

(19) 日本国特許庁(JP)

(12) 公表特許公報(A)

(11) 特許出願公表番号

特表2006-525537

(P2006-525537A)

(43) 公表日 平成18年11月9日(2006.11.9)

(51) Int. Cl.	F I	テーマコード (参考)
G10L 11/00 (2006.01)	G10L 11/00 402C	5B075
H04N 5/91 (2006.01)	H04N 5/91 N	5C053
G06F 17/30 (2006.01)	G06F 17/30 170E	5D015
G10L 15/02 (2006.01)	G06F 17/30 220A	5D077
G11B 27/10 (2006.01)	G10L 15/02 300D	5D110

審査請求 未請求 予備審査請求 未請求 (全 30 頁) 最終頁に続く

(21) 出願番号 特願2006-506452 (P2006-506452)
 (86) (22) 出願日 平成16年4月2日(2004.4.2)
 (85) 翻訳文提出日 平成17年10月12日(2005.10.12)
 (86) 国際出願番号 PCT/IB2004/001068
 (87) 国際公開番号 W02004/090752
 (87) 国際公開日 平成16年10月21日(2004.10.21)
 (31) 優先権主張番号 60/462,777
 (32) 優先日 平成15年4月14日(2003.4.14)
 (33) 優先権主張国 米国(US)
 (31) 優先権主張番号 60/509,800
 (32) 優先日 平成15年10月8日(2003.10.8)
 (33) 優先権主張国 米国(US)

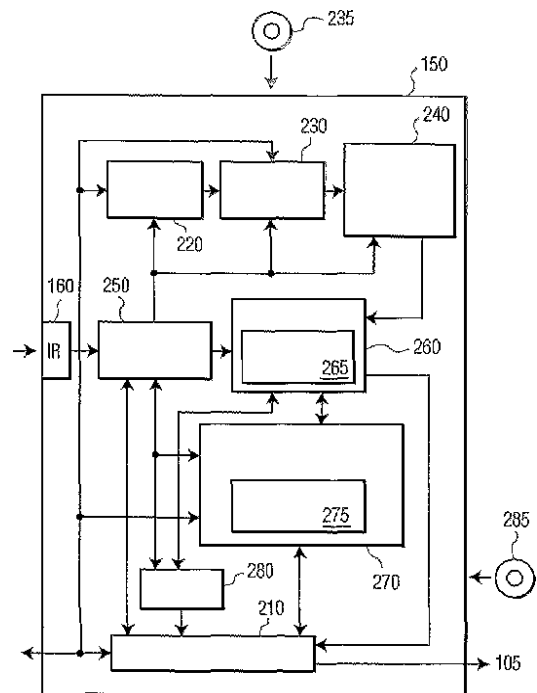
(71) 出願人 590000248
 コーニンクレッカ フィリップス エレクトロニクス エヌ ヴィ
 Koninklijke Philips Electronics N.V.
 オランダ国 5621 ペーアー アインドーフエン フルーネヴァウツウェッハ 1
 Groenewoudseweg 1, 5621 BA Eindhoven, The Netherlands
 (74) 代理人 100070150
 弁理士 伊東 忠彦
 (74) 代理人 100091214
 弁理士 大貫 進介

最終頁に続く

(54) 【発明の名称】 コンテンツ解析を用いて音楽ビデオを要約する方法及び装置

(57) 【要約】

コンテンツ解析を用いてマルチメディア・ストリーム(505)における音楽ビデオ(507)をセグメント化し、要約する方法及び装置を備える。音楽ビデオ(507)は、マルチメディア・ストリームに関する複数のコンテンツ特徴を評価することによってマルチメディア・ストリーム(505)においてセグメント化される。複数のコンテンツ特徴は、顔存在特徴と、ビデオテキスト存在特徴と、色ヒストグラム特徴と、オーディオ特徴と、カメラ・カット特徴と、少なくとも1つの音楽ビデオのトランスクリプトから得られるキーワードの解析とうちの少なくとも2つを有する。複数のコンテンツ特徴を、ベイジアン確率ネットワークなどのパターン認識エンジン(1000)又は1つ若しくは複数のビデオ・セグメント化ルール(1115)を用いて処理して、マルチメディア・ストリーム(505)における音楽ビデオ(505)を識別する。コーラスは、少なくとも1つの音楽ビデオ(507)において音楽ビデオ(507)のトランスクリプト(T)を用いてトランスクリプトにおける語の繰り返しに基づいて検出される。抽出されたコーラスは、音楽ビデオ(507)の要約を自動的に生成す



【特許請求の範囲】

【請求項 1】

マルチメディア・ストリームにおける音楽ビデオをセグメント化する方法であって：
少なくとも1つの音楽ビデオを有するマルチメディア・ストリームを受信する工程；
該マルチメディア・ストリームからの前記少なくとも1つの音楽ビデオを、前記マルチメディア・ストリームに関する複数のコンテンツ特徴を評価することによってセグメント化する工程；及び

前記少なくとも1つの音楽ビデオを識別する工程を備えることを特徴とする方法。

【請求項 2】

請求項1記載の方法であって、更に、前記少なくとも1つの音楽ビデオの要約を生成する工程を備えることを特徴とする方法。 10

【請求項 3】

請求項1記載の方法であって、前記少なくとも1つの音楽ビデオの要約がユーザに対して、個人化された選好に基づいて提示されることを特徴とする方法。

【請求項 4】

請求項1記載の方法であって、前記少なくとも1つの音楽ビデオをユーザによって、個人化された選好に基づいて取り出し得ることを特徴とする方法。

【請求項 5】

請求項1記載の方法であって、パターン認識エンジンを用いて前記複数のコンテンツ特徴を処理して前記少なくとも1つの音楽ビデオを識別することを特徴とする方法。 20

【請求項 6】

請求項1記載の方法であって、ベイジアン確率ネットワークを用いて前記複数のコンテンツ特徴を処理して前記少なくとも1つの音楽ビデオを識別することを特徴とする方法。

【請求項 7】

請求項1記載の方法であって、1つ又は複数のビデオ・セグメント化ルールを用いて前記複数のコンテンツ特徴を処理して前記少なくとも1つの音楽ビデオを識別することを特徴とする方法。

【請求項 8】

請求項1記載の方法であって、前記複数のコンテンツ特徴は、前記マルチメディア・ストリームにおける顔の提示におけるパターンを評価するよう顔存在特徴を有することを特徴とする方法。 30

【請求項 9】

請求項1記載の方法であって、前記複数のコンテンツ特徴は、ビデオテキストが前記マルチメディア・ストリームにおいて現れる時点を判定するビデオテキスト存在特徴を有することを特徴とする方法。

【請求項 10】

請求項1記載の方法であって、前記複数のコンテンツ特徴は、前記マルチメディア・ストリームの色コンテンツにおけるパターンを評価するよう色ヒストグラム特徴を有することを特徴とする方法。

【請求項 11】

請求項1記載の方法であって、前記複数のコンテンツ特徴は、前記マルチメディア・ストリームにおけるカメラのカット及び動きにおけるパターンを評価するようカメラ・カット特徴を有することを特徴とする方法。 40

【請求項 12】

請求項1記載の方法であって、前記複数のコンテンツ特徴は、前記少なくとも1つの音楽ビデオのトランスクリプトから得られるキーワードの解析を有することを特徴とする方法。

【請求項 13】

請求項1記載の方法であって、前記複数のコンテンツ特徴は、前記マルチメディア・ストリームから直接導き出される低レベル特徴の解析を有することを特徴とする方法。 50

【請求項 14】

請求項13記載の方法であって、前記低レベル特徴は、いくつかのエッジと、いくつかの形状と、いくつかの局所的な動きと、いくつかの大局的な動きとのうちの1つ又は複数を有することを特徴とする方法。

【請求項 15】

請求項1記載の方法であって、前記複数のコンテンツ特徴は、オーディオ特徴を有することを特徴とする方法。

【請求項 16】

請求項15記載の方法であって、前記オーディオ特徴は、1つのマルチメディア・ストリームを評価することを特徴とする方法。 10

【請求項 17】

請求項15記載の方法であって、前記オーディオ特徴は、メル周波数ケプストラム係数(MFCC)と、線形予測係数(LPC)と、ピッチにおける変動と、帯域幅における変動と、音量における変動と、音色における変動とのうちの1つ又は複数を評価することを特徴とする方法。

【請求項 18】

請求項1記載の方法であって、更に、前記少なくとも1つの音楽ビデオを識別する情報を外部ソースから得る工程を備えることを特徴とする方法。

【請求項 19】

少なくとも1つの音楽ビデオにおいてコーラスを検出する方法であって：

前記少なくとも1つの音楽ビデオを備えるマルチメディア・ストリームを受信する工程；

前記少なくとも1つの音楽ビデオに関連したトランスクリプトをアクセスする工程；及び

前記コーラスを前記トランスクリプトにおける語の繰り返しの基づいて検出する工程を備えることを特徴とする方法。 20

【請求項 20】

請求項19記載の方法であって、前記トランスクリプトは、クローズド・キャプション情報から得られることを特徴とする方法。 30

【請求項 21】

請求項19記載の方法であって、前記コーラスは、前記少なくとも1つの音楽ビデオの要約の自動的な生成に用いられることを特徴とする方法。

【請求項 22】

請求項19記載の方法であって、更に、前記繰り返される語を検出し、クラスタ化する工程を備えることを特徴とする方法。

【請求項 23】

請求項19記載の方法であって、前記検出する工程は更に、前記マルチメディア・ストリームに関する別のコンテンツ特徴に基づくことを特徴とする方法。

【請求項 24】

請求項19記載の方法であって、更に、前記少なくとも1つの音楽ビデオを識別する情報を外部ソースから得る工程を備えることを特徴とする方法。 40

【請求項 25】

マルチメディア・ストリームにおける音楽ビデオをセグメント化する装置であって：
メモリ；及び

少なくとも1つの音楽ビデオを備えるマルチメディア・ストリームを受信する工程と、
該マルチメディア・ストリームに関する複数のコンテンツ特徴をパターン認識エンジンに適用して、前記マルチメディア・ストリームからの前記少なくとも1つの音楽ビデオをセグメント化する工程と、

前記少なくとも1つの音楽ビデオを識別する工程とを行うよう動作する、前記メモリに 50

結合される少なくとも1つのコントローラを備えることを特徴とする装置。

【請求項26】

請求項25記載の装置であって、前記パターン認識エンジンがベイジアン確率ネットワークであることを特徴とする装置。

【請求項27】

請求項25記載の装置であって、前記パターン認識エンジンがニューラル・ネットワークであることを特徴とする装置。

【請求項28】

請求項25記載の装置であって、前記パターン認識エンジンが自己回帰移動平均法を用いることを特徴とする装置。

10

【請求項29】

請求項25記載の装置であって、前記複数のコンテンツ特徴が、顔存在特徴と、ビデオテキスト存在特徴と、色ヒストグラム特徴と、カメラ・カット特徴と、前記少なくとも1つの音楽ビデオのトランスクリプトから得られるキーワードの解析とのうちの少なくとも2つを有することを特徴とする装置。

【請求項30】

マルチメディア・ストリームにおける音楽ビデオをセグメント化する装置であって：
メモリ；及び

少なくとも1つの音楽ビデオを備えるマルチメディア・ストリームを受信する工程と、
該マルチメディア・ストリームに関する複数のコンテンツ特徴を1つ又は複数のビデオ
・セグメント化ルールに適用して、前記マルチメディア・ストリームからの前記少なくと
も1つの音楽ビデオをセグメント化する工程と、

20

前記少なくとも1つの音楽ビデオを識別する工程とを行うよう動作する、前記メモリに
結合される少なくとも1つのコントローラを備えることを特徴とする装置。

【請求項31】

請求項30記載の装置であって、前記複数のコンテンツ特徴が、顔存在特徴と、ビデオテキスト存在特徴と、色ヒストグラム特徴と、カメラ・カット特徴と、前記少なくとも1つの音楽ビデオのトランスクリプトから得られるキーワードの解析とのうちの少なくとも2つを有することを特徴とする装置。

【請求項32】

請求項30記載の装置であって、前記1つ又は複数のビデオ・セグメント化ルールは、前記複数のコンテンツ特徴に対する閾値を規定して、ビデオ・セグメントが生起した時点を判定することを特徴とする装置。

30

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、一般的に、ビデオ要約手法に関し、特に、音楽ビデオの索引付け及び要約を行う方法及び装置に関する。

【背景技術】

【0002】

音楽ビデオ番組は、フューズ(Fuse)、VH1、MTV、MTV2を有するいくつかのテレビジョン・チャンネル上で利用可能である。www.buymusic.comなどのいくつかの人気ウェブ・サイトが、個々の楽曲のオーディオ部分をユーザがブラウジングし、取得することを可能にする一方、ビデオ・レコーダや他のビデオ・ベースのアプリケーションは、複数の音楽ビデオを備えている番組を有する番組全体をユーザが取得することを可能にするに過ぎない。個々の音楽ビデオを自動的に取得する方法は現在、全く存在しない。よって、1つ又は複数の音楽ビデオを備える番組全体を視聴者が記録する場合、記録は、広告やコメンタリなどの非音楽ビデオ部分も全て有することになる。音楽ビデオを視聴するには、視聴者は、所望の音楽ビデオ部分に達するまで非音楽ビデオ部分にわたって記録を早送りすることを要する。更に、ビデオ再生装置の大量の記録容量が、広告や他のトークなどの無用のマテリア

40

50

ルを記録するのに用いられる。

【0003】

ハイライト部分などの、番組の特定の部分に対する高レベルのアクセスを備えるよう、コンテンツ解析手法が提案又は示唆されている。ニュース、スポーツ及び映画を有する、多くの種類の番組についてビデオ要約手法が開発されている。例えば、「インフォメディア・プロジェクト (InforMedia Project)」は、音声認識、自然言語の理解及び字幕テキストに主に基づいて各ビデオの短い要約を作成するデジタル・ビデオ・ライブラリ・システムである (A. Hauptmann 及び M. Smithによる「Text, Speech, and Vision for Video Segmentation: The InforMedia Project」, American Association for Artificial Intelligence (AAAI), Fall, 1995 Symposium on Computational Models for Integrating Language and Vision (1995)参照。)

10

【0004】

しかし、音楽の解析及び取り出しの分野における研究は主に、オーディオの特徴に集中している。例えば、B. Logan及びS. Chuによる「Music Summarization Using Key Phrases」, Int'l Conf. on Acoustics, Speech and Signal Processing, 2000は、ポピュラー音楽のセレクションにおけるキーフレーズを見つけてオーディオ・サムネールを生成するアルゴリズムを開示している。J. Footeによる「Visualizing Music and Audio Using Self Similarity」 Proc. ACM Multimedia '99, 77-80, Nov. 1999は、オーディオの斬新さの尺度の応用としてオーディオの「要約化 (gisting)」を紹介した。このオーディオの斬新さの得点は、オーディオから抽出される特徴に基づいてオーディオのフレームを比較する類似度マトリックスに基づいている。

20

【発明の開示】

【発明が解決しようとする課題】

【0005】

よって、音楽コンテンツ解析が、研究の活発な分野である一方、音楽ビデオの解析及び要約を行う手法の改善に対する必要性がなお存在している。マルチメディア・データ・ストリームにおいて音楽ビデオをセグメント化し、適切な音楽ビデオ情報を備える要約を音楽ビデオ毎に作成する方法及び装置に対する更なる必要性が存在している。

【課題を解決するための手段】

【0006】

概括的には、コンテンツ解析を用いてマルチメディア・ストリームにおける音楽ビデオをセグメント化し、要約する方法及び装置が備えられる。音楽ビデオは、マルチメディア・ストリームにおいて本発明によってマルチメディア・ストリームに関する複数のコンテンツ特徴を評価することによってセグメント化される。複数のコンテンツ特徴は、顔存在特徴と、ビデオテキスト存在特徴と、色ヒストグラム特徴と、オーディオ特徴と、カメラ・カット特徴と、少なくとも1つの音楽ビデオのトランスクリプトから得られるキーワードの解析とのうちの少なくとも2つを有する。複数のコンテンツ特徴は、ベイジアン確率ネットワーク (Bayesian Belief Network) などのパターン認識エンジン又は、1つ若しくは複数のビデオ・セグメント化ルールを用いて処理されて、マルチメディア・ストリームにおける音楽ビデオを識別する。

30

40

【0007】

本発明の一特徴によれば、顔存在特徴は、マルチメディア・ストリームにおける顔の表示におけるパターンを評価する。当初、いくつかの考えられる顔種類ラベルの1つが、各画像フレームに割り当てられる。画像フレームは更に、割り当てられた顔種類ラベルに基づいてクラスタ化され、パターンが、顔種類ラベルのクラスタにおいて解析されて、ビデオ境界を検出する。本発明の別の特徴では、色ヒストグラム特徴は、マルチメディア・ストリームの色コンテンツにおけるパターンを評価する。色ヒストグラムが画像フレーム毎に取得され、画像フレームが更に、ヒストグラムに基づいてクラスタ化される。パターンをヒストグラムのクラスタにおいて解析してビデオ境界を検出する。カメラ・カット特徴は、マルチメディア・ストリームにおけるカメラのカット及び動きを評価する。オーディ

50

オ特徴は、マルチメディア・ストリームのオーディオ・コンテンツにおけるパターンを評価するよう明らかにされる。例えば、1つのマルチメディア・ストリームを評価して、増加する音量と減少する音量との各々によって示される楽曲の始めと終わりとを検出する。

【0008】

本発明の別の特徴によれば、コーラスが、少なくとも1つの音楽ビデオにおいて検出される。受信マルチメディア・ストリームにおける音楽ビデオに関連したトランスクリプトがアクセスされ、コーラスが、トランスクリプトにおける語の繰り返しに基づいて検出される。トランスクリプトは、例えば、クロズド・キャプション情報から取得し得る。抽出コーラスを用いて、音楽ビデオの要約を自動的に要約し得る。生成された要約は、ユーザ選好によってユーザに提示することが可能であり、音楽ビデオをユーザ選好によって取り出すのに用い得る。

10

【発明を実施するための最良の形態】

【0009】

本発明、更には、本発明の更なる特徴及び利点、の更に徹底的な理解は、以下の詳細な説明及び添付図面を参照することによって得られる。

【実施例】

【0010】

図1は、本発明の一実施例による例示的なビデオ再生装置150及びテレビジョン受信機105を示す。ビデオ再生装置150は、ケーブル・テレビジョン・サービス・プロバイダ、局所アンテナ、インターネット・サービス・プロバイダ(ISP)、DVDプレーヤ又はVHSテープ・プレーヤなどの外部ソースから入力テレビジョン信号を受信する。ビデオ再生装置150は、視聴者選択チャンネルからのテレビジョン信号をテレビジョン受信機105に送信する。チャンネルは、ユーザによって手作業で選択されてもよく、ユーザによって先行してプログラムされる記録装置によって自動的に選択されてもよい。あるいは、チャンネル及びビデオ番組は、ユーザの個人視聴履歴における番組プロファイルからの情報に基づいて記録装置によって自動的に選択してもよい。本発明は例示的なテレビジョン受信機の意味合いで説明しているが、本発明の例示の実施例は何れかの種類のビデオ表示システムにおいて用いるよう容易に修正し得るということが当業者は認識するものである。

20

【0011】

レコード・モードでは、ビデオ再生装置150は入力無線周波数(RF)テレビジョン信号を復調してベースバンド・ビデオ信号を生成し、ベースバンド・ビデオ信号は、ビデオ再生装置150内の記憶媒体上又はビデオ再生装置150に接続される記憶媒体上に記憶される。再生モードでは、ビデオ再生装置150は、ユーザによって選択される、記憶されているベースバンド・ビデオ信号(すなわち、番組)を記憶媒体から読み取り、それをテレビジョン受信機105に送信する。ビデオ再生装置150は、デジタル信号を受信し、記録し、デジタル信号と相互作用し、デジタル信号を再生することができるタイプのビデオ・レコーダを備え得る。

30

【0012】

ビデオ再生装置150は、記録テープを利用するタイプ、ハード・ディスクを利用するタイプ、ソリッドステート・メモリを利用するタイプや何れかの他の種類の記憶装置を利用するタイプのビデオ・レコーダを備え得る。ビデオ再生装置150がビデオ・カセット・レコーダ(VCR)である場合、ビデオ再生装置150は入力テレビジョン信号を磁気カセット・テープに記憶し、入力テレビジョン信号を磁気カセット・テープから取り出す。ビデオ再生装置150が、リプレイ(Replay)TV(商標)レコーダやティーボ(TiVo)(商標)レコーダなどのディスク・ドライブ・ベースの装置である場合、ビデオ再生装置150は入力テレビジョン信号を磁気カセット・テープではなくコンピュータ磁気ハード・ディスクに記憶し、入力テレビジョン信号を磁気カセット・テープではなくコンピュータ磁気ハード・ディスクから取り出し、記憶されているテレビジョン信号をハード・ディスクから取り出す。更に別の実施例では、ビデオ再生装置150は、局所読み取り/書き込み(R/W)デジタル多用途ディスク(DVD)又は読み取り/書き込み(R/W)コンパクト・ディスク(CD-RW)と

40

50

の間で記憶及び取り出しを行い得る。局所記憶媒体は、固定型（例えば、ハード・ディスク・ドライブ）であってもよく、取り外し可能型（例えば、DVD、CR-ROM）であってもよい。

【0013】

ビデオ再生装置150は、ユーザによって操作されるリモコン装置125から（チャンネル・アップ、チャンネル・ダウン、音量アップ、音量ダウン、記録、再生、早送り（FF）、逆送りなどの）コマンドを受信する赤外線（IR）センサ160を備える。テレビジョン受信機105は、画面110、赤外線（IR）センサ115及び（点線が示す）1つ又は複数の手動のつまみ120を備える通常のテレビジョン受信機である。IRセンサ115は、ユーザによって操作されるリモコン装置125から（音量アップ、音量ダウン、電源オン、電源オフなどの）コマンドも受信する。

10

【0014】

なお、ビデオ再生装置150は、特定の種類の入力テレビジョン信号を特定の種類のソースから受信することに限定されない。上記のように、外部ソースは、ケーブル・サービス・プロバイダ、通常のRF放送アンテナ、衛星放送アンテナ、インターネット接続や、DVDプレイヤー又はVHSテープ・プレイヤーなどの別の局所記憶装置であり得る。一部の実施例では、ビデオ再生装置150は、記録することもできない場合がある一方、取り外し可能なDVD又はCD-ROMから取り出されるテレビジョン信号を再生することに限定され得る。よって、入力信号は、デジタル信号、アナログ信号又はインターネット・プロトコル(IP)パケットであり得る。

20

【0015】

しかし、本発明の原理を説明するうえで単純にかつ明確にする目的で、以下の記載は概括的には、ビデオ再生装置150が（アナログ及び/又はデジタルの）入力テレビジョン信号をケーブル・サービス・プロバイダから受信する実施例に関するものである。しかし、本発明の原理が、無線放送テレビジョン信号、局所記憶システム、MPEGデータを備えるIPパケットの入力ストリームなどとともに用いるよう容易に形成し得るということを当業者は分かるものである。音楽ビデオがテレビジョン受信機105の画面100上に表示される場合、音楽ビデオの始めは通常、テキスト・キャプション180（ビデオテキスト）をビデオ画像の下に表示する。テキスト・キャプション180は通常、楽曲のタイトル、アルバムの名前、アーティスト又はグループの名前、リリース日付や他の同様な情報を通常、備えている。テキスト・キャプション180も通常、音楽ビデオの終わりに表示される。テキスト・キャプション180は、ビデオテキスト・ブロック180としても呼ぶこととする。音楽ビデオ要約コントローラ270は、記憶されている音楽ビデオ要約ファイル360の全てのリスト190をアクセスし、テレビジョン受信機105の画面110上にリスト190を表示することができる。すなわち、リスト190は、(1)マルチメディア・データ・ストリームにおいて検出された音楽ビデオ全ての音楽ビデオ要約ファイルと(2)音楽ビデオ各々をレコーディングしたアーティスト又はグループの識別情報を表示する。リモコン装置125及びIRセンサ160を用いて、ユーザは、「音楽ビデオ要約再生」制御信号を音楽ビデオ要約コントローラ270に送信して、リスト190において次に再生する音楽ビデオ要約ファイルを選択する。このようにして、ユーザは、音楽ビデオ要約ファイルが再生される順序を選択する。

30

40

【0016】

図2は、本発明の一実施例による例示的なビデオ再生装置150を更に詳細に示す。ビデオ再生装置150は、IRセンサ160、ビデオ・プロセッサ210、MPEG2符号器220、ハード・ディスク・ドライブ230、MPEG2復号器/NTSC符号器240及びビデオ・レコーダ（VR）コントローラ250を備える。ビデオ再生装置150は更に、フレーム・グラバ265を備えるビデオ装置260、クローズド・キャプション復号器275を備える音楽ビデオ要約コントローラ270及びメモリ280を備える。フレーム・グラバ265は、ビデオ・フレームをMPEG2復号器/NTSC符号器240の出力から捕捉し、記憶する。クローズド・キャプション復号器265は、MPEG2復号器/NTSC符号器240のNTSC出力信号におけるクローズド・キャプション・テキストを復号する。図2では、クローズド・キャプション復号器275は音楽ビデオ要約コントローラ270の中に

50

あるものとして示しているが、クロズド・キャプション復号器275が音楽ビデオ要約コントローラ270内にある必要はない。

【0017】

VRコントローラ250は、ビュー・モード、記録モード、再生モード、早送り（FF）モード、逆送りモードや他の同様な機能を有する、ビデオ再生装置150の動作全体を制御する。音楽ビデオ要約コントローラ270は、音楽ビデオ要約の作成、記憶及び再生を本発明の原理によって制御する。

【0018】

ビュー・モードでは、VRコントローラ250は、ビデオ信号をハード・ディスク・ドライブ230上に記憶すること（又はビデオ信号をハード・ディスク・ドライブ230から取り出すこと）の有無にかかわらず、ケーブル・サービス・プロバイダからの入力テレビジョン信号をビデオ・プロセッサ210によって復調させ、処理させ、テレビジョン受信機105に送信させる。ビデオ・プロセッサ210は、入力テレビジョン信号をケーブル・サービス・プロバイダから受信し、ユーザ選択チャンネルに合わせ、選択RF信号を、テレビジョン受信機105上で表示するのに適切なベースバンド・テレビジョン信号（例えば、スーパー・ビデオ信号）に変換する無線周波数（RF）フロントエンド回路を備える。ビデオ・プロセッサ210も、再生モード中に（ビデオ装置260のビデオ・バッファ256においてバッファリングした後）通常のNTSC信号をMPEG2復号器/NTSC符号器240から受信し、ベースバンド・テレビジョン信号をテレビジョン受信機105に送信することができる。

【0019】

記録モードでは、VRコントローラ250は、入力テレビジョン信号をハード・ディスク・ドライブ230上に記憶させる。VRコントローラ250の制御下で、MPEG2符号器220は、入力アナログ・テレビジョン信号をケーブル・サービス・プロバイダから受信し、ハード・ディスク・ドライブ230に記憶するよう、受信RF信号をMPEG2形式に変換する。あるいは、ビデオ再生装置150が、MPEG2データを送信するソースに結合される場合、入力MPEG2データはMPEG2符号器220をバイパスし、ハード・ディスク・ドライブ230上に直接、記憶してよい。

【0020】

再生モードでは、VRコントローラ250は、記憶されているテレビジョン信号（すなわち、番組）をMPEG2復号器/NTSC符号器240にストリーミングするようハード・ディスク・ドライブ230に指示し、MPEG2復号器/NTSC符号器はハード・ディスク・ドライブ230からのMPEG2データを、例えば、ビデオ・プロセッサ210がテレビジョン受信機105に送信するスーパー・ビデオ（S-ビデオ）信号に変換する。

【0021】

なお、MPEG2符号器220と、MPEG2復号器/NTSC符号器240とについてMPEG2標準を選択していることは、図示の目的に過ぎない。本発明の別の実施例では、MPEG符号器及び復号器は、MPEG-1標準と、MPEG-2標準と、MPEG-4標準とのうちの1つ又は複数のものや、1つ又は複数の他の種類の標準に準拠し得る。

【0022】

本明細書及び特許請求の範囲の目的で、ハード・ディスク・ドライブ230は、読み取り/書き込みデジタル多用途ディスク（DVD-RW）、書き換え可能CD-ROM、VCRテープなど用の通常の磁気ディスク・ドライブ及び光ディスク・ドライブを有するが、それらに限定されない、読み取り可能でありかつ書き込み可能である何れかの大容量記憶装置を備えるものとして定義している。実際に、ハード・ディスク・ドライブ230は、ビデオ再生装置150に永久に組み込まれているという通常の意味合いで固定されていなくてよい。むしろ、ハード・ディスク・ドライブ230は、記録されたビデオ番組を記憶する目的でビデオ再生装置150専用の何れかの大容量記憶装置を備える。よって、ハード・ディスク・ドライブ230は、いくつかの読み取り/書き込みDVD又は書き換え可能CD-ROMを保持するジューク・ボックス装置（図示せず）などの、付属型周辺ドライブ又は（組み込み型若しくは付属型の）取り外し可能ディスク・ドライブを備え得る。図2の略図に示すように、この種の取り外し可能ディスク・ドライブは、書き換え可能CD-ROMディスク235を受け入れ、読み取るこ

10

20

30

40

50

とが可能である。

【0023】

更に、本発明の効果的実施例では、ハード・ディスク・ドライブ230は、例えば、ユーザのホーム・パーソナル・コンピュータ(PC)内のディスク・ドライブ又はユーザのインターネット・サービス・プロバイダ(ISP)にあるサーバ上のディスク・ドライブを有する、ビデオ再生装置150がネットワーク接続(例えば、インターネット・プロトコル(IP)接続)を介してアクセス及び制御を行い得る外部の大容量記憶装置を有し得る。

【0024】

VRコントローラ250は、ビデオ・プロセッサ210によって受信されるビデオ信号に関する情報をビデオ・プロセッサ210から取得する。ビデオ再生装置150がビデオ番組を受信するということをVRコントローラ250が判定する場合、VRコントローラ250は、ビデオ番組が、記録するよう選択されたものであるかを判定する。ビデオ番組が記録する対象である場合、VRコントローラ250は、ビデオ番組をハード・ディスク・ドライブ230上に前述の方法で記憶させる。ビデオ番組が記録する対象でない場合、VRコントローラ250は、前述の方法で、ビデオ番組を、ビデオ・プロセッサ210によって処理させ、テレビジョン受信機105に送信させる。

【0025】

本発明の例示的实施例では、メモリ280は、ランダム・アクセス・メモリ(RAM)又は、ランダム・アクセス・メモリ(RAM)と読み取り専用メモリ(ROM)との組み合わせを備え得る。メモリ280は、フラッシュ・メモリなどの非揮発性ランダム・アクセス・メモリ(RAM)を備え得る。テレビジョン受信機105の別の効果的実施例では、メモリ280は、ハード・ディスク・ドライブ(図示せず)などの大容量記憶データ装置を備え得る。メモリ280は、読み取り/書き込みDVD又は書き換え可能CD-ROMを読み取る付属型周辺ドライブ又は(組み込み型若しくは付属型の)取り外し可能なディスク・ドライブも有し得る。図2に略図で示したように、この種の取り外し可能ディスク・ドライブは、書き換え可能CD-ROMディスク285を受け入れ、読み取ることが可能である。

【0026】

図3は、本発明の音楽ビデオ要約コンピュータ・ソフトウェア300を備える特定のメモリ部分を示す。メモリ280は、オペレーティング・システム・インタフェース・プログラム310、音楽ビデオ・セグメント化アプリケーション320、音楽ビデオ識別アプリケーション330、音楽ビデオ要約アプリケーション340、音楽ビデオ要約ブロック350及び音楽ビデオ要約ファイル360を備える。音楽ビデオ要約コントローラ270と音楽ビデオ要約コンピュータ・ソフトウェア300は併せて、本発明を行うことができる音楽ビデオ要約制御システムを備える。オペレーティング・システム・インタフェース・プログラム310は、音楽ビデオ要約コンピュータ・ソフトウェア300の動作と、VRコントローラ250及び音楽ビデオ要約コントローラ270のオペレーティング・システムとの関係をとる。

【0027】

図4は、音楽ビデオ要約ブロック群350を本発明の効果的な実施例の一部として示す。本発明の音楽ビデオ要約コントローラ270は、音楽ビデオに関して取得する情報を音楽ビデオ要約ブロック(例えば、音楽ビデオ要約ブロック410)に記憶する。図4に示すように、音楽ビデオ要約ブロック群350はN個の音楽ビデオ要約ブロック(410, 470, ... 480)を備え、そのときNは整数である。図4に示す例示的な音楽ビデオ要約ブロック410は、各音楽ビデオ要約ブロックが備え得る種類の情報を示す。例示的な音楽ビデオ要約ブロック410は、タイトルブロック420と、アルバム・ブロック430と、アーティスト・ブロック440と、レコーディング・スタジオ450・ブロックと、リリース日付ブロック460との各々を備える。これらの分類は例示的なものであり、網羅的なものでない。すなわち、他の種類の情報(図示せず)が本発明の音楽ビデオ要約ブロックに記憶される場合もある。

【0028】

音楽ビデオを備えるマルチメディア・データ・ストリームを音楽ビデオ要約コントローラ270が受信することとする。以下に更に十分に説明するように、音楽ビデオ要約コント

10

20

30

40

50

ローラ270は、(1)マルチメディア・データ・ストリームにおける音楽ビデオをセグメント化し、残りのマルチメディア・データ・ストリームからそれらを分離する工程と、(2)セグメント化音楽ビデオの各々を識別し、各音楽ビデオの主題である楽曲に関する情報を得る工程と、(3)テキスト・セグメントと、オーディオ・セグメントと、ビデオ・セグメントとを備える音楽ビデオ毎に音楽ビデオ要約ファイルを作成する工程と、(4)音楽ビデオ要約ファイルを記憶する工程と、(5)ユーザ要求に応じて、音楽ビデオ要約ファイルをユーザによって選択される順序で表示する工程とを行うことができる。

【0029】

一実施例では、音楽ビデオ要約コントローラ270は、各音楽ビデオの始めと終わりとを見つけることによってマルチメディア・データ・ストリームにおける音楽ビデオをセグメント化する。本発明の一特徴によれば、音楽ビデオは、顔の存在若しくは顔の識別などの1つ若しくは複数の画像特徴又は、新しい楽曲の始めを通常、示唆する、非音楽成分から音楽成分への、オーディオ成分における変化を検出するオーディオ分類手法などの1つ若しくは複数のオーディオ特徴を用いてセグメント化される。別の変形では、セグメント化処理は、スーパー・ヒストグラム（又は色クラスタ化手法）を用いて、新たな楽曲の始めを示唆する場合もある、暗い画像から明るい画像への変化などの、色における変化を検出する。

【0030】

更に別の変形では、音楽ビデオ要約コントローラ270は、音楽ビデオ・セグメント化アプリケーション320におけるコンピュータ命令を実行して、音楽ビデオの始めと終わりでビデオテキスト・ブロック180をサーチする。2つのビデオテキスト・ブロック180が同一の場合、それらの間のビデオ部分は、2つのビデオテキスト・ブロック180によって識別される音楽ビデオを表す。音楽ビデオがテレビジョン受信機105の画面110上に表示される場合、音楽ビデオの始めは通常、テキスト・キャプション180をビデオ画像の下部に表示する。テキスト・キャプション180は通常、楽曲のタイトル、アルバムの名前、アーティスト又はグループの名前、リリース日付や他の同様な情報を備える。テキスト・キャプション180は通常、音楽ビデオの終わりでも表示される。テキスト・キャプション180は、ビデオテキスト・ブロック180としても呼ばれる。

【0031】

音楽ビデオ要約コントローラ270が新たな音楽ビデオをセグメント化する場合、音楽ビデオ要約コントローラ270は、音楽ビデオ識別アプリケーション330におけるコンピュータ命令を実行して、例えば音楽ビデオを識別する情報をビデオテキスト・ブロック180から抽出する。音楽ビデオ要約コントローラ270は、Lalitha Agnihotri, Nevenka Dimitrova及びHerman Elenbassによる、「Video Stream Classifiable Symbol Isolation Method and System」と題する西暦1999年11月17日付出願の米国特許出願公開第09/441,943号に記載されている類の方法を用いて取得し得る。

【0032】

音楽ビデオ要約コントローラ270は、楽曲、アルバム、アーティスト又はレコーディング会社の包括的なリストを見つけて、音楽ビデオ要約コントローラ270がビデオテキスト・ブロック180から取得する情報と比較するよう、メモリ280におけるデータベース（図示せず）をアクセス（するか、インターネット上にあるデータベースをアクセス）してよい。別個の音楽ビデオ毎の音楽ビデオ情報は、別個の音楽ビデオ要約ブロック（例えば、音楽ビデオ要約ブロック410）に記憶される。

【0033】

一部の 경우에는、音楽ビデオ要約コントローラ270は、何れのビデオテキスト・ブロック180を位置特定することも識別することもできない場合がある。そのような場合には、音楽ビデオ要約コントローラ270は、楽曲の数行のトランスクリプトを歌詞のトランスクリプトのデータベースと比較してテキストの一致を見つけ得る。音楽ビデオ要約コントローラ270は、楽曲の数行のテキストを表す「サーチ列」を選択する。一実施例では、「サーチ列」テキストは、クローズド・キャプション復号器275から取得し得る。音楽ビデオ

10

20

30

40

50

要約コントローラ270は更に、メモリ280における歌詞のデータベース（図示せず）をアクセス（するか、www.lyrics.comなどのインターネット上にある歌詞のデータベースをアクセス）して、包括的な歌詞リストを見つける。音楽ビデオ要約コントローラ270は更に、「サーチ列」テキストと、歌詞のデータベースのトランスクリプトと比較して楽曲の識別情報を見つける。楽曲の識別情報が判定された後は、アーティストの名前や他の情報をデータベースから容易にアクセスすることが可能である。「サーチ列」テキストを歌詞のデータベースと比較することによって音楽ビデオ要約コントローラ270が音楽ビデオ情報をサーチし、位置特定する方法は、以下に、図7を参照しながら更に十分に説明することとする。

【0034】

前述のように、音楽ビデオ要約コントローラ270は音楽ビデオ情報を取得し、音楽情報を音楽ビデオ要約ブロック350に記憶する。次に、音楽ビデオ要約ブロック（例えば、音楽ビデオ要約ブロック410）毎に、音楽ビデオ要約コントローラ270は歌詞をアクセスし、楽曲の「コーラス」を歌詞から識別する。楽曲のコーラスは通常、歌詞のデータベースにおけるコーラスとして識別される。あるいは、数回繰り返される歌詞部分を楽曲のコーラスとしての役目を担うものとして選択する場合もある。このことは、クローズド・キャプション復号器275を用いるか、オーディオ・トラック部分を比較して同様なオーディオ・パターンを見つけることによって達成し得る。本発明の別の特徴によれば、音楽ビデオのコーラス部分は、別個のデータベースをアクセスすることを必要とすることなく、コーラスを示唆する場合が多い、句の繰り返しに関連したトランスクリプトを解析することによって明らかにされる。トランスクリプトは、例えば、クローズド・キャプション情報から取得し得る。

【0035】

楽曲の「コーラス」は、楽曲の特性を、大半のリスナに対して楽曲の最初の数行が明らかにする以上に明らかにする。音楽ビデオ要約コントローラ270はそして、歌詞のトランスクリプトにおけるコーラスを、コーラスに相当する、マルチメディア・ファイルのオーディオ及びビデオの部分と突合することが可能である。音楽ビデオ要約コントローラ270は更に、音楽ビデオ要約ファイル360におけるコーラスに相当する、マルチメディア・ファイルのオーディオ及びビデオの部分の複製を音楽ビデオ要約ファイル360に入れる。

【0036】

音楽ビデオ要約コントローラ270は音楽ビデオ毎の各音楽ビデオ要約ファイルをメモリ280に記憶する。ユーザ要求の受信に応じて、音楽ビデオ要約コントローラ270は、特定の音楽ビデオ要約ファイル360をアクセスし、テレビジョン受信機105を介して（オーディオ及びビデオの部分有する）音楽ビデオ要約ファイル360を再生することができる。あるいは、音楽ビデオ要約コントローラ270は、記憶されている音楽ビデオ要約ファイル360の全てのリスト190をアクセスし、リスト190をテレビジョン受信機105の画面110上に表示することができる。すなわち、リスト190は、(1)マルチメディア・データ・ストリームにおいて検出された音楽ビデオの全ての音楽ビデオ要約ファイルと、(2)各音楽ビデオをレコーディングしたアーティスト又はグループの識別情報とを表示する。リスト190は任意的には、ユーザ選好によって提示して、リストに提示される情報のコンテンツを個人化し得る。リモコン装置125及びIRセンサ160を用いて、ユーザは「音楽ビデオ要約再生」制御信号を音楽ビデオ要約コントローラ270に送って、リスト190において次に再生する音楽ビデオ要約ファイルを選択する。このようにして、ユーザは、音楽ビデオ要約ファイルが再生される順序を選択する。

【0037】

図5は、音楽ビデオの索引付け及び要約を行う、本発明によって用いられる手法の概要を備える流れ図500である。図5に示すように、音楽ビデオ要約コントローラ270はまず、工程510中に、音楽ビデオ507を備える受信マルチメディア・ストリーム505をオーディオ成分、ビデオ成分及びトランスクリプト成分に分離する。音楽ビデオ要約コントローラ270は、工程520中に、以下に更に説明するいくつかの特徴をオーディオ成分、ビデオ成分及

10

20

30

40

50

びトランスクリプト成分から抽出する。トランスクリプトは、例えば、クローズド・キャプション情報から取得され、ソフトウェアによってテキスト行毎にタイム・スタンプが挿入される場合がある。この時点で、全ての特徴が、楽曲の境界を何ら示さずに、タイム・スタンプされたデータ・ストリームを備える。

【0038】

楽曲の当初の境界は、工程530中に、図10及び図11に関して以下に更に説明する方法で、視覚特徴、聴覚特徴及びテキスト特徴を用いて判定される。その後、当初の境界とトランスクリプト情報とを用いて、コーラス位置とコーラスのキーフレーズが、工程540中に、図11に関して以下に説明するように判定される。工程545及び工程550中には、コーラス情報に基づいて、ウェブ・サイトからの情報を用いて楽曲のタイトル、アーティスト名、ジャンル及び歌詞が判定される。

10

【0039】

次に工程560中に、楽曲の境界は、例えば、取得された歌詞と、オーディオ分類と、(色情報に基づいた)視覚的なシーン境界と、オーバーレイされたテキストのうちの1つ又は複数を用いて確認される。本発明は、ウェブ・サイト上の歌詞とトランスクリプト中の歌詞とが常に完全に一致する訳ではないということを考慮に入れている。歌詞に基づいて、歌詞の境界が当初の境界の情報と歌詞とを用いて合わせられる。あるいは、トランスクリプト情報が入手可能でない場合、タイトルのページを、抽出ビデオテキストに対する光学式文字認識(OCR)手法を用いて解析して、アーティスト名、楽曲のタイトル、年情報及びレコード・レーベル情報などのビデオ情報を見つけることが可能であり、ウェブ情報を用いてOCR工程からの出力を検証することが可能である。この情報によって、楽曲の歌詞をウェブ・サイトから得ることが可能であり、テキスト情報を用いてコーラス検出手法を行うことが可能である(この場合の問題は、ダウンロードされるこれらの歌詞がタイム・スタンプされておらず、アラインメントの問題が存在するということである。)。好ましくは、トランスクリプトは、音声・テキストオーディオ解析を用いて得られる。一変形では、ダウンロードされるトランスクリプトと、音声・テキスト生成器によって生成されるトランスクリプトとを集約してより正確なトランスクリプトを得ることが可能である。

20

【0040】

楽曲毎の境界と、視聴覚特徴とを得れば、楽曲は次に、工程565と工程570との各々の間に、以下に図14に関して説明したように最善の代表的なフレームと、楽曲の要約に最善のビデオ・クリップとを判定することによって要約される。最善の代表的なフレームは、アーティストからのクローズアップ、楽曲情報を備えているタイトル画像、アーティスト、レーベル、アルバム及び年を備える。楽曲の要約は、工程575中に、楽曲要約ライブラリに記憶される。ユーザは、工程580中に、番組の要約を、例えば、ウェブ・ベースの音楽ビデオ取り出しアプリケーションを用いてアクセスすることが可能である。

30

【0041】

本発明による音楽ビデオ要約は、個々の楽曲の識別及び要約に基づいている。番組レベルでは、要約は楽曲のリストを備える。次のレベルでは、各楽曲は、楽曲を表すタイトル、アーティストや特定のマルチメディア構成要素を備える。

【0042】

境界の検出

音楽ビデオ要約は、2つの種類の境界の検出を有する。第1に、楽曲の境界を自動的に検出することを要する。その後、コーラスの境界を検出することを要する。図5に関して前述したように、本発明は、境界の検出を視覚特徴、オーディオ特徴及びトランスクリプト特徴を用いて行う。視覚特徴は、ビデオテキストの存在と、顔の検出(及び/又は識別)と、急なカットと、色ヒストグラムとを備える。

40

【0043】

ビデオテキストの存在を用いた、境界の検出

ビデオテキストの存在を用いて境界の検出を行うのに適切な手法を詳細に記載しているものがある(例えば、内容を本明細書及び特許請求の範囲に援用する、N. Dimitrova他

50

による「MPEG-7 VideoText Description Scheme for Superimposed Text」 Int'l Signal Processing and Image Communications Journal (Sept., 2000)又は、「System and Method for Indexing and Summarizing Music Videos」と題する西暦2002年6月20日付願の米国特許出願公報第10/176,239号(代理人管理番号US020206)を参照。)。

【0044】

ビデオテキストの検出は境界を検出するうえで信頼性の高い方法を備えるが、それは、読み取り及び認識を容易にする方法で、アーティストやタイトルなどのビデオテキスト情報が各音楽ビデオの始めと終わりにて提示されるからである。よって、楽曲の始めにビデオテキストが存在することは、楽曲の間の境界を描写することに寄与する。ビデオテキスト検出性能は、例えば、テキスト・ボックスが楽曲のタイトル情報を備えているようにするか、画面の左下部分などの特定の位置にテキスト・ボックスがあるようにすることによって向上させ得る。楽曲が既に始まったということの1つの標識として楽曲のタイトル・ページを用いて、楽曲の始めを判定することが可能である。

10

【0045】

顔検出(又は顔識別)を用いた、境界の検出

本発明の一特徴によれば、考えられる、楽曲の境界は、画像フレームにおける顔の検出に基づいて識別し得る。図6は、本発明の特徴を組み入れた例示的な顔特徴解析処理600の流れ図である。図6に示すように、顔特徴解析処理600はまず、工程610中に、考えられるいくつかの顔種類ラベルのうちの1つを各画像フレームに割り当てる。例えば、顔特徴解析処理600は、フレームが主に、肩のショット(S)、体全体のショット(F)、顔のクローズアップ(C)又は複数の個人(M)を有するか否かに基づいてラベルを各フレームに割り当て得る。割り当てられる顔種類ラベルの例示的なタイムラインのイメージ図を以下に説明する図12に備える。画像フレームは次に、工程620中に、割り当てられる顔種類ラベルに基づいてクラスタ化される。最後に、工程630中に、パターンが顔種類ラベルのクラスタにおいて解析されてビデオ境界を検出する。プログラム制御は次に終結する。工程630中に行われるパターン解析は、以下に、図10及び図11に関して更に説明する。

20

【0046】

このようにして、経時的に、顔特徴解析処理600は、(フレームが同じビデオの一部であるということを示唆する)同質の画像シーケンス・パターンを探すことになる。そのようなパターンからの逸脱は、新たなビデオ・マテリアル又は非ビデオ・マテリアルが始まったということを示唆するものである。顔検出及び顔ラベリングを行ううえで適切な手法を詳細に記載するものがある(例えば、その内容を本明細書及び特許請求の範囲に援用する、N. Dimitrova他による「Video Classification Using Object Tracking, International Journal of Image and Graphics」 Special issue on Image and Video Databases, Vol. 1, No. 3, (Aug. 2001)参照。)。

30

【0047】

主演アーティストを見つけるうえで顔は極めて重要である一方、音楽ビデオは、ビデオ顔検出を行うことが難しいジャンルである。顔の存在は、例えば、特殊効果や、種々の色による照明が理由で、ビデオにおいて適切に検出されない場合がある。更に、例えば、出演者が踊っているか眠っている場合には、顔は多くの場合、対角線の位置又は水平方向の位置にある。

40

【0048】

別の変形では、任意的には、顔識別も行って、各フレームにおいて識別されるアーティストに基づいて識別ラベルを周知の方法で割り当てることが可能である。新たなアーティストが画像シーケンスにおいて現れることは、新たなビデオの始まりを示唆している。顔識別の性能は任意的には、人気アーティスト又は期待アーティストの顔画像を備えるデータベースを用いることによって改善することが可能である。

【0049】

急なカット(カメラ・チェンジ)を用いた、境界の検出

50

本発明の一特徴によれば、考えられる、楽曲の境界は、画像シーケンスにおけるカメラ・チェンジの 패턴の検出に基づいて識別することが可能である。図7は、本発明の特徴を組み入れた例示的なカメラ・チェンジ解析処理700の流れ図である。図7に示すように、カメラ・チェンジ解析処理700はまず、工程710で、ビデオ・シーケンスにおけるカメラ・カットの頻度を判定する。カメラ・カットの頻度を判定するのに適切な手法を詳細に記載しているものがある（例えば、その内容を本明細書及び特許請求の範囲に援用する「Significant Scene Detection and Frame Filtering for a Visual Indexing System」と題する、米国特許第6137544号明細書参照。）。

【0050】

その後、工程730中に、カメラ・チェンジ解析処理700は、カメラ・カット頻度データにおけるパターンを解析してビデオ境界を検出する。工程730中に行われるパターン解析は更に以下に図10及び図11に関して説明する。なお、カット・チェンジは、音楽ビデオにおいては非常に頻繁に起こる。実際に、平均カット距離は楽曲の間よりもコマーシャル・ブレイクの間の方が高いということを示している。これは極めて珍しいことであるが、それは、大半の他のジャンルの場合、コマーシャル・ブレイクが番組よりも低いカット距離を表すからである。別の変形では、更なるカメラ・チェンジ・ラベルを備えて、パン、チルト及びズームなどのカメラの動きの種類を特徴付けることが可能である。

【0051】

色ヒストグラムを用いた、境界の検出

本発明の別の特徴によれば、考えられる、楽曲の境界は、色変化特徴に基づいて識別し得る。例示的实施例ではスーパー・ヒストグラム手法を用いて、同様な色を表すフレーム・ファミリーを推論している。図8は、本発明の特徴を組み入れた例示的な色ヒストグラム解析処理800の流れ図である。図8に示すように、色ヒストグラム解析処理800はまず、工程810中に、画像フレーム毎の色ヒストグラムを取得する。一般的に、色ヒストグラムは、相当するフレームの色成分を特徴付ける署名としてみなし得る。画像フレームは更に、工程820中に、（図12に示すような）ヒストグラムに基づいてクラスタ化される。最後に、パターンを、工程830中に、ヒストグラムのクラスタにおいて解析してビデオ境界を検出する。プログラム制御はそして、終結する。工程830中に行われるパターン解析は、図10及び図11に関して更に以下に説明する。クラスタ化段階中に検討される、画像フレームの履歴は、例えば、一分に限定される場合があるが、それは、同様な色を備えている何れかの先行フレームは適切でない場合があるからである。

【0052】

このようにして、経時的に、色ヒストグラム解析処理800は、（フレームが同じビデオの一部であるということを示唆する）同質の画像シーケンス・パターンを探すことになる。そのようなパターンから逸脱していることは、新たなビデオ・マテリアル又は非ビデオ・マテリアルが始まったということを示唆するものである。例えば、特定の楽曲は、フィルミングのスタイルによって、ビデオにわたってドミナント・カラー配色を有し得る。更に、各楽曲間のコマーシャル・ブレイクは通常、別々のドミナント・カラー配色を表すものである。色ヒストグラムによって、同様な色を表すフレーム・ファミリーを識別することが可能になる。一般的に、新たな楽曲が登場するにつれ、色パレットが変化し、新たな楽曲のフレームが新たなファミリーにクラスタ化される。よって、色ヒストグラム手法は、音楽ビデオの、潜在的な始まり及び終わりを検出するうえで有用である。

【0053】

色ヒストグラムを更に詳細に記載するものがある（例えば、その各々の内容を本明細書及び特許請求の範囲に援用する、L. Agnihotri及びN. Dimitrovaによる「Video Clustering Using Superhistograms in Large Video Archives」 Visual 2000, Lyons, France (November, 2000)や、N. Dimitrova他による、「Superhistograms for Video Representation」 IEEE ICIP, 1999, Kobe, Japanを参照。）。

【0054】

10

20

30

40

50

オーディオ特徴を用いた、境界の検出

本発明の別の特徴によれば、潜在的な楽曲の境界をオーディオ特徴に基づいて識別することが可能である。図9は、本発明の特徴を組み入れた、例示的なオーディオ特徴解析処理900の流れ図である。図9に示すように、オーディオ特徴解析処理900はまず、工程910中に、いくつかの考えられるオーディオ種類ラベルの1つを各オーディオ・フレームに割り当てる。なお、オーディオ・フレームの持続時間は、画像フレームの持続時間とは異なり得る。例えば、オーディオ特徴解析処理900は、ラベルを各オーディオ・フレームに、オーディオ・フレームが主に1)音楽、2)音声、3)背景音楽を備えている音声、4)複数の個人の話し声、5)雑音、6)雑音を備えている音声、7)無音、8)音量の増加や9)音量の減少を備えているか否かに基づいて割り当て得る。オーディオ・フレームは次に、工程920中に、割り当てられるオーディオ種類ラベルに基づいてクラスタ化される。最後に、工程930中に、オーディオ種類ラベルのクラスタにおけるパターンを解析してビデオ境界を検出する。そして、プログラム制御は終結する。工程930中に行われるパターン解析は、図10及び図11に関して以下に更に説明する。例えば、楽曲の始めを示す、楽曲の始めと終わりでの無音、又は音量の上昇を探してもよく、楽曲の終わりを示す、音量の減少を探してもよい。

10

【0055】

このようにして、経時的に、オーディオ特徴解析処理900は、(オーディオ特徴が同じビデオの一部であるということを示唆する)同質のオーディオ・シーケンス・パターンを探ることになる。そのようなパターンから逸脱していることは、新たなビデオ・マテリアル又は非ビデオ・マテリアルが始まったということを示唆するものである。オーディオのセグメント化及び分類を行ううえで適切な手法を詳細に記載しているものがある(例えば、内容を本明細書及び特許請求の範囲に援用する、D. Li他による「Classification of General Audio Data for Content-Based Retrieval」 Pattern Recognition Letters 2000 (2000)を参照。)

20

【0056】

トランスクリプト特徴を用いた、境界の検出

本発明の別の特徴によれば、潜在的な、楽曲の境界は、例えば、クローズド・キャプション情報から取得し得るオーディオ・トランスクリプトに基づいて明らかにし得る。一般的に、段落をテキスト・トランスクリプトにおいて、キーワード解析(又は自己相関解析)を用いて識別する。特に、語のヒストグラムを取得し、解析して、新たな楽曲を検出する。新たなキーワード群を識別することは、新たなビデオ・マテリアル又は非ビデオ・マテリアルが始まったということを示唆するものである。トランスクリプトの「段落化(paragraphing)」を行ううえで適切な手法を詳細に記載しているものがある(例えば、内容を本明細書及び特許請求の範囲に援用する、N. Stokes他による「Segmenting Broadcast News Streams Using Lexical Chains」 Proc. of Starting Artificial Intelligence Researchers Symposium (STAIRS) (2002)参照。)

30

【0057】

低レベル特徴を用いた、境界の検出

上記特徴に加えて、本発明は、各画像フレームにおけるいくつかのエッジ又は形状や、いくつかの局所的な動き又は大局的な動きなどの、コンテンツから直接導き出されるいくつかの低レベル特徴を直接使い、何れかの、パターン及びそのようなパターンからの逸脱をこれらの低レベル特徴において監視することも可能である。更に、メル周波数ケプストラム係数(MFCC)、線形予測係数(LPC)、ピッチ変動、帯域幅、音量や音色などの低レベルのオーディオ特徴も解析することが可能である。

40

【0058】

視覚特徴、オーディオ特徴及びトランスクリプト特徴の解析

前述のように、本発明は、図5乃至図9に関して上記で説明した、視覚特徴、オーディオ特徴及びトランスクリプト特徴を用いた、境界の検出を行う。図10に示す1つの例示的実施例では、視覚特徴、オーディオ特徴及びトランスクリプト特徴を、ベイジアン確率ネッ

50

トワーク (BBN) 1000などのパターン認識エンジンを用いて監視してビデオ・ストリームを個々のビデオにセグメント化する。図11に示す別の実施例では、視覚特徴、オーディオ特徴及びトランスクリプト特徴を、ルール・ベースの発見的処理1100を用いて処理してビデオ・ストリームを個々のビデオにセグメント化する。一般的に、何れの例示的实施例も、上記の種々の特徴全てからの近似の境界を用いてビデオをセグメント化する。

【0059】

図10は、本発明の特徴を組み入れた例示的なベイジアン確率ネットワーク (BBN) 1000を示す。BBN1000は、視覚特徴、オーディオ特徴及びトランスクリプト特徴を監視してビデオ・ストリームを個々のビデオにセグメント化する。一般的に、ベイジアン確率ネットワークは複雑なパターンを認識し、予め規定された活動を学習し、認識するのに用いられている。ベイジアン確率ネットワーク1000は、セグメント化情報によって既にラベリングされたビデオ・シーケンスを用いて公知の方法で訓練される。

10

【0060】

図10に示すように、ベイジアン確率ネットワーク1000は、本発明によって監視される別々の特徴によって各々が関連付けられる、複数の状態1010-1乃至1010-Nを有する第1層1010を備える。各状態に対する入力は、特定のウィンドウにわたる平均特徴値である。例えば、顔存在特徴の場合、入力は、例えば、先行する20秒のウィンドウにわたる各画像における顔の数の変動が、現行の20秒のウィンドウと比較して存在するか否かであり得る。同様に、色ヒストグラム特徴の場合、入力は、例えば、新たなクラスタが現行のウィンドウにおいて検出されたか否かであり得る。

20

【0061】

ベイジアン確率ネットワーク1000は、ビデオの始め又は終わりに関連した遷移 P_{trans} に現行の時間ウィンドウが相当するという確率を相当する状態毎に、状態と関連した単一の特徴に基づいて判定する第2層1020を備える。例えば、確率 $P_{facechg}$ は顔変動特徴データによって示唆されるような顔変動の確率を示す。最後のレベル1030では、ベイジアン確率ネットワーク1000は、ベイズ推論を用いて、楽曲のブレークが存在したか否かを、監視される特徴の各々にわたる確率に基づいて判定する。更なる変形では、ニューラル・ネットワーク又は自己回帰移動平均 (ARMA) 法を用いて楽曲の境界を予測し得る。

【0062】

現行の時間ウィンドウが状態1030でのセグメントに相当するか否かを判定する条件付確率は：

30

【0063】

【数1】

$$P(x|e^t) = \sum_{\phi_{mn}} P(x|\phi_{mn}) \prod_{i=1}^{|\phi|} P(\phi_i|e)$$

として計算し得る。

40

【0064】

上記式は、条件付確率を計算する一般的なケースを表す。図10に表すモデルの場合、確率は

【0065】

【数 2】

$$P(x_i | e^{\theta}) = \sum_{\theta_{mn}} P(x_i, v, f, a, c, t, a)$$

$$= \sum_{\theta_{mn}} P(x_i | v)P(v)p(x_i | f)P(f)p(x_i | a)P(a)p(x_i | c)P(c)p(x_i | t)P(t)p(x_i | a)P(a)$$

として計算し得る。このとき、vはビデオテキストであり、aは急なカットであり、cは色であり、tはトランスクリプトであり、aはオーディオに関する解析である。 10

【0066】

図11は、ビデオ・セグメント化処理1100の例示的实施方法を表す流れ図である。前述のように、ビデオ・セグメント化処理1100は、視覚特徴、オーディオ特徴及びトランスクリプト特徴をルール・ベースの発見的手法を用いて処理してビデオ・ストリームを個々のビデオにセグメント化する。図11に示すように、ビデオ・セグメント化処理1100はまず、工程1110中に、監視されたビデオ特徴値、オーディオ特徴値、及びトランスクリプト特徴値を評価する。その後、ビデオ・セグメント化処理1100は、工程1120中に、1つ又は複数の予め規定されたビデオ・セグメント化ルール1115を特徴値に適用する。例えば、特定のアプリケーションは、ビデオテキスト存在に対する確率値と色ヒストグラム特徴に対する確率値が両方とも所定の閾値を超える場合にビデオ・セグメントを識別することとするということの規定するビデオ・セグメント化ルールを規定し得る。更なる例では、ビデオ・セグメント化ルールは、ビデオテキスト存在に対する確率値と少なくともN個の他の監視される特徴に対する確率値が所定の閾値を超える場合にビデオ・セグメントを識別することとするということの規定し得る。 20

【0067】

工程1130中には、新たなビデオが検出されたかを判定する検査が行われる。工程1130中に、新たなビデオは検出されなかったということが判定される場合、プログラム制御は工程1110に戻って、上記の方法で画像ストリームを監視し続ける。しかし、工程1130中に、新たなビデオが検出されたということが判定される場合、新たなビデオ・セグメントが工程1140で更に識別される。そして、プログラム制御は、適宜、終結してもよく、工程1110に戻って、上記の方法で、画像ストリームを監視し続けてもよい。 30

【0068】

ベイジアン確率ネットワーク1000又はビデオ・セグメント化処理1100による、監視される特徴の処理は、トランスクリプトが、視覚ストリームやオーディオ・ストリームよりも後から始まるということ考慮に入れ得る。視覚的な観点から、楽曲の始めから数秒後に通常現れるビデオテキストのタイトル・ページも取得する。始めの境界は、楽曲の視覚的な色境界と、オーディオ領域における音楽分類の始めと合わせられる。

【0069】

図12は、割り当てられた顔種類ラベル1210、色ヒストグラム・クラスタ1220及びビデオテキスト存在1230のタイムラインのイメージ図を備える。図12に示すように、監視される特徴の各々の特徴データは、ビデオ・セグメントを検出するために合わせられる。本発明は、ベイジアン確率ネットワーク1000又はビデオ・セグメント化処理1100を用いて、2つのビデオ間又はビデオ・マテリアルと非ビデオ・マテリアルとの間の遷移1240を、個々の特徴各々が示唆する遷移期間に基づいて識別する。 40

【0070】

コーラスの検出

楽曲のコーラスを判定するために、従来の研究は音楽のオーディオ特徴を中心においてきた。楽曲におけるセグメントの繰り返しを見つけるうえでの共通的な手法には、自己相関解析を行うということがある。コーラスは、ポピュラー楽曲では少なくとも2回繰り返 50

される。通常、大半の楽曲では3回以上繰り返される。

【0071】

本発明の更なる特徴によれば、楽曲のコーラスは、トランスクリプト（クローズド・キャプション情報）を用いて検出される。一般的に、コーラスは、誤の繰り返しを有する楽曲部分を検出することによって識別される。なお、クローズド・キャプションは完全でなく、例えば、タイプミス又は欠落を有し得る。図13は、コーラス検出処理1300の例示的な実施方法の流れ図である。図13に示し、以下に記載するように、コーラス検出処理1300は、工程1310中にクローズド・キャプションに対してキーフレーズの検出を行い、工程1320中に、潜在的なコーラスの検出と、後解析とを行い、工程1330中にコーラス候補の確認を行い、工程1340中に不規則なコーラスの検出を行うことによってコーラス・セグメントを認識する。最後に、工程1350中に、自己相関解析を行って何れかのコーラスを識別する。

10

【0072】

キーフレーズの識別（工程1310）

コーラスは、最も多く繰り返される歌詞を楽曲において有する。このようなフレーズを検出し、クラスタ化することによって、コーラス・セグメントの時間的位置を識別することが可能である。コーラスを有する潜在的な部分を選択するよう、フレーズが楽曲において存在するタリー（総数）が集計される。このフレーズは、トランスクリプトから得られ、テレビジョン画面上のテキスト行全体又は、コマ若しくはピリオドなどのデリミタによって分割された行部分を表す。新たなフレーズ毎に、フレーズがタリーにおいて存在しているか否かが判定され、そのフレーズに対するカウンタが増やされる。さもなければ、新たなピンが新たなフレームについて作成され、そのピンについてカウンタが1に初期化される。この処理は、各楽曲のテキスト全体について繰り返される。楽曲の終わりでは、繰り返されるフレーズが、キーフレーズとして指定される。

20

【0073】

候補コーラスの検出（工程1320）

コーラス・セグメントの潜在的候補は、キーフレーズの2つ以上の生起を有するものである。このセグメントを見つけるために、キーフレーズの各々が生起する時点でのタイム・スタンプが識別される。キーフレーズのタイム・スタンプ毎に、潜在的なコーラスが指定される。この潜在的コーラスと別のコーラスとの間がn秒間以内である場合、それらは併合される。いくつかの楽曲の検査に基づけば、コーラスの長さが30秒(n=30)を超えることはほとんどないものとみなす。

30

コーラス候補の確認（工程1330）

2つ以上のキーフレーズを有する候補のみがコーラスとして選択される。4つ以上のコーラスが選択される場合、以下に定義する、最高のキーフレーズ密度を有する3つのコーラスが判定される。

【0074】

密度 = (コーラスにおけるキーフレーズ数) / (コーラスの持続時間)

不規則なコーラスの検出と、後解析（工程1340）

要約するには、正確に判定するのは1つのコーラスだけでよい。ユーザに提示されることになる「キーコーラス」が識別される。楽曲中では種々のコーラスの持続時間に関して大きな変異性が存在する（15秒乃至30秒はまれではない。）。この変異性によって、コーラスの位置及び長さを予測することが難しくなる。3つのコーラスのうちの中間の長さのコーラスが選択される。第1のコーラスは、第1のコーラスとともに楽曲への「リード」も得るよう、他のコーラスよりも好ましい場合がある。更に、楽曲内のコーラスの配置は変化する。最後のコーラスの解析は、他のコーラスとの距離が適度なコーラスを選択するのに用いられる。

40

【0075】

自己相関解析（工程1350）

オーディオ・コンテンツ解析では、研究者は、コーラスを見つけるために自己相関を用

50

いている（例えば、内容を本明細書及び特許請求の範囲に援用する、J. Footeによる「Visualizing Music and Audio Using Self Similarity」 Proc. ACM Multimedia '99, 77-80, Orlando, Florida (Nov., 1999)を参照。）。自己相関解析を、本発明によってトランスクリプトに対して用いて楽曲の構造を視覚化する。自己相関関数を見つけるために、トランスクリプトにおける語の全てが2次元でレイアウトされ、何れの次元上の語も同じであるか否かによってマトリックスが1と0とで充填される。このマトリックスは更に、対角線方向に描かれ、このビューにおけるピークを判定し、それはその場合、コーラスが楽曲において生起する位置を示す。

【0076】

音楽ビデオの要約

音楽ビデオの要約は、種々のメディア（オーディオ、ビデオ及びトランスクリプト）におけるビデオから導き出されるコンテンツ構成要素を備える。例示的な実施方法では、ベイジアン確率ネットワークを用いて、音楽ビデオの一般的なコンテンツ構成要素と音楽イベントの遷移とを捕捉し、作曲の構造を捕捉する。BBNは、例えば、器楽とヴァース（V）のイベント及びコーラス（C）イベントを有するものとして楽曲をモデリングするのに用い得る。特定の楽曲における音楽イベントの順序は、例えば、VCCVCCであり得る。しかし、多くの楽曲は、コーラスとヴァースとの間のブリッジ部分などのより複雑な構造を有している場合があり、多くの楽曲では、コーラスの繰り返しもない一方、楽曲全体が1つの単一の継ぎ目のないヴァースである。BBN手法によって、音楽イベントのうちの1つが欠落していても、妥当な要約がなお得られる。

【0077】

図14は、要約を構成する、ビデオからの構成要素を見つけるのに用いられる関数をモデリングするのに用い得るベイジアン確率ネットワーク1400を示す。重要なセグメントを判定する条件付確率は、

【0078】

【数3】

$$P(x | e^p) = \sum_{\varphi_{mn}} P(x | \varphi_{mn}) \prod_{i=1}^{|\varphi|} P(\varphi_i | e)$$

30

として計算することが可能である。

【0079】

上記式は、条件付確率を計算する一般的なケースを表す。図14に示すモデルの場合、確率は：

【0080】

【数4】

$$\begin{aligned} P(x_i | e^p) &= \sum_{\varphi_{mn}} P(x_i, t, c, h, m) \\ &= \sum_{\varphi_{mn}} p(x_i | t) P(t) p(x_i | c) P(c) p(x_i | h) P(h) p(x_i | m) P(m) \end{aligned}$$

40

として計算することが可能であり、そのとき、

【0081】

【数5】

$$\varphi = \{title, closeup, chorus, music\}$$

である。

50

【 0 0 8 2 】

例示的な実施例では4つのメディア構成要素が存在するので、mの値は4である。nの値は、メディア構成要素毎に、確率が呈し得る値の数によって変わってくる。例えば、P(title)は、0と1との間の値であり得るものであり、この値は、0.1のステップで、テキストが及ぶ、画像の割合によって変わってくる。よって、この場合、nは10である。ことによると、動き、オーディオ・テクスチャ、リード楽器/歌手ハイライトなどの更なる特徴を親ノードに備え得る。

【 0 0 8 3 】

選択基準は、メディア構成要素毎の要約において提示する対象のコンテンツを決定する。要約は、

【 0 0 8 4 】

【 数 6 】

$$\psi_{Visual}P(x|e^p) = \begin{cases} 1 & \text{if } P(x_i|e^p) \geq \theta_1 \\ & \& P(vtext) \geq \theta_2 \\ & \& P(face) \geq \theta_3 \\ 0 & \text{if } P(x_i|e^p) < \theta_1 \end{cases} \quad \psi_{Audio}P(x|e^p) = \begin{cases} 1 & \text{if } P(x_i|e^p) \geq \theta_1 \\ & \& P(music) \geq \theta_4 \\ 0 & \text{if } P(x_i|e^p) < \theta_1 \end{cases}$$

$$\psi_{Transcript}P(x|e^p) = \begin{cases} 1 & \text{if } P(x_i|e^p) \geq \theta_1 \\ & \& P(chorus) \geq \theta_5 \\ 0 & \text{if } P(x_i|e^p) < \theta_1 \end{cases}$$

として定義される選択関数からの出力である。

【 0 0 8 5 】

音楽ビデオの要約は、上記選択関数の全ての出力を有する集合である。

【 0 0 8 6 】

【 数 7 】

$$S = \{\psi_{Audio}P(x|e^p), \psi_{Video}P(x|e^p), \psi_{Transcript}P(x|e^p)\}$$

ビデオから導き出されるこれらの構成要素に加えて、アーティスト、タイトルやアルバムなどの高レベルの情報を加えてもよい。この高レベル情報は、例えば、インターネットから抽出して要約を完成し得る。

【 0 0 8 7 】

当然、ベイジアン確率ネットワークは、要約に重要な構成要素の選択をモデリングする1つの方法に過ぎない。H. Sundaram他による「A Utility Framework for the Automatic Generation of Audio-Visual Skims」ACM Multimedia 2002, Juan Les Pin (December 1-5, 2002)記載のスングラム(Sundaram)による利用度最大化フレームワークや、Yu-Fei Ma他による「A User Attention Model for Video Summarization」ACM Multimedia 2002, Juan Les Pin (December 1-5, 2002)記載のマ(Ma)による、要約のためのユーザ注目モデルの適用を考えてもよい。これらのモデルは、要約のための生成モデルであり、アルゴリズムの設計者が重要であると決定することをモデリングしている。非監督マシン学習手法を音楽ビデオの視覚化及び要約に適用して、固有の構造パターン及びハイライトを見つけ得る。

【 0 0 8 8 】

要約は、上記のユーザ・インタフェース及び情報の種類について個人化することが可能である。ユーザは、要約を自らが受け取りたいインタフェースの種類と、提示される要約のうちの特定のコンテンツとを選び得る。情報の多寡や情報の配置などの違いはユーザ設

10

20

30

40

50

定によって変えることが可能である。ユーザは、要約に備えることとするものも選ぶことが可能である。ユーザは、短いアンケートに記入して、見たい情報の種類を示すことが可能である。

【0089】

公知のように、本明細書及び特許請求の範囲記載の方法及び装置は、その上にコンピュータ判読可能コード手段を実施させたそのコンピュータ判読可能媒体をそれ自体が備えるその製造物品として流通させ得る。コンピュータ判読可能プログラム・コード手段は、本明細書及び特許請求の範囲記載の方法を行うか本明細書及び特許請求の範囲記載の装置を作成する工程の全部又は一部をコンピュータ・システムとともに行うよう動作可能である。コンピュータ判読可能媒体は、記録可能媒体（例えば、フロッピー（登録商標）・ディスク、ハード・ドライブ、コンパクト・ディスク又はメモリ・カード）であってよく、伝送媒体（例えば、光ファイバ、ワールドワイド・ウェブ、ケーブル、若しくは、時分割多元アクセス、符号分割多元アクセスを用いた無線チャネル、又は他の無線周波数チャネル）であってよい。コンピュータ・システムとともに用いるのに適切な情報を記憶することが可能な、何れかの公知の媒体又は開発された媒体を用い得る。コンピュータ判読可能コード手段は、磁気媒体上の磁気変動又はコンパクト・ディスクの表面上の高さ変動などの、命令及びデータをコンピュータが読み取ることが可能にする何れかの機構である。

10

【0090】

本明細書及び特許請求の範囲記載のコンピュータ・システム及びサーバは各々、本明細書及び特許請求の範囲開示の方法、工程及び機能を実施するよう、関連したプロセッサを構成することになるメモリを備える。メモリは分散型であっても局所型であってもよく、プロセッサは分散型であっても単一型であってもよい。メモリは、電気メモリとして実施しても、磁気メモリとして実施しても、光メモリとして実施してもよく、これらの種類や他の種類の記憶装置の何れかの組み合わせとして実施してもよい。更に、「メモリ」の語は、関連したプロセッサによってアクセスされるアドレス指定可能な空間におけるアドレスとの間での読み書きができる何れかの情報を有するのに十分広く解されることとする。この定義によれば、ネットワーク上の情報はメモリ内部になおあるが、それは関連したプロセッサがネットワークからこの情報を取り出すことが可能であるからである。

20

【0091】

本明細書及び特許請求の範囲に示し、説明した実施例及び変形が、本発明の原理を示しているに過ぎず、種々の修正を当業者によって本発明の範囲及び趣旨から逸脱することなく実施し得ることとする。

30

【図面の簡単な説明】

【0092】

【図1】本発明が動作し得る例示的な従来のビデオ表示システムを示す図である。

【図2】本発明の一実施例による、図1の例示的なビデオ表示システムにおける音楽ビデオの索引付け及び要約を行うシステムを示す図である。

【図3】本発明の特徴を組み入れた音楽ビデオ要約処理を備えるメモリを示す図である。

【図4】本発明の実施例とともに用いられる音楽ビデオ要約ブロックを備えるメモリを示す図である。

40

【図5】本発明の特徴を組み入れた、音楽の索引付け及び要約の処理の例示的实施方法を示す図である。

【図6】本発明の特徴を組み入れた例示的な顔特徴解析処理の流れ図を示す図である。

【図7】本発明の特徴を組み入れた例示的なカメラ・チェンジ解析処理の流れ図を示す図である。

【図8】本発明の特徴を組み入れた例示的な色ヒストグラム解析処理の流れ図を示す図である。

【図9】本発明の特徴を組み入れた例示的なオーディオ処理の流れ図を示す図である。

【図10】本発明の特徴を組み入れた例示的なベイジアン確率ネットワークを示す図である。

50

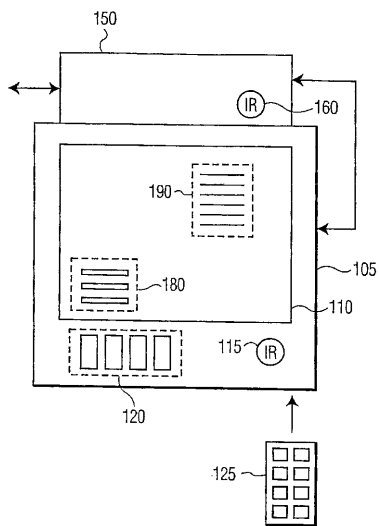
【図11】ビデオ・セグメント化処理の例示的な実施方法を表す流れ図である。

【図12】本発明によって監視される種々の特徴の例示的なタイムラインのイメージを備える図である。

【図13】コーラス検出処理の例示的な実施例の流れ図である。

【図14】ビデオからの構成要素を見つけて要約を自動的に生成するのに用い得るベイジアン確率ネットワークを示す図である。

【図1】



従来技術

【図2】

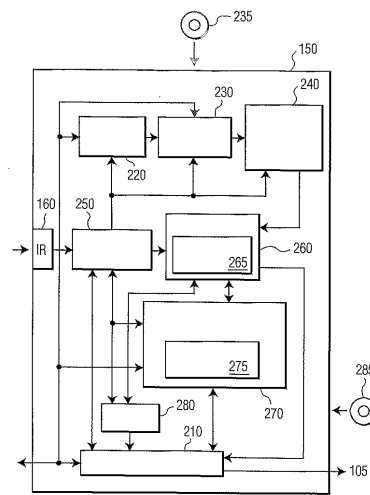


FIG. 2

【 図 3 】

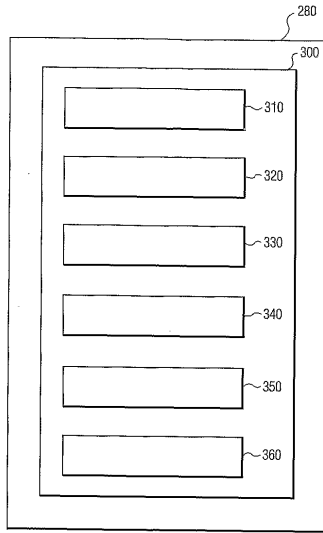


FIG. 3

【 図 4 】

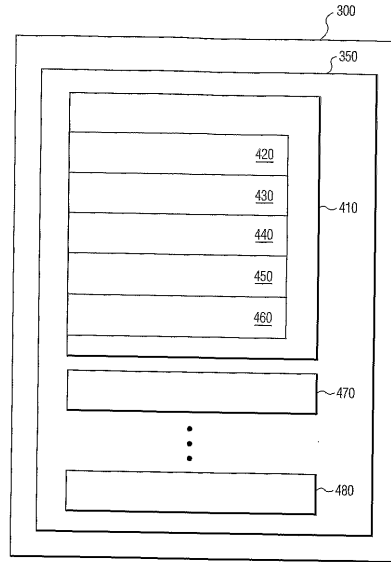


FIG. 4

【 図 5 】

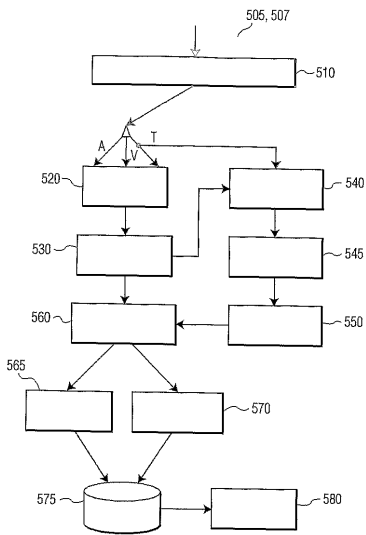


FIG. 5

【 図 6 】

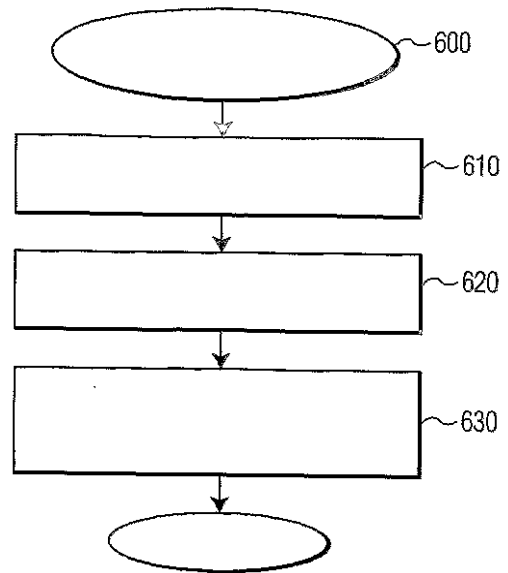
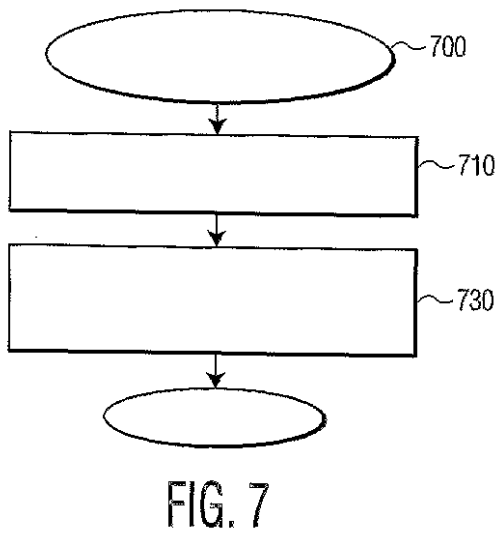
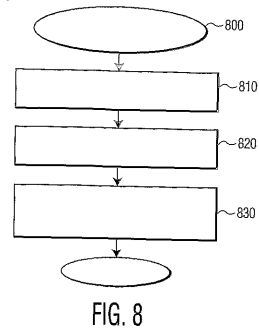


FIG. 6

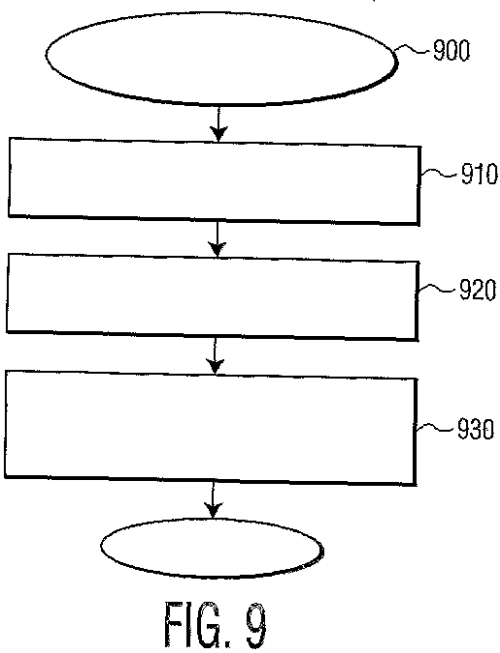
【 図 7 】



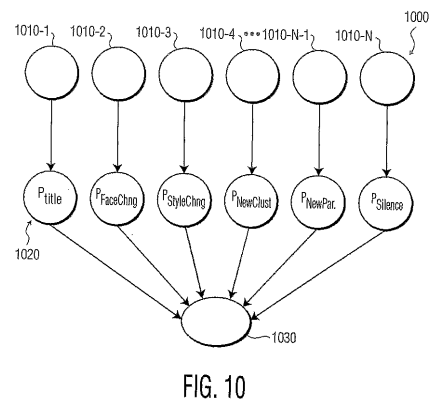
【 図 8 】



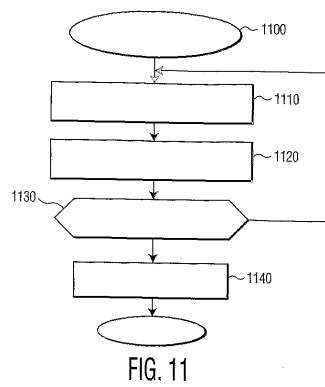
【 図 9 】



【 図 10 】



【 図 11 】



【 図 1 2 】

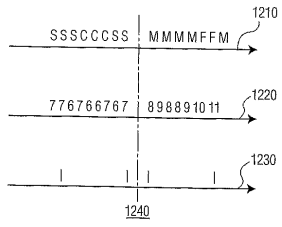


FIG. 12

【 図 1 4 】

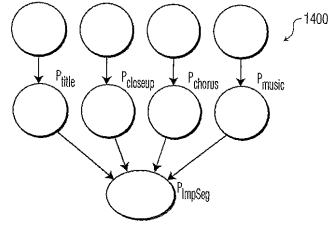


FIG. 14

【 図 1 3 】

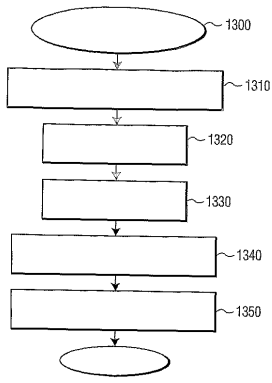


FIG. 13

【 国際調査報告 】

INTERNATIONAL SEARCH REPORT		International Application No PCT/IB2004/001068
A. CLASSIFICATION OF SUBJECT MATTER IPC 7 G06F17/30		
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) IPC 7 G06F		
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched		
Electronic data base consulted during the international search (name of data base and, where practical, search terms used) EPO-Internal, WPI Data, INSPEC		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	WO 2004/001626 A (KONINKL PHILIPS ELECTRONICS NV ; PHILIPS CORP (US)) 31 December 2003 (2003-12-31) cited in the application page 6, line 18 - page 16, line 6 ----- -/--	1,2,5,7, 9,18-25, 30-32
<input checked="" type="checkbox"/> Further documents are listed in the continuation of box C. <input checked="" type="checkbox"/> Patent family members are listed in 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 document 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 9 July 2004		Date of mailing of the international search report 26/07/2004
Name and mailing address of the ISA European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016		Authorized officer Güttlich, J

INTERNATIONAL SEARCH REPORT

International Application No

PCT/IB2004/001068

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	JASINSCHI R S ET AL: "Integrated multimedia processing for topic segmentation and classification" PROCEEDINGS 2001 INTERNATIONAL CONFERENCE ON IMAGE PROCESSING. ICIP 2001. THESSALONIKI, GREECE, OCT. 7 - 10, 2001, INTERNATIONAL CONFERENCE ON IMAGE PROCESSING, NEW YORK, NY : IEEE, US, vol. VOL. 1 OF 3. CONF. 8, 7 October 2001 (2001-10-07), pages 366-369, XP010563359 ISBN: 0-7803-6725-1 the whole document	1-18, 25-27, 29-32
Y	WO 00/45291 A (KONINKL PHILIPS ELECTRONICS NV) 3 August 2000 (2000-08-03) page 13, line 4 - page 18, line 3	1-18, 25-27, 29-32
Y	DIMITROVA N ET AL: "Personalizing video recorders using multimedia processing and integration" PROCEEDINGS OF 9TH ACM INTERNATIONAL CONFERENCE ON MULTIMEDIA, 2001, pages 564-567, XP002287757 OTTAWA, CA the whole document	1-18, 25-27, 29-32
Y	US 5 805 733 A (NORMILE JAMES ET AL) 8 September 1998 (1998-09-08)	5,27
A	column 4, line 55 - line 57; figure 1	1,19,25, 30
Y	NAPHADE M R ET AL: "AUDIO-VISUAL QUERY AND RETRIEVAL: A SYSTEM THAT USES DYNAMIC PROGRAMMING AND RELEVANCE FEEDBACK" JOURNAL OF ELECTRONIC IMAGING, SPIE + IS&T, US, vol. 10, no. 4, October 2001 (2001-10), pages 861-870, XP001104666 ISSN: 1017-9909	10
A	page 863, paragraph 4.1 - page 865, paragraph 4.4.3	1,13-15, 17
A	US 2002/164151 A1 (DIMITROVA NEVENKA ET AL) 7 November 2002 (2002-11-07) paragraph '0003! - paragraph '0006! paragraph '0027! - paragraph '0051!	1,19,25, 30
	-/--	

INTERNATIONAL SEARCH REPORT

International Application No

PCT/IB2004/001068

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	FOOTE J: "VISUALIZING MUSIC AND AUDIO USING SELF-SIMILARITY" ACM MULTIMEDIA, PROCEEDINGS OF THE INTERNATIONAL CONFERENCE, NEW YORK, NY, US, October 1999 (1999-10), pages 77-80, XP002265508 cited in the application the whole document	1,5,25
A	LI D ET AL: "Classification of general audio data for content-based retrieval" PATTERN RECOGNITION LETTERS, NORTH-HOLLAND PUBL. AMSTERDAM, NL, vol. 22, no. 5, April 2001 (2001-04), pages 533-544, XP004233004 ISSN: 0167-8655 cited in the application page 535 - page 539	1,15,17
A	ABDEL-MOTTALEB M ET AL: "MPEG -7: a content description standard beyond compression" 42ND MIDWEST SYMPOSIUM ON CIRCUITS AND SYSTEMS, vol. 2, 8 August 1999 (1999-08-08), pages 770-777, XP010511065 cited in the application page 771, paragraph 3.0 - page 776, paragraph 4.3	1,19,25, 30
A	DIMITROVA N, ZHANG H J: "Applications of video-content analysis and retrieval" MULTIMEDIA, IEEE, vol. 9, no. 3, July 2002 (2002-07), pages 42-55, XP002287758 ISSN: 1070-986 page 52, right-hand column - page 53, left-hand column; figure 8	1,19,25, 30

INTERNATIONAL SEARCH REPORT

 International Application No
 PCT/IB2004/001068

Patent document cited in search report		Publication date	Patent family member(s)	Publication date
WO 2004001626	A	31-12-2003	US 2003233929 A1	25-12-2003
			WO 2004001626 A1	31-12-2003
WO 0045291	A	03-08-2000	US 6608930 B1	19-08-2003
			CN 1295690 T	16-05-2001
			WO 0045291 A1	03-08-2000
			EP 1066577 A1	10-01-2001
			JP 2002536853 T	29-10-2002
			US 6614930 B1	02-09-2003
			US 6731788 B1	04-05-2004
US 5805733	A	08-09-1998	NONE	
US 2002164151	A1	07-11-2002	EP 1428147 A2	16-06-2004
			WO 02089008 A2	07-11-2002

フロントページの続き

(51) Int. Cl.		F I		テーマコード (参考)
G 1 1 B 27/00	(2006.01)	G 1 0 L 11/00	1 0 1 H	
		G 1 1 B 27/10	A	
		G 1 1 B 27/00	D	

(81) 指定国 AP(BW, GH, GM, KE, LS, MW, MZ, 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, HU, IE, IT, LU, MC, NL, 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, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW

(74) 代理人 100107766

弁理士 伊東 忠重

(72) 発明者 アグニホトリ, ラリタ

アメリカ合衆国 ニューヨーク州 1 0 5 1 0 - 8 0 0 1 ブライアクリフ・マナー ピー・オー
・ボックス 3 0 0 1

(72) 発明者 デイミトロワ, ネヴェンカ

アメリカ合衆国 ニューヨーク州 1 0 5 1 0 - 8 0 0 1 ブライアクリフ・マナー ピー・オー
・ボックス 3 0 0 1

(72) 発明者 ケンダー, ジョン

アメリカ合衆国 ニューヨーク州 1 0 5 1 0 - 8 0 0 1 ブライアクリフ・マナー ピー・オー
・ボックス 3 0 0 1

F ターム(参考) 5B075 ND14 ND16 NS01 QT04

5C053 FA14 FA21 FA23 FA30 GA11 GB09 GB12 GB38 JA16 LA07
LA14

5D015 CC11 CC13

5D077 AA22 AA23 CA11 CA12 CA20 DC11 DC22 DC40 DE05 DE11
EA33

5D110 AA13 AA14 AA27 DA02 DA04 DA12 DA17 DA19 DC22 DD01
EA07 EA08

【要約の続き】

るのに用い得る。