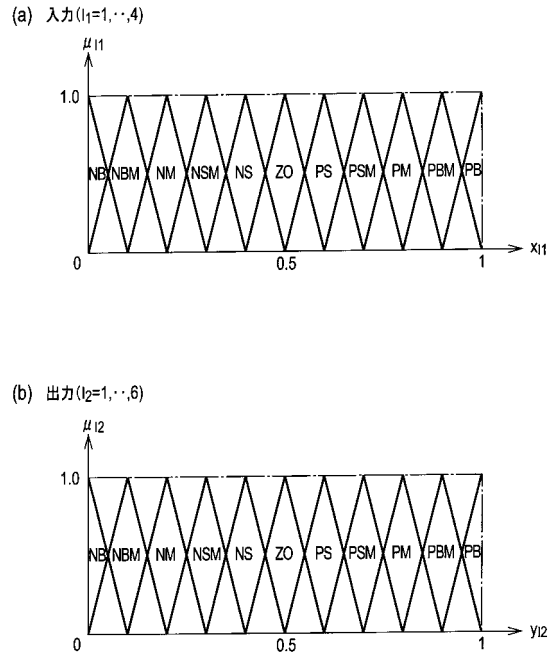
【 図 1 7 】



【 図 1 8 】



【 図 1 9 】



【 図 2 0 】

ファジィ制御規則1(Si,Sp,Mu,No)
(i1=1,...,4;それぞれSi,Sp,Mu,Noに対応)

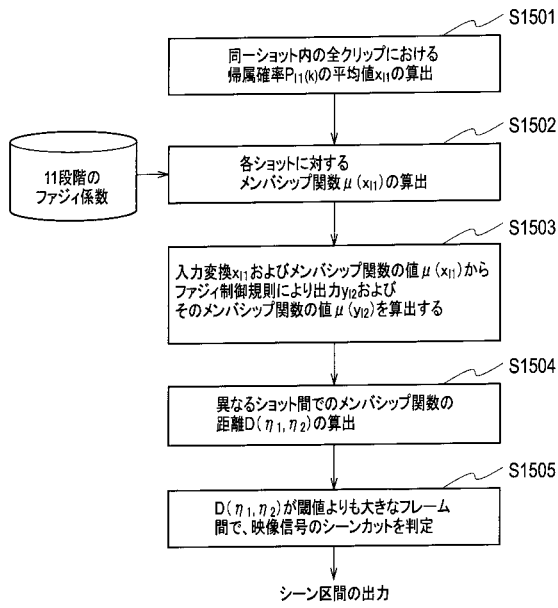
R_{11}^1	: IF "x ₁₁ is NB" THEN "y ₁₁ is NB"
R_{11}^2	: IF "x ₁₁ is NBM" THEN "y ₁₁ is NBM"
R_{11}^3	: IF "x ₁₁ is NM" THEN "y ₁₁ is NM"
R_{11}^4	: IF "x ₁₁ is NSM" THEN "y ₁₁ is NSM"
R_{11}^5	: IF "x ₁₁ is NS" THEN "y ₁₁ is NS"
R_{11}^6	: IF "x ₁₁ is ZO" THEN "y ₁₁ is ZO"
R_{11}^7	: IF "x ₁₁ is PS" THEN "y ₁₁ is PS"
R_{11}^8	: IF "x ₁₁ is PSM" THEN "y ₁₁ is PSM"
R_{11}^9	: IF "x ₁₁ is PM" THEN "y ₁₁ is PM"
R_{11}^{10}	: IF "x ₁₁ is PBM" THEN "y ₁₁ is PBM"
R_{11}^{11}	: IF "x ₁₁ is PB" THEN "y ₁₁ is PB"

【 図 2 1 】

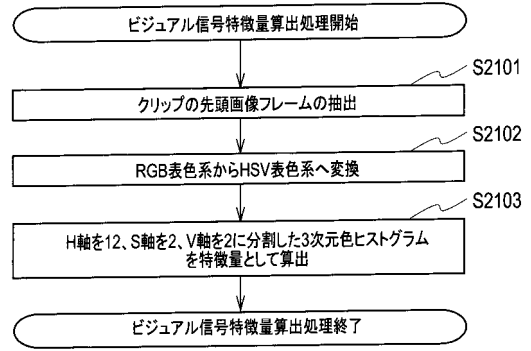
ファジィ制御規則2(SpMu,SpNo)

R_5^1	: IF "x ₁ is NB" and "x ₃ is NBM" and "x ₄ is NBM" THEN "y ₅ is NB"
R_5^2	: IF "x ₁ is NB" and "x ₂ is PS" and "x ₃ is NSM" THEN "y ₅ is PM"
R_5^3	: IF "x ₁ is NB" and "x ₂ is PM" and "x ₃ is NBM" THEN "y ₅ is PBM"
R_5^4	: IF "x ₁ is NB" and "x ₂ is PSM" and "x ₃ is NM" THEN "y ₅ is PB"
R_6^1	: IF "x ₁ is NB" and "x ₃ is NBM" and "x ₄ is NBM" THEN "y ₆ is NB"
R_6^2	: IF "x ₁ is NB" and "x ₂ is PS" and "x ₄ is NSM" THEN "y ₆ is PM"
R_6^3	: IF "x ₁ is NB" and "x ₂ is PM" and "x ₄ is NBM" THEN "y ₆ is PBM"
R_6^4	: IF "x ₁ is NB" and "x ₂ is PSM" and "x ₄ is NM" THEN "y ₆ is PB"

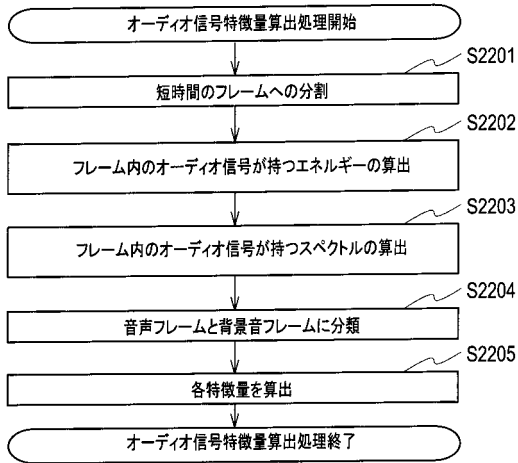
【 図 2 2 】



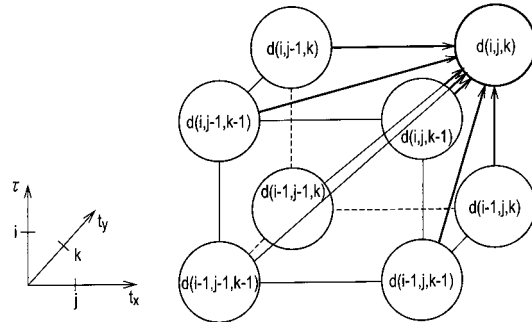
【 図 2 3 】



【 図 2 4 】

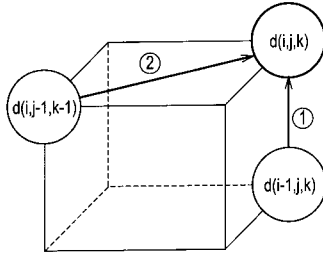


【 図 2 5 】

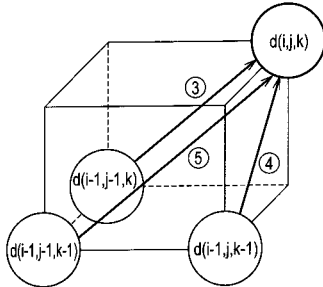


【 図 2 6 】

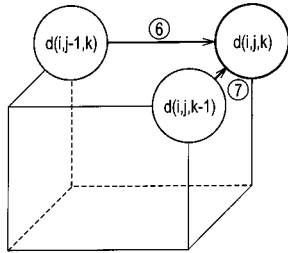
(a) バス#1,#2



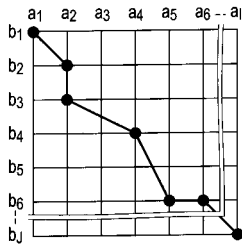
(b) バス#3~#5



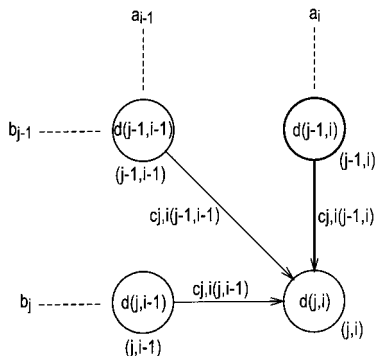
(c) バス#6,#7



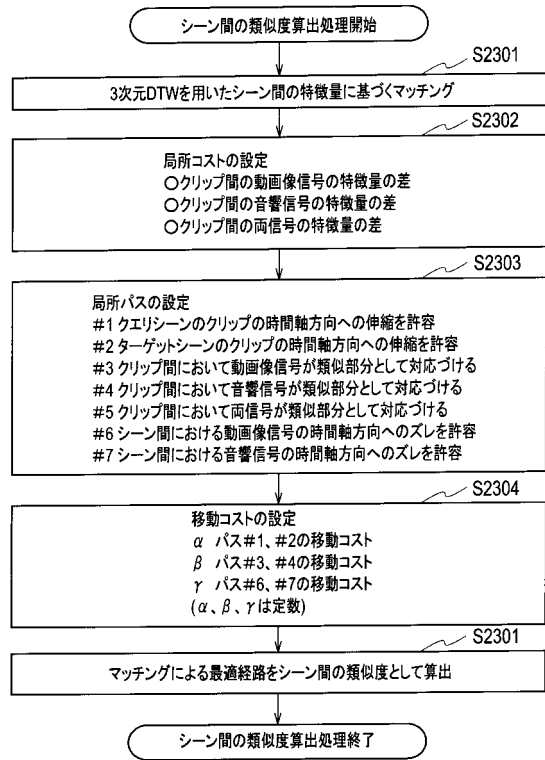
【 図 2 8 】



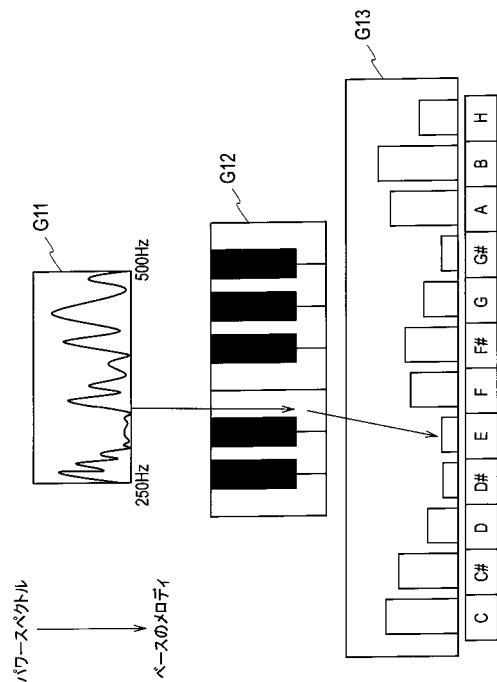
【 図 2 9 】



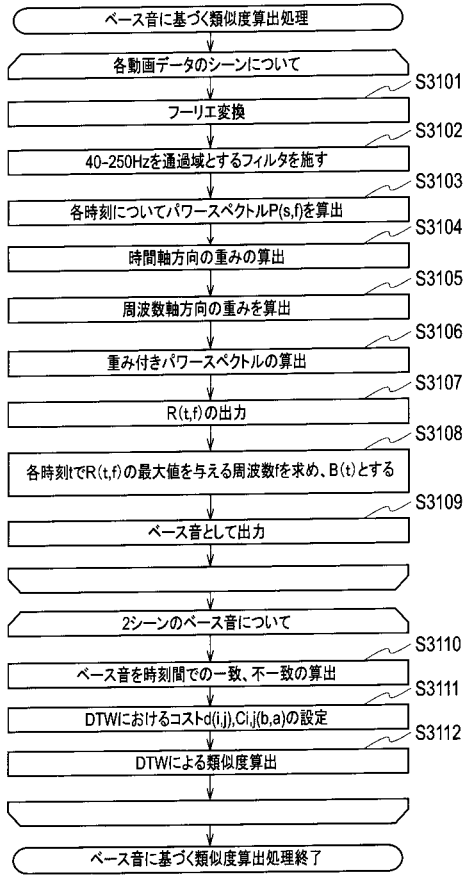
【 図 2 7 】



【 図 3 0 】



【 図 3 1 】

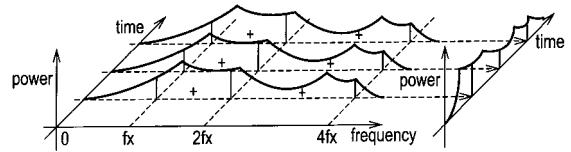


【 図 3 2 】

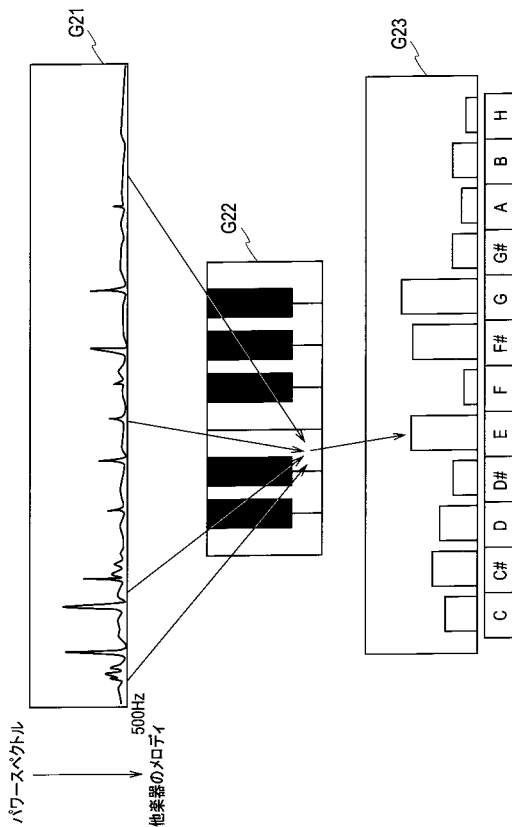
各音名の基本周波数

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

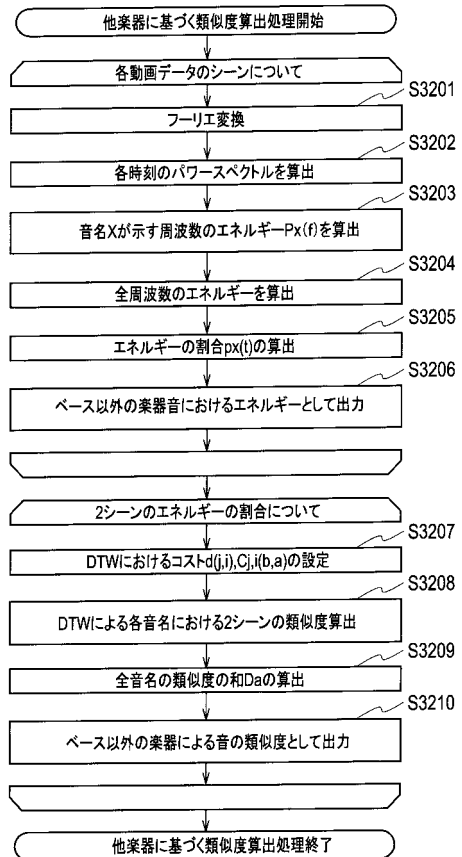
【 図 3 3 】



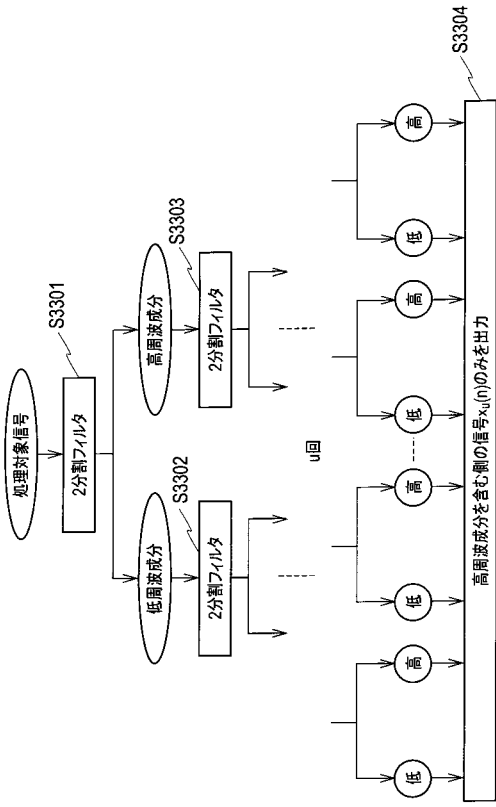
【 図 3 4 】



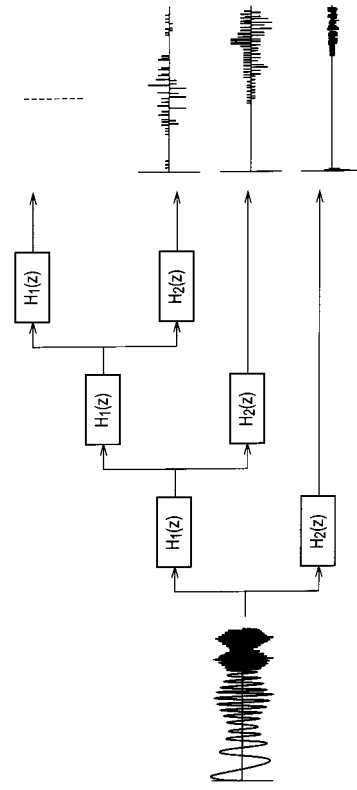
【 図 3 5 】



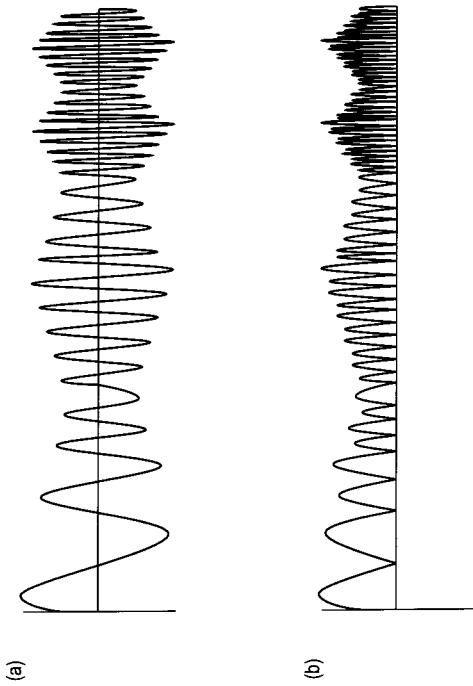
【 図 3 6 】



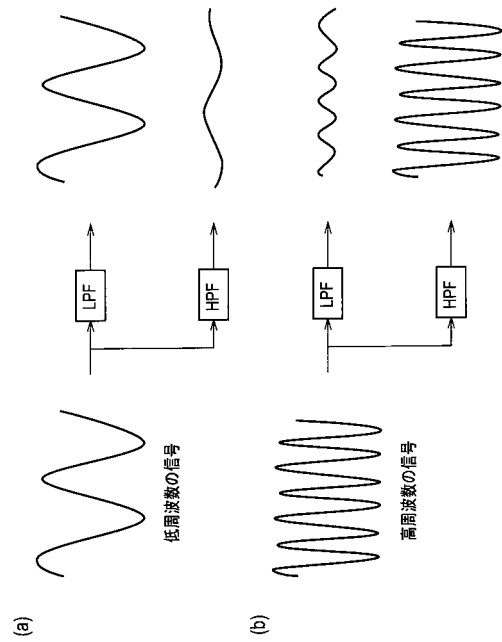
【 図 3 7 】



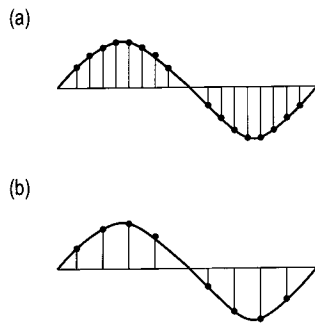
【 図 3 8 】



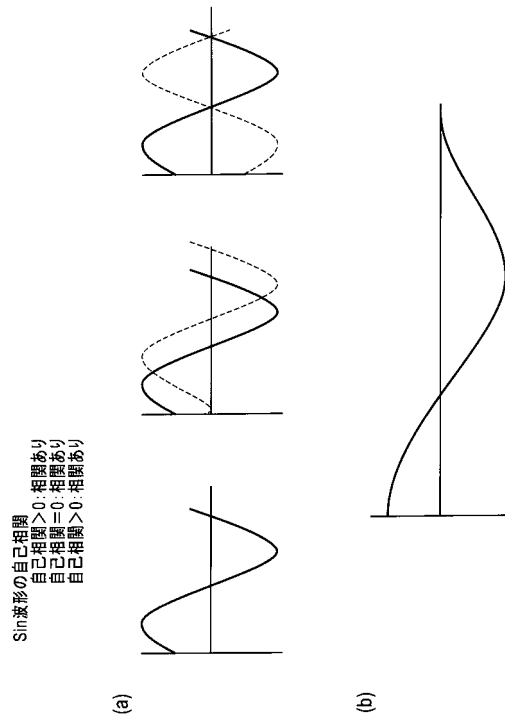
【 図 3 9 】



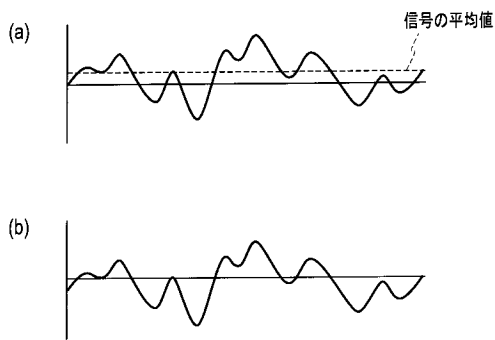
【 図 4 0 】



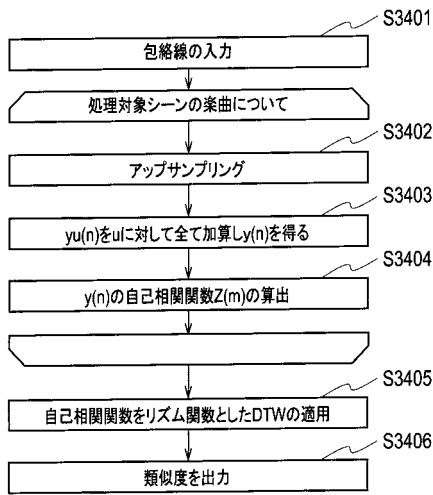
【 図 4 2 】



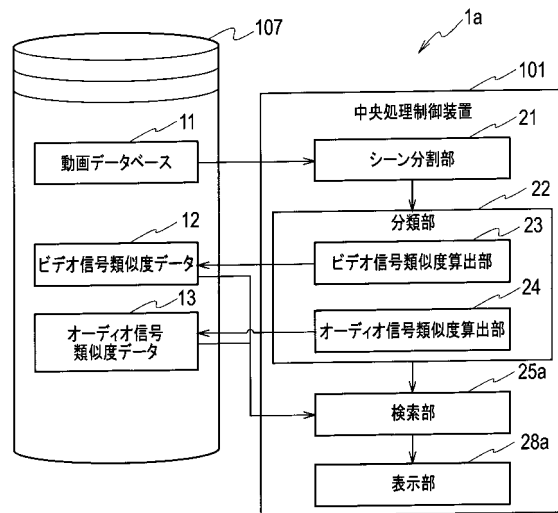
【 図 4 1 】



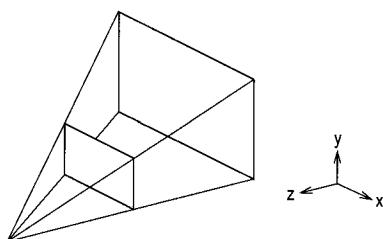
【 図 4 3 】



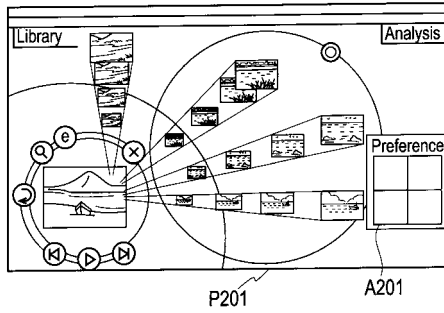
【 図 4 5 】



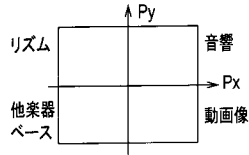
【 図 4 4 】



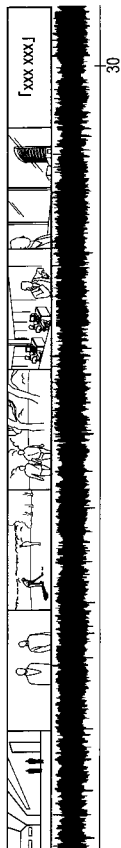
【 図 4 6 】



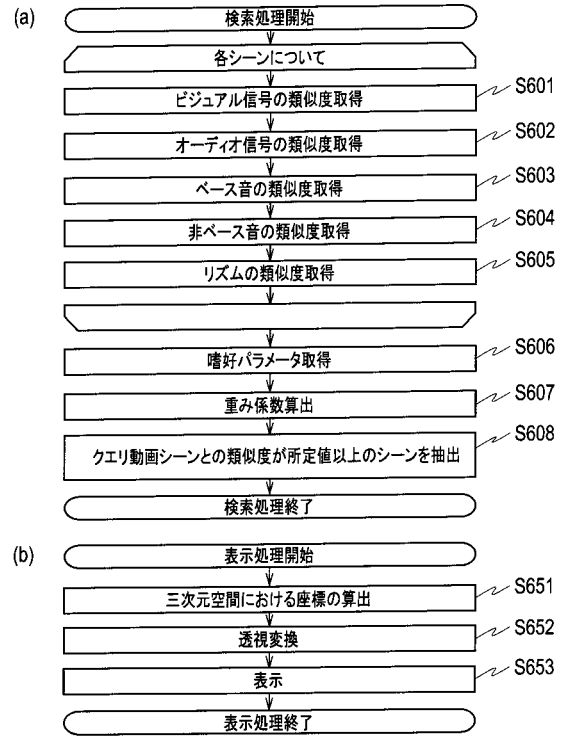
【 図 4 7 】



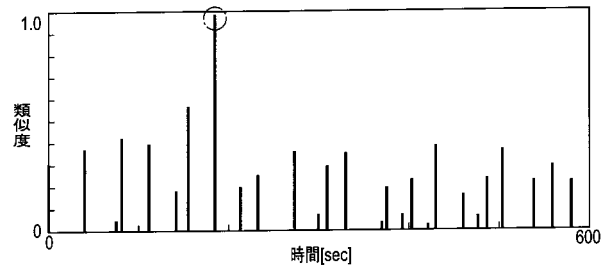
【 図 4 9 】



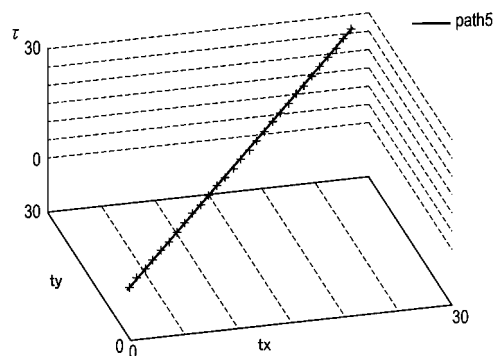
【 図 4 8 】



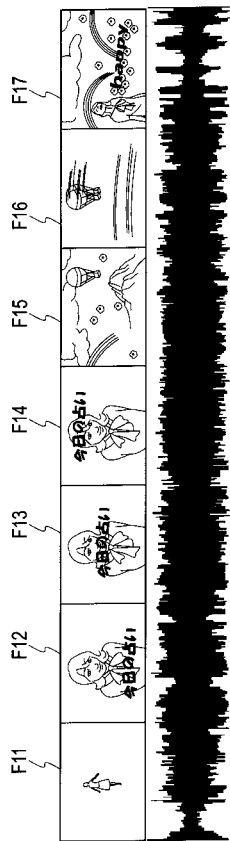
【 図 5 0 】



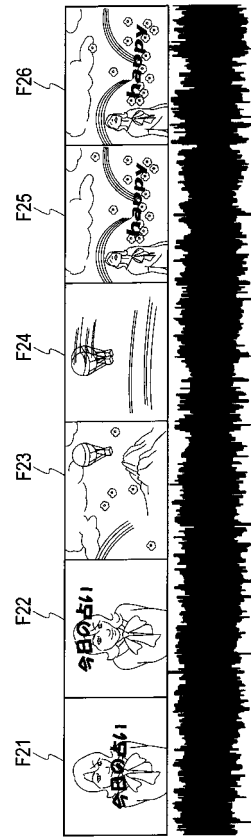
【 図 5 1 】



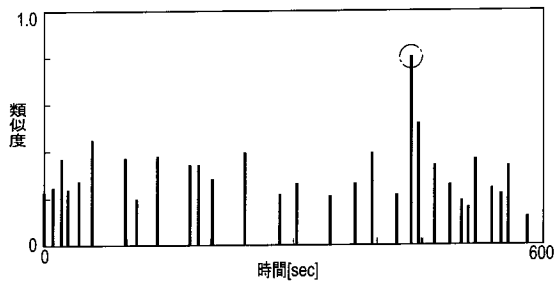
【 図 5 2 】



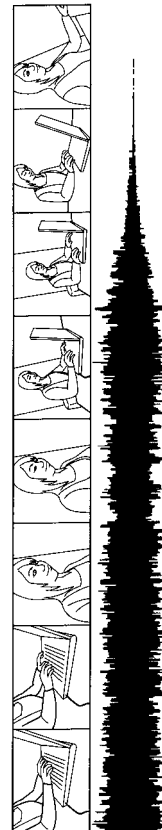
【 図 5 3 】



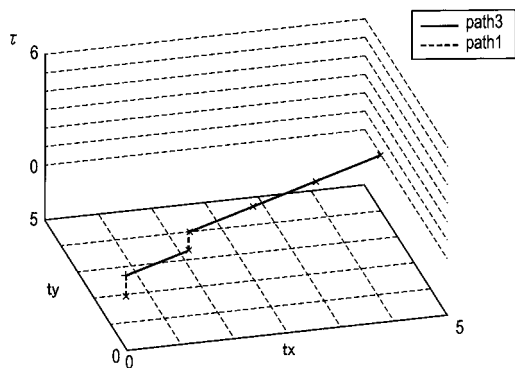
【 図 5 4 】



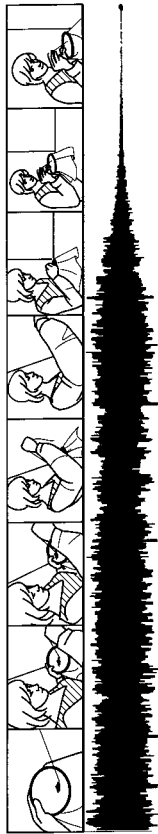
【 図 5 6 】



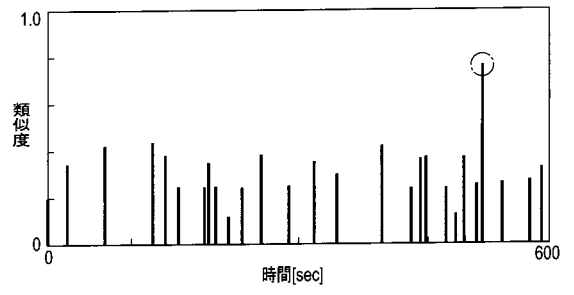
【 図 5 5 】



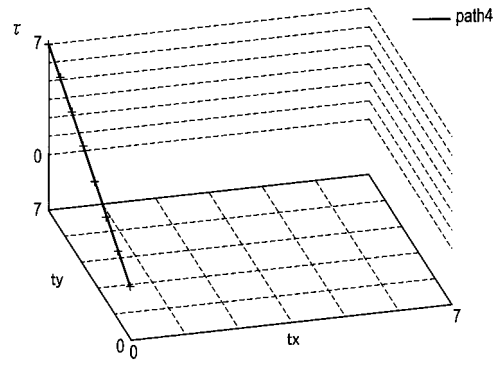
【 図 5 7 】



【 図 5 8 】



【 図 5 9 】



【 国際調査報告 】

INTERNATIONAL SEARCH REPORT		International application No. PCT/JP2009/055315
A. CLASSIFICATION OF SUBJECT MATTER H04N5/76(2006.01)i, G06F17/30(2006.01)i, G10L11/00(2006.01)i According to International Patent Classification (IPC) or to both national classification and IPC		
B. FIELDS SEARCHED Minimum documentation searched (classification system followed by classification symbols) H04N5/76, G06F17/30, G10L11/00 Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched Jitsuyo Shinan Koho 1922-1996 Jitsuyo Shinan Toroku Koho 1996-2009 Kokai Jitsuyo Shinan Koho 1971-2009 Toroku Jitsuyo Shinan Koho 1994-2009 Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	JP 2006-14084 A (Hiroshima University), 12 January, 2006 (12.01.06), Par. Nos. [0058] to [0061], [0107] to [0110]	1-4, 6-7, 9-12, 14-15, 17-22
Y	(Family: none)	5, 8, 13, 16
Y	JP 2008-5167 A (National University Corporation Hokkaido University), 10 January, 2008 (10.01.08), Par. No. [0009] (Family: none)	5, 8, 13, 16
A	JP 2005-252859 A (KDDI Corp.), 15 September, 2005 (15.09.05), Full text; Figs. 1 to 6 (Family: none)	1-22
<input checked="" type="checkbox"/> Further documents are listed in the continuation of Box C. <input type="checkbox"/> See patent family annex.		
* Special categories of cited documents: "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier application or patent but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed		"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "&" document member of the same patent family
Date of the actual completion of the international search 26 May, 2009 (26.05.09)		Date of mailing of the international search report 02 June, 2009 (02.06.09)
Name and mailing address of the ISA/ Japanese Patent Office		Authorized officer Telephone No.
Facsimile No.		

INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2009/055315

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JP 2004-159331 A (Microsoft Corp.), 03 June, 2004 (03.06.04), Full text; Figs. 1 to 8 & US 2004/0085341 A1 & EP 1416490 A1	1-22

国際調査報告		国際出願番号 PCT/JP2009/055315									
A. 発明の属する分野の分類 (国際特許分類 (IPC)) Int.Cl. H04N5/76(2006.01)i, G06F17/30(2006.01)i, G10L11/00(2006.01)i											
B. 調査を行った分野 調査を行った最小限資料 (国際特許分類 (IPC)) Int.Cl. H04N5/76, G06F17/30, G10L11/00											
最小限資料以外の資料で調査を行った分野に含まれるもの <table border="0"> <tr> <td>日本国実用新案公報</td> <td>1922-1996年</td> </tr> <tr> <td>日本国公開実用新案公報</td> <td>1971-2009年</td> </tr> <tr> <td>日本国実用新案登録公報</td> <td>1996-2009年</td> </tr> <tr> <td>日本国登録実用新案公報</td> <td>1994-2009年</td> </tr> </table>				日本国実用新案公報	1922-1996年	日本国公開実用新案公報	1971-2009年	日本国実用新案登録公報	1996-2009年	日本国登録実用新案公報	1994-2009年
日本国実用新案公報	1922-1996年										
日本国公開実用新案公報	1971-2009年										
日本国実用新案登録公報	1996-2009年										
日本国登録実用新案公報	1994-2009年										
国際調査で使用した電子データベース (データベースの名称、調査に使用した用語)											
C. 関連すると認められる文献											
引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求項の番号									
X	JP 2006-14084 A (国立大学法人広島大学) 2006.01.12, 【0058】 - 【0061】, 【0107】 - 【0110】 (ファミリーなし)	1-4, 6-7, 9-12, 14-15, 17-22									
Y		5, 8, 13, 16									
Y	JP 2008-5167 A (国立大学法人北海道大学) 2008.01.10, 【0009】 (ファミリーなし)	5, 8, 13, 16									
<input checked="" type="checkbox"/> C欄の続きにも文献が列挙されている。		<input type="checkbox"/> パテントファミリーに関する別紙を参照。									
* 引用文献のカテゴリー		の日の後に公表された文献									
「A」特に関連のある文献ではなく、一般的技術水準を示すもの		「T」国際出願日又は優先日後に公表された文献であって出願と矛盾するものではなく、発明の原理又は理論の理解のために引用するもの									
「E」国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの		「X」特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの									
「L」優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献 (理由を付す)		「Y」特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの									
「O」口頭による開示、使用、展示等に言及する文献		「&」同一パテントファミリー文献									
「P」国際出願日前で、かつ優先権の主張の基礎となる出願											
国際調査を完了した日 26.05.2009		国際調査報告の発送日 02.06.2009									
国際調査機関の名称及びあて先 日本国特許庁 (ISA/JP) 郵便番号100-8915 東京都千代田区霞が関三丁目4番3号		特許庁審査官 (権限のある職員) 木方 庸輔	5C 9649								
		電話番号 03-3581-1101 内線 3541									

国際調査報告

国際出願番号 PCT/JP2009/055315

C (続き) . 関連すると認められる文献		
引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求項の番号
A	JP 2005-252859 A (KDDI 株式会社) 2005.09.15, 全文, 【図1】 - 【図6】 (ファミリーなし)	1-22
A	JP 2004-159331 A (マイクロソフト コーポレーション) 2004.06.03, 全文, 【図1】 - 【図8】 & US 2004/0085341 A1 & EP 1416490 A1	1-22

フロントページの続き

(81)指定国 AP(BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), EA(AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), EP(AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, SE, SI, SK, TR), OA(BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG), AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW

(注)この公表は、国際事務局(WIPO)により国際公開された公報を基に作成したものである。なおこの公表に係る日本語特許出願(日本語実用新案登録出願)の国際公開の効果は、特許法第184条の10第1項(実用新案法第48条の13第2項)により生ずるものであり、本掲載とは関係ありません。