

(19) 日本国特許庁(JP)

(12) 公開特許公報(A)

(11) 特許出願公開番号

特開2008-178090
(P2008-178090A)

(43) 公開日 平成20年7月31日(2008.7.31)

(51) Int.Cl.		F I		テーマコード (参考)
HO4N 5/91 (2006.01)		HO4N 5/91	N	5C053
HO4N 5/765 (2006.01)		HO4N 5/91	L	5C122
HO4N 5/222 (2006.01)		HO4N 5/91	Z	
		HO4N 5/91	C	
		HO4N 5/222	Z	

審査請求 未請求 請求項の数 6 O L (全 16 頁)

(21) 出願番号 特願2007-324399 (P2007-324399)
 (22) 出願日 平成19年12月17日 (2007.12.17)
 (31) 優先権主張番号 特願2006-345752 (P2006-345752)
 (32) 優先日 平成18年12月22日 (2006.12.22)
 (33) 優先権主張国 日本国 (JP)

(71) 出願人 000004329
 日本ビクター株式会社
 神奈川県横浜市神奈川区守屋町3丁目12番地
 (74) 代理人 100083806
 弁理士 三好 秀和
 (74) 代理人 100101247
 弁理士 高橋 俊一
 (72) 発明者 若林 つきみ
 神奈川県横浜市神奈川区守屋町3丁目12番地 日本ビクター株式会社内
 Fターム(参考) 5C053 FA14 GB06 GB09 GB11 JA22
 JA23 LA01 LA14
 5C122 EA42 EA61 GA21 GA31

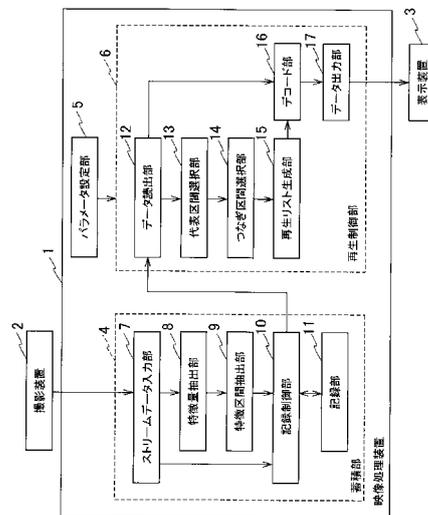
(54) 【発明の名称】 映像処理装置

(57) 【要約】

【課題】 ユーザにより撮影された映像について、映像内容を把握でき、かつ見易い要約映像を生成することができる映像処理装置を提供する。

【解決手段】 代表区間選択部 13 は、データストリームの特徴量を基に映像データの特徴的なシーンに対応する代表区間を選択し、つなぎ区間選択部 14 は、代表区間の導入部となるつなぎ区間を選択し、再生リスト生成部 15 は、代表区間とつなぎ区間を用いて要約映像を生成する。

【選択図】 図 1



【特許請求の範囲】**【請求項 1】**

第 1 の映像データ、この第 1 の映像データに同期した第 1 の音声データ、および前記第 1 の映像データについての所定の撮影条件を示す撮影情報を含む第 1 のデータストリームを入力する入力手段と、

前記第 1 のデータストリーム中の前記第 1 の映像データ、前記第 1 の音声データ、および前記撮影情報から所望の特徴量を抽出する特徴量抽出手段と、

この特徴量抽出手段で抽出した前記特徴量に基づいて、前記第 1 の映像データから特徴的なシーンを含む特徴区間を抽出する特徴区間抽出手段と、

この特徴区間抽出手段で抽出した各特徴区間について、前記特徴量を用いて、前記各特徴区間の特徴を評価するための評価値を算出し、この評価値に基づいて前記各特徴区間の中から代表区間を選択する代表区間選択手段と、

前記第 1 の映像データにおける前記代表区間以外の区間から、前記代表区間の導入部となるつなぎ区間を選択するつなぎ区間選択手段と、

前記第 1 の映像データから選択された前記代表区間と前記つなぎ区間とを撮影時刻順に並べて要約映像を生成する要約映像生成手段と

を備えることを特徴とする映像処理装置。

【請求項 2】

第 2 の映像データおよびこの第 2 の映像データに同期した第 2 の音声データを含む第 2 のデータストリームを格納する第 1 の記録手段を備え、

前記つなぎ区間選択手段は、前記第 1 の記録手段に予め格納された前記第 2 のデータストリーム中の前記第 2 の映像データから前記つなぎの区間を選択することを特徴とする請求項 1 に記載の映像処理装置。

【請求項 3】

第 3 の音声データを含む第 3 のデータストリームを格納する第 2 の記録手段と、

前記代表区間の映像データに同期した音声データにおける背景音のレベルを調整する調整手段と、

前記特徴量に基づいて、前記代表区間の映像データに対応付ける挿入音声データを前記第 3 のデータストリーム中の前記第 3 の音声データから選択する挿入音声選択手段と、

前記代表区間の映像データに同期した音声データと、前記挿入音声選択手段で選択した前記挿入音声データとを合成する音響合成手段と

を備えることを特徴とする請求項 1 または 2 に記載の映像処理装置。

【請求項 4】

通信ネットワークに接続し、データ送受信を行う通信手段を備え、

前記つなぎ区間選択手段は、前記通信手段および前記通信ネットワークを介して外部の映像音声データベースに接続され、前記映像音声データベースに予め格納された第 2 の映像データおよびこの第 2 の映像データに同期した第 2 の音声データを含む第 2 のデータストリーム中の前記第 2 の映像データから前記つなぎの区間を選択することを特徴とする請求項 1 に記載の映像処理装置。

【請求項 5】

前記代表区間の映像データに同期した音声データにおける背景音のレベルを調整する調整手段と、

前記通信手段および前記通信ネットワークを介して前記映像音声データベースに接続され、前記特徴量に基づいて、前記代表区間の映像データに対応付ける挿入音声データを、前記映像音声データベースに予め格納された第 3 の音声データを含む第 3 のデータストリーム中の前記第 3 の音声データから選択する挿入音声選択手段と、

前記代表区間の映像データに同期した音声データと、前記挿入音声選択手段で選択した前記挿入音声データとを合成する音響合成手段と

を備えることを特徴とする請求項 4 に記載の映像処理装置。

【請求項 6】

10

20

30

40

50

通信ネットワークに接続し、データ送受信を行う通信手段を備え、

前記通信手段は、前記通信ネットワークに接続された外部の撮影装置から前記第1のデータストリームを受信して前記入力手段に供給するとともに、前記要約映像のデータを前記通信ネットワークに接続された外部の表示装置に送信することを特徴とする請求項1に記載の映像処理装置。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、映像処理装置に係り、特に映像コンテンツの要約映像を生成する映像処理装置に関する。

10

【背景技術】

【0002】

近年、家庭用ビデオカメラの普及により、誰でも気軽に身近なイベントや風景を映像として記録保存することができるようになった。しかしながら、こうした所謂撮りっ放し映像は、撮影直後は楽しく見るものの、後々まで映像コンテンツとして鑑賞され活用される機会は少ない。また、一般のユーザにより撮影された映像は、失敗や不要な場面を多く含み、同じような場面が何度も映っているなど冗長度が高い。そのため、撮影直後に関係者でイベントを振り返るには好適であるが、後々の鑑賞に堪える映像コンテンツとするには、撮影映像を素材として整理し、選択された素材を製作意図に沿ってつなぎ合わせる編集作業を要する。

20

【0003】

動画像編集作業を行うためのパソコン用ソフトウェアも販売されてはいるが、これはパソコンや画像編集作業を趣味とする一部のマニア向けであり、一般ユーザにとってはパソコンを駆使しての映像編集は依然として面倒で敷居の高い作業である。

【0004】

このような状況を背景に、自動で映像データや音声データを編集して要約映像を生成する技術が各種提案されている。

【0005】

従来の要約手法としては、例えば、シーンを一定時間毎にスキップして再生する方法が考えられている。また、特許文献1には、フレーム間の画像の変化からシーン変化を検出し、動画像のシーンの長さ、シーン内の画像の変化度合を基に重要シーンを選択し要約再生する技術が開示されている。また、特許文献2には、動画像データを複数のシーンに分割し、複数の条件から再生シーンを選択する技術が開示されている。

30

【特許文献1】特開平6-149902号公報

【特許文献2】特開2002-142189号公報

【発明の開示】

【発明が解決しようとする課題】

【0006】

しかしながら、シーンを一定時間毎にスキップして再生する要約手法では、必ずしも映像の内容を把握し易いシーンが提示されるとは限らなかった。また、特許文献1に開示された技術のように、動きのあるシーンを重要選択するような設定では、動きの激しいシーンが次々と現れ、目まぐるしい要約映像になりがちである。

40

【0007】

また、特許文献2に開示された技術においては、画面の明るさや高周波成分など、シーンの評価に複数の条件を設定しているものの、各条件の評価結果からシーンを選択する基準は設定モード毎に一定であり、ある設定モードにおいては、重要と判定されるシーンと対極にあるような条件のシーンは選択されないため、やはり、同種のシーンが集まることになる。

【0008】

映画やテレビ番組の映像に対しては、予め専門家によるシーン編集が施されているため

50

、アクション区間に隣接するシーンを適宜組み合わせることにより、ある程度ストーリー性のある要約も可能であるが、一般のユーザがイベントや風景をスナップショット的に撮影した素材映像においては、このような手法は有効ではない。

【0009】

このように、上述のような従来技術では、ある基準で選び出した重要シーンを発生時間順に次々と提示するのみであって、ユーザにとって必ずしも見易い要約映像とはならないという問題があった。

【0010】

そこで本発明は、ユーザにより撮影された映像について、映像内容を把握でき、かつ見易い要約映像を生成することができる映像処理装置を提供することを目的とする。

10

【課題を解決するための手段】

【0011】

上記目的を達成するため、本発明の映像処理装置は、第1の映像データ、この第1の映像データに同期した第1の音声データ、および前記第1の映像データについての所定の撮影条件を示す撮影情報を含む第1のデータストリームを入力する入力手段と、前記第1のデータストリーム中の前記第1の映像データ、前記第1の音声データ、および前記撮影情報から所望の特徴量を抽出する特徴量抽出手段と、この特徴量抽出手段で抽出した前記特徴量に基づいて、前記第1の映像データから特徴的なシーンを含む特徴区間を抽出する特徴区間抽出手段と、この特徴区間抽出手段で抽出した各特徴区間について、前記特徴量を用いて、前記各特徴区間の特徴を評価するための評価値を算出し、この評価値に基づいて前記各特徴区間の中から代表区間を選択する代表区間選択手段と、前記第1の映像データにおける前記代表区間以外の区間から、前記代表区間の導入部となるつなぎ区間を選択するつなぎ区間選択手段と、前記第1の映像データから選択された前記代表区間と前記つなぎ区間とを撮影時刻順に並べて要約映像を生成する要約映像生成手段とを備えることを特徴とする。

20

【0012】

また、本発明の映像処理装置は、第2の映像データおよびこの第2の映像データに同期した第2の音声データを含む第2のデータストリームを格納する第1の記録手段を備え、前記つなぎ区間選択手段は、前記第1の記録手段に予め格納された前記第2のデータストリーム中の前記第2の映像データから前記つなぎの区間を選択することを特徴とする。

30

【0013】

また、本発明の映像処理装置は、第3の音声データを含む第3のデータストリームを格納する第2の記録手段と、前記代表区間の映像データに同期した音声データにおける背景音のレベルを調整する調整手段と、前記特徴量に基づいて、前記代表区間の映像データに対応付ける挿入音声データを前記第3のデータストリーム中の前記第3の音声データから選択する挿入音声選択手段と、前記代表区間の映像データに同期した音声データと、前記挿入音声選択手段で選択した前記挿入音声データとを合成する音響合成手段とを備えることを特徴とする。

【0014】

また、本発明の映像処理装置は、通信ネットワークに接続し、データ送受信を行う通信手段を備え、前記つなぎ区間選択手段は、前記通信手段および前記通信ネットワークを介して外部の映像音声データベースに接続され、前記映像音声データベースに予め格納された第2の映像データおよびこの第2の映像データに同期した第2の音声データを含む第2のデータストリーム中の前記第2の映像データから前記つなぎの区間を選択することを特徴とする。

40

【0015】

また、本発明の映像処理装置は、前記代表区間の映像データに同期した音声データにおける背景音のレベルを調整する調整手段と、前記通信手段および前記通信ネットワークを介して前記映像音声データベースに接続され、前記特徴量に基づいて、前記代表区間の映像データに対応付ける挿入音声データを、前記映像音声データベースに予め格納された第

50

3の音声データを含む第3のデータストリーム中の前記第3の音声データから選択する挿入音声選択手段と、前記代表区間の映像データに同期した音声データと、前記挿入音声選択手段で選択した前記挿入音声データとを合成する音響合成手段とを備えることを特徴とする。

【0016】

また、本発明の映像処理装置は、通信ネットワークに接続し、データ送受信を行う通信手段を備え、前記通信手段は、前記通信ネットワークに接続された外部の撮影装置から前記第1のデータストリームを受信して前記入力手段に供給するとともに、前記要約映像のデータを前記通信ネットワークに接続された外部の表示装置に送信することを特徴とする。

10

【発明の効果】

【0017】

本発明の映像処理装置によれば、データストリームの特徴量を基に映像データの特徴的なシーンに対応する代表区間を選択するとともに、代表区間の導入部となるつなぎ区間を選択して要約映像を生成するので、ユーザが映像内容を把握でき、かつ見易い要約映像を生成することができる。

【発明を実施するための最良の形態】

【0018】

以下、本発明の映像処理装置を実施するための最良の形態について、図面を参照して説明する。

20

【0019】

(第1の実施の形態)

図1は本発明の第1の実施の形態に係る映像処理装置の構成を示すブロック図である。図1に示すように、本発明の第1の実施の形態に係る映像処理装置1は、撮影装置2および表示装置3と接続して、映像および音声のコンテンツを蓄積、管理、再生する形態で実施して効果的である。具体的には、図2に示すように、映像処理装置1は、テレビモニタ22に接続した家庭用コンテンツ蓄積装置21に内蔵され、家庭用ビデオカメラ23を接続してデータ蓄積、視聴する使用形態に好適である。

【0020】

本発明の第1の実施の形態に係る映像処理装置1は、図1に示すように、蓄積部4と、パラメータ設定部5と、再生制御部6とを備える。蓄積部4は、ストリームデータ入力部7と、特徴量抽出部8と、特徴区間抽出部9と、記録制御部10と、記録部11とを備える。

30

【0021】

ストリームデータ入力部7は、撮影装置2で撮影録画された映像データ、この映像データに同期した音声データ、および撮影日時や画質などの撮影情報を含むデータストリームを取得し、取得したデータストリームから各データを分離する。データストリームは様々な形式が利用できるが、本実施の形態では、録画開始/停止のショット毎に映像データ、音声データ、撮影情報を含むデータファイルが作成され保存される形式を例に用いる。ユーザ設定可能なタイトル情報も、メタデータとしてデータファイルに保存される。

40

【0022】

特徴量抽出部8は、各ショットの分離された映像データ、音声データ、および撮影情報の各データから特徴量を抽出する。ここで、映像データの特徴量としては、例えば、映像のエッジ分布、色および輝度の分布、および、それらの変化量を用いることができる。音声データの特徴量としては、例えば、音声の全体パワーおよび周波数分布を用いることができる。撮影情報の特徴量としては、撮影時間情報や、撮影装置2がGPS受信機を搭載している場合は撮影位置情報などを用いることができる。

【0023】

また、単純な映像および音響特性のみでなく、映像情報を基にオブジェクトを抽出し、オブジェクトサイズ、オブジェクトの動き情報を数値化し、特徴量としてもよい。また、

50

音声情報を基に鳥の声、人の声など音源の種類を分類し、それぞれのレベルを特徴量としてもよい。

【0024】

特徴区間抽出部9は、特徴量抽出部8で抽出した特徴量に基づいて、ショット内の特徴区間および平均区間を抽出し、特徴区間の位置等を示す特徴区間情報、平均区間の位置等を示す平均区間情報を生成して出力する。

【0025】

記録制御部10は、ストリームデータ入力部7から入力されるデータストリームと、特徴量抽出部8で抽出した特徴量、および特徴区間抽出部9からの特徴区間情報、平均区間情報に対応付けて、HDD等からなる記録部11に保存する。

10

【0026】

パラメータ設定部5は、再生時において、ユーザより選択されたタイトル、再生モード、再生時間をパラメータとして設定する。

【0027】

再生制御部6は、データ読出部12と、代表区間選択部13と、つなぎ区間選択部14と、再生リスト生成部15と、デコード部16と、データ出力部17とを備える。

【0028】

データ読出部12は、ユーザより指定されたデータストリームおよび対応する特徴量、特徴区間情報、平均区間情報を記録制御部10を介して記録部11より読み出す。

【0029】

代表区間選択部13は、特徴区間情報に基づき、設定された再生モードにより、特徴量を用いて各特徴区間を評価し、代表区間を選択する。つなぎ区間選択部14は、代表区間以外の区間から、代表区間の導入部となるつなぎ区間を選択する。

20

【0030】

再生リスト生成部15は、選択された代表区間と対応するつなぎ区間とを代表区間の撮影時刻順に並び替え、要約映像を生成するための再生リストを生成する。デコード部16は、再生リストのデータストリームをデコードしてデータ出力部17へ出力する。データ出力部17は、再生リストの映像データ、音声データを表示装置3へ出力する。

【0031】

次に、第1の実施の形態に係る映像処理装置1の動作について説明する。

30

【0032】

まず、ストリームデータ入力部7は、撮影装置2で撮影録画された映像データ、この映像データに同期した音声データ、および撮影情報を含むデータストリームを取得すると、取得したデータストリームから各データを分離する。次に、特徴量抽出部8は、各ショットの分離された映像データ、音声データ、および撮影情報の各データから特徴量を抽出する。

【0033】

次に、特徴区間抽出部9は、特徴量抽出部8で抽出した特徴量に基づいて、ショット内の特徴区間および平均区間を抽出する。ここで、映像の特徴量として輝度情報を用い、ショット内のフレーム全体での輝度平均値、ショット中の輝度変化、および輝度変化のレベルから特徴区間を抽出する例を説明する。

40

【0034】

L_t を時刻 t におけるフレームの輝度平均値、 L_m をショット全体での輝度平均値とし、閾値 $Th_1 = k_1 \sigma$ (k_1 は係数、 σ はショット内での輝度の標準偏差)とすると、下記の(数式1)を満たす時刻 t を特徴的なフレームとし、下記の(数式2)で求められる P_t を特徴量の得点とする。

【0035】

$$|L_t - L_m| > Th_1 \quad (\text{数式1})$$

$$P_t = (|L_t - L_m| - Th_1) / Th_1 \quad (\text{数式2})$$

そして、特徴区間抽出部9は、特徴的なフレームが連続するところを特徴区間として抽

50

出する。時刻 t_1 から時刻 t_2 の間が特徴区間として抽出される場合、特徴区間の得点は、時刻 $t_1 \sim t_2$ 間の得点の平均とする。

【0036】

一方、特徴区間抽出部 9 は、 $Th_2 = k_2$ ($k_2 < k_1$) とし、下記の (数式 3) を満たす時刻 t をショット内の平均的なフレームとし、平均的なフレームが安定して連続するところを平均区間として抽出する。

【0037】

$$|Lt - Lm| < Th_2 \quad (\text{数式 3})$$

なお、特徴区間抽出部 9 は、手振れ、ピンボケなどのミスショット区間は特徴量を基に検出し、特徴区間設定から予め除去する。

10

【0038】

図 3 に特徴区間および平均区間抽出の概念図を示す。図 3 において、特徴量 A について特徴区間 3 1 A および平均区間 3 1 B、特徴量 B について特徴区間 3 2 A₁、3 2 A₂ および平均区間 3 1 B、特徴量 C について特徴区間 3 3 A および平均区間 3 3 B、特徴量 D について特徴区間 3 4 A および平均区間 3 4 B が抽出され、抽出されたそれぞれの区間について得点が算出された例を示している。

【0039】

そして、特徴区間抽出部 9 は、特徴区間の位置情報および得点を含む特徴区間情報、平均区間の位置情報および得点を含む平均区間情報を生成して出力する。

【0040】

20

次に、記録制御部 10 は、ストリームデータ入力部 7 から入力されるデータストリームと、特徴量抽出部 8 で抽出した特徴量、および特徴区間抽出部 9 からの特徴区間情報、平均区間情報に対応付けて記録部 11 に保存する。

【0041】

再生時には、パラメータ設定部 5 は、ユーザより選択されたタイトル、再生モード、再生時間をパラメータとして設定する。そして、再生制御部 6 のデータ読出部 12 は、ユーザより指定されたデータストリームおよび対応する特徴量と特徴区間情報、平均区間情報を、記録制御部 10 を介して記録部 11 より読み出す。

【0042】

次に、代表区間選択部 13 は、特徴区間情報に基づき、設定された再生モードにより、各特徴区間の得点を評価し、代表区間を選択する。ここで、 n 種類の特徴量を用いて特徴区間を評価する際に、 i 番目の特徴量に関する時刻 t を含む特徴区間の得点を Pit とすると、設定された再生モード m での特徴区間の評価値 Smt は、下記の (数式 4) により求めることができる。

30

【数 1】

$$Smt = \sum_{i=1}^n (Cmi \times Pit) \quad (\text{数式 4})$$

【0043】

40

ここで、 Cmi は再生モード m での i 番目の特徴量の重み付け係数である。例えば、子供など人物中心であれば、画面中央部の肌色、音声の中高域成分の特徴量に重み付けして評価値を算出することができる。

【0044】

そして、代表区間選択部 13 は、特徴区間を評価値の上位から並べて代表区間候補とし、設定した再生時間を超えない範囲で、代表区間候補の上位から代表区間を選択する。代表区間は、見易さを損なわない最短の長さおよび指定再生時間に応じた最長の長さを範囲設定し、範囲内の長さとするよう特徴量を基に調整してもよい。

【0045】

次に、つなぎ区間選択部 14 は、評価値が上位の代表区間から順に、特徴量に基づいて

50

、代表区間と背景の親和性が高く中心部に大きな動きの少ない落ち着いた映像区間を短時間、つなぎ区間として選択する。つなぎ区間は、代表区間と同一ショット内の映像区間を用い、代表区間の導入部に位置付ける。例えば、代表区間と同一ショット内の平均区間から動きの少ない区間を2～3秒選択する。また、簡易にショット内の落ち着いた映像区間を選択する方法として、一般ユーザに多く見られる、ショット開始時に周囲の落ち着いた映像から撮り始めるという撮影行動特性を利用し、ショット開始時点から一定時間を選択してもよい。

【0046】

また、指定タイトル以外に、予め記録部11に蓄積された、映像データおよび音声データを含むデータストリームを用い、代表区間と背景の特徴量の親和性が高いショットの中からつなぎ区間を選択してもよい。また、予めつなぎ用の映像を対応する特徴量と共にデータベース化し、代表区間の特徴量と親和性の高いつなぎ区間を選択してもよい。

10

【0047】

次に、再生リスト生成部15は、図4に示すように、選択された代表区間40A, 41A, ...と、それぞれに対応するつなぎ区間40B, 41B, ...とを代表区間の撮影時刻順に並び替え、要約映像を生成するための再生リストを生成する。そして、デコード部16は、再生リストのデータストリームをデコードし、データ出力部17で表示装置3へ要約映像の映像データ、音声データを出力する。

【0048】

ここで、つなぎ区間と代表区間との接続部分では、ディゾルブやフェードイン・フェードアウト効果を用いて、各区間を滑らかに接続するようにしてもよい。また、図5に示すように、映像をつなぎ区間51B, 53Bから代表区間51A, 53Aに切り替えるより先に、音声をつなぎ区間52B, 54Bから代表区間52A, 54Aに切り替えるように調整してもよい。このようにすることで、ユーザは予め音声で代表区間への移行を認識することにより映像を違和感なく視聴でき、短時間で内容を把握し易くなる。

20

【0049】

上記説明のように、第1の実施の形態によれば、データストリームの特徴量を基に映像データの特徴的なシーンに対応する代表区間を選択するとともに、代表区間の導入部となるつなぎ区間を選択し、選択した代表区間とつなぎ区間とを並べて要約映像を生成するので、ユーザが映像内容を把握でき、かつ見易い要約映像を生成することができる。

30

【0050】

なお、上記第1の実施の形態で説明したように、今日普及している家庭用ビデオカメラにおいては、録画/停止の撮影ショット毎にインデックスを生成し記録する形式が一般的であるので、強いてショットチェンジ検出を行う必要のない場合が多いが、旧型の撮影機器で撮影したインデックス情報のない映像データを処理する際には、フレーム間の相関など公知の手法を用いて予めショット区切を検出し、上記第1の実施の形態の処理を適用すればよい。

【0051】

(第2の実施の形態)

図6は本発明の第2の実施の形態に係る映像処理装置の構成を示すブロック図である。なお、図6に示す第2の実施の形態の構成要素において、第1の実施の形態と同様の構成要素については同一番号を付し、詳細な説明は省略する。

40

【0052】

図6に示すように、第2の実施の形態に係る映像処理装置1Aは、第1の実施の形態に係る映像処理装置1に対し、再生制御部6にBGM合成部61を追加した構成である。また、映像処理装置1Aの記録部11には予めBGM音声データを含むデータストリームが格納されている。

【0053】

BGM合成部61は、代表区間の映像データに同期した音声データにおける背景音のレベルを調整する周波数特性調整部62と、代表区間の映像データに対応付けるBGM音声

50

データを選択する BGM 選択部 63 と、代表区間の映像データに同期した音声データと BGM 選択部 63 で選択した BGM 音声データとを合成する音響合成部 64 とを備える。

【0054】

このように構成された第 2 の実施の形態に係る映像処理装置 1A において、ストリームデータ入力から再生リスト生成までの区間選択手順は、第 1 の実施の形態と同様である。

【0055】

再生リストが生成されると、周波数特性調整部 62 は、ショットの特徴量に基づき、代表区間の音声データに対して、代表区間に現れる特徴的な周波数成分を保持し、継続的な背景音の周波数成分を小さくするようレベル調整する。

【0056】

BGM 選択部 63 は、ショットの特徴量に基づき、記録部 11 に格納された BGM 音声データから適当な BGM 音声データを選択する。ここで、映像データの色合や複雑さ情報を BGM 音声データの音響特徴量と対応付けたテーブルを予め用意し、参照するようにしてもよい。また、好ましくは、映像処理装置 1A に地域情報をインデックスとして検索可能な BGM データベースを内蔵またはネットワーク接続にて備え、撮影位置情報を含むデータストリームに対しては、撮影位置情報を地域情報に変換して BGM データベースを検索し、撮影場所に好適な BGM を選択するようにしてもよい。また、パラメータ設定部 5 にユーザが好みの BGM を選択する機能を設けてもよい。

【0057】

そして、音響合成部 64 は、周波数特性調整部 62 で調整された代表区間の音声データと、BGM 選択部 63 で選択された BGM 音声データとを合成して出力する。また、音響合成部 64 は、代表区間の特徴的な周波数成分が存在する区間は BGM の音量を下げ、代表区間の特徴的な音声内容が十分聞こえるよう調整する。

【0058】

図 7 は BGM 合成処理の概念図である。屋外で撮影する場合は、意図した撮影対象以外の音声背景が多く含まれる。背景音は周囲の雰囲気を感じるためには有効であるが、そのままの音量で継続的に聞くのには適さない音も多い。野外での風切音や車の騒音などがしばしば継続的な背景音に該当する。図 7 では、低周波数成分のレベルを徐々に下げ、風切音を低減させ、BGM のレベルを徐々に上げ、合成する例を示している。鳥のさえずり声や人の話し声が入る部分では、BGM のレベルを下げ、特徴的な音声背景が十分聞こえるようにする。

【0059】

また、BGM の選曲は、ショットの特徴量に基づいて、背景の特徴量が類似する区間では同じ BGM を使い、場面が転換する区間、すなわち背景の特徴量が大きく異なる区間で BGM を変更するようにしてもよい。図 7 では、第 2 のショットの映像から第 3 のショットの映像に切り替わるときに、第 1 の BGM から第 2 の BGM に切り替える例を示している。

【0060】

このように、第 2 の実施の形態によれば、映像データに同期した音声データのうち、背景音に該当する周波数成分のレベルを調整し、映像データに好適な BGM 音声データと組み合わせ再生することにより、ユーザが映像に同期した音声の内容を容易に把握でき、かつ聴き易く心地よい音声を提供することができる。

【0061】

(第 3 の実施の形態)

図 8 は本発明の第 3 の実施の形態に係る映像処理装置の構成を示すブロック図である。なお、図 8 に示す第 3 の実施の形態の構成要素において、第 1 および第 2 実施の形態と同様の構成要素については同一番号を付し、詳細な説明は省略する。

【0062】

図 8 に示すように、第 3 の実施の形態に係る映像処理装置 1B は、第 1 の実施の形態に係る映像処理装置 1 に対し、情報通信接続部 18 を追加し、つなぎ区間選択部 14 に映像

10

20

30

40

50

検索部 141 を追加した構成である。

【0063】

情報通信接続部 18 は、通信ネットワーク 20 経由で映像処理装置 1B と外部の映像音声データベース 30 とを接続する。

【0064】

映像検索部 141 は、情報通信接続部 18 を介してネットワーク接続された映像音声データベース 30 にアクセスし、予め映像音声データベース 30 に格納された、映像データおよび音声データを含むデータストリームからつなぎ区間の映像を取得する。

【0065】

このように構成された第 3 の実施の形態に係る映像処理装置 1B において、ストリームデータ入力から代表区間選択までの動作は、第 1 の実施の形態と同様である。

10

【0066】

つなぎ区間選択部 14 は、評価値が上位の代表区間から順に、特徴量に基づいて、代表区間と背景の親和性が高く中心部に大きな動きの少ない落ち着いた映像区間を短時間、つなぎ区間として選択する。ここで、つなぎ区間選択部 14 は、映像検索部 141 により、情報通信接続部 18 を介してネットワーク接続された映像音声データベース 30 にアクセスし、予め映像音声データベース 30 に格納された、映像データおよび音声データを含むデータストリームからつなぎ区間の映像データを検索して取得し、このつなぎ区間の映像データをデコード部 16 に供給する。

【0067】

映像検索部 141 では、つなぎ区間を選択するための映像音声データベース 30 への検索条件を設定する。再生対象のデータストリームの映像データおよび音声データの特徴量に関しては、各特徴量について映像音声データベース 30 のタグと対応させ、代表区間と親和性の高い範囲を検索条件として設定する。撮影情報の特徴量に関しては、映像音声データベース 30 の対応するタグおよび、関連するタグに検索条件を設定する。

20

【0068】

例えば、再生対象のデータストリームのタイトル情報として記録しているキーワードをタイトルタグのみに設定して検索を行う場合、タイトルタグには大きな括りのキーワードを用い、シーン毎の見出しタグやテキスト注釈タグに検索条件を設定する。同様に、撮影位置情報についても、撮影場所としてのタグの他、タイトルやシーン見出しに検索条件を設定する。また、撮影日時情報は、季節や朝、昼、夜等、類似の特徴をもつ映像を絞り込める検索範囲を設定する。

30

【0069】

次に、再生リスト生成部 15 は、選択された代表区間とそれぞれに対応するつなぎ区間とを代表区間の撮影時刻順に並び替え、要約映像を生成するための再生リストを生成する。そして、デコード部 16 は、再生リストのデータストリームをデコードし、データ出力部 17 で表示装置 3 へ要約映像の映像データ、音声データを出力する。

【0070】

このように、第 3 の実施の形態によれば、つなぎ区間には撮影時の映像と異なる映像ソースを用いるので、ユーザにとって新鮮味のある要約映像を提供することができる。

40

【0071】

なお、つなぎ区間の映像としては、動画像の短時間の区間の他、静止画素材を用い、スクロール、拡大など適当な演出効果を加えて表示するようにしてもよい。

【0072】

また、つなぎ区間の検索を容易にするために、入力されたデータストリームの記録部 11 への記録時に、特徴量情報を映像音声データベース 30 のタグ情報と対応する形式で記録することが好ましい。図 9 は入力されたデータストリームの特徴量と映像音声データベース 30 のタグ情報の対応付け例を示す図である。図 9 に示すように、撮影情報の特徴量である撮影日時、タイトル、撮影位置情報を、それぞれ作成日、タイトル、撮影場所のタグと対応付け、映像データの特徴量である背景色、エッジ分布、動き情報を、それぞれ色

50

分布、エッジ分布、動きアクティビティと対応付け、音声データの特徴量である音声パワー、周波数分布を、それぞれ音声パワー/無音、音響効果のタグと対応付けることができる。

【0073】

(第4の実施の形態)

図10は本発明の第4の実施の形態に係る映像処理装置の構成を示すブロック図である。なお、図10に示す第4の実施の形態の構成要素において、第1～第3の実施の形態と同様の構成要素については同一番号を付し、詳細な説明は省略する。

【0074】

図10に示すように、第4の実施の形態に係る映像処理装置1Cは、第2の実施の形態に係る映像処理装置1Aに対し、第3の実施の形態に係る映像処理装置1Bと同様の、情報通信接続部18と、つなぎ区間選択部14の映像検索部141とを追加した構成である。

10

【0075】

第2の実施の形態では、記録部11に予めBGM音声データを含むデータストリームが格納され、BGM選択部63は、記録部11に格納されたBGM音声データから適当なBGM音声データを選択したが、第4の実施の形態では、BGM選択部63は、情報通信接続部18を介してネットワーク接続された映像音声データベース30にアクセスし、予め映像音声データベース30に格納された、BGM音声データを含むデータストリームから適当なBGM音声データを選択して取得する。

20

【0076】

第4の実施の形態に係る映像処理装置1Cにおいて、ストリームデータ入力から再生リスト生成までの区間選択手順は、第3の実施の形態と同様である。

【0077】

また、映像データに同期した音声データのうち、背景音に該当する周波数成分のレベルを調整し、映像データに好適なBGM音声データと組み合わせて再生する手順は、映像音声データベース30からBGM音声データを選択することの他は、第2の実施の形態と同様である。

【0078】

第4の実施の形態によれば、つなぎ区間に撮影時の映像と異なる映像ソースを用いることで、ユーザにとって新鮮味のある要約映像を提供することができ、また、背景音に該当する周波数成分のレベルを調整し、映像データに好適なBGM音声データと組み合わせて再生することにより、ユーザが映像に同期した音声の内容を容易に把握でき、かつ聴き易く心地よい音声を提供することができる。

30

【0079】

(第5の実施の形態)

図11は本発明の第5の実施の形態に係る映像処理装置の構成を示すブロック図である。なお、図11に示す第5の実施の形態の構成要素において、第1～第4の実施の形態と同様の構成要素については同一番号を付し、詳細な説明は省略する。

【0080】

図11に示すように、第5の実施の形態に係る映像処理装置1Dは、第1の実施の形態に係る映像処理装置1に対し、情報通信接続部18を追加し、デコード部16をストリーム編集部19に置き換えた構成である。

40

【0081】

情報通信接続部18は、通信ネットワーク20経由で、映像処理装置1Dと外部の撮影装置70、表示装置80とを接続する。

【0082】

ストリーム編集部19は、撮影映像のストリームデータの再生リストに該当する区間を編集し、要約映像のストリームデータを生成する。

【0083】

50

撮影装置 70 は、撮影映像の映像データ、この映像データに同期した音声データ、および撮影情報を含むデータストリームを生成する撮影部 71 と、撮影部 71 で生成されたデータストリームを格納する記録部 72 と、通信ネットワーク 20 経由で映像処理装置 1D と接続する情報通信接続部 73 とを備える。

【0084】

表示装置 80 は、通信ネットワーク 20 経由で映像処理装置 1D と接続する情報通信接続部 81 と、映像処理装置 1D から送信される要約映像のデータストリームをデコードするデコード部 82 と、要約映像を表示する表示部 83 とを備える。

【0085】

次に、第 5 の実施の形態に係る映像処理装置 1D の動作について説明する。

10

【0086】

撮影装置 70 の撮影部 71 で撮影され記録部 72 に格納された映像データ、この映像データに同期した音声データ、撮影情報を含むデータストリームは、情報通信接続部 73 を介して通信ネットワーク 20 経由で映像処理装置 1D に送られる。

【0087】

映像処理装置 1D のストリームデータ入力部 7 は、情報通信接続部 18 を介して通信ネットワーク 20 経由で撮影装置 70 からデータストリームを取得する。ストリームデータ入力部 7 は、取得したデータストリームから各データを分離する。

【0088】

その後、第 1 の実施の形態と同様の処理により、特徴量抽出部 8 で各ショットの特徴量を抽出し、特徴区間抽出部 9 で特徴区間情報、平均区間情報を生成する。そして、記録制御部 10 により、ストリームデータ入力部 7 から入力されるデータストリームと、特徴量抽出部 8 で抽出した特徴量、および特徴区間抽出部 9 からの特徴区間情報、平均区間情報とを対応付けて記録部 11 に保存する。

20

【0089】

再生時には、パラメータ設定部 5 は、通信ネットワーク 20 経由で、ユーザより選択されたタイトル、再生モード、再生時間をパラメータとして設定する。ユーザは表示装置 80 を操作することによりパラメータを指定することができる。

【0090】

そして、第 1 の実施の形態と同様の処理により、代表区間選択部 13 で代表区間を選択し、つなぎ区間選択部 14 でつなぎ区間を選択し、再生リスト生成部 15 で要約映像を生成するための再生リストを生成する。

30

【0091】

その後、ストリーム編集部 19 は、ストリームデータの再生リストに該当する区間を編集し、要約映像のデータストリームを情報通信接続部 18 へ出力する。情報通信接続部 18 は、通信ネットワーク 20 を介して表示装置 80 に要約映像のデータストリームを送信する。

【0092】

表示装置 80 の情報通信接続部 81 は、通信ネットワーク 20 を介して映像処理装置 1D で作成された要約映像のデータストリームを受信すると、この要約映像のデータストリームをデコード部 82 に送る。デコード部 82 は、要約映像のデータストリームをデコードし、表示部 83 により、デコードされた映像および音声を表示する。

40

【0093】

このように、第 5 の実施の形態によれば、撮影映像のデータストリームを通信ネットワーク 20 経由で映像処理装置 1D に送ることにより、時間、場所の制約なく、撮影映像のデータストリーム、特徴量、および特徴区間情報、平均区間情報に対応付けて蓄積することができる。また、特徴区間の要約映像データを通信ネットワーク 20 経由で表示装置 80 に伝送することにより、時間、場所の制約なく要約映像を視聴することができる。例えば、旅行中に撮影した映像の要約を当日中に知人が自宅で視聴する等、リアルタイムに近い情報共有が可能となる。

50

【 0 0 9 4 】

なお、撮影装置 7 0 において撮影映像のデータストリームを送信する際に、映像タイトルとともに要約映像配信先の表示装置 8 0 を指定し、映像処理装置 1 D では蓄積処理が完了後に指定された表示装置 8 0 へ新たに視聴可能となった撮影タイトル情報を通知するようにしてもよい。

【 図面の簡単な説明 】

【 0 0 9 5 】

【 図 1 】 本発明の第 1 の実施の形態に係る映像処理装置の構成を示すブロック図である。

【 図 2 】 本発明の映像処理装置の好適な使用形態を示す図である。

【 図 3 】 特徴区間および平均区間抽出の概念図である。

10

【 図 4 】 再生リスト生成の概念図である。

【 図 5 】 つなぎ区間から特徴区間への切り替えタイミングの例を示す概念図である。

【 図 6 】 本発明の第 2 の実施の形態に係る映像処理装置の構成を示すブロック図である。

【 図 7 】 B G M 合成処理の概念図である。

【 図 8 】 本発明の第 3 の実施の形態に係る映像処理装置の構成を示すブロック図である。

【 図 9 】 入力されたデータストリームの特徴量と映像音声データベースのタグ情報の対応付け例を示す図である。

【 図 1 0 】 本発明の第 4 の実施の形態に係る映像処理装置の構成を示すブロック図である。

。

【 図 1 1 】 本発明の第 5 の実施の形態に係る映像処理装置の構成を示すブロック図である。

20

。

【 符号の説明 】

【 0 0 9 6 】

1 , 1 A ~ 1 D 映像処理装置

2 , 7 0 撮影装置

3 , 8 0 表示装置

4 蓄積部

5 パラメータ設定部

6 再生制御部

7 ストリームデータ入力部

30

8 特徴量抽出部

9 特徴区間抽出部

1 0 記録制御部

1 1 , 7 2 記録部

1 2 データ読出部

1 3 代表区間選択部

1 4 つなぎ区間選択部

1 5 再生リスト生成部

1 6 , 8 2 デコード部

1 7 データ出力部

40

1 8 , 7 3 , 8 1 情報通信接続部

1 9 ストリーム編集部

2 0 通信ネットワーク

3 0 映像音声データベース

6 1 B G M 合成部

6 2 周波数特性調整部

6 3 B G M 選択部

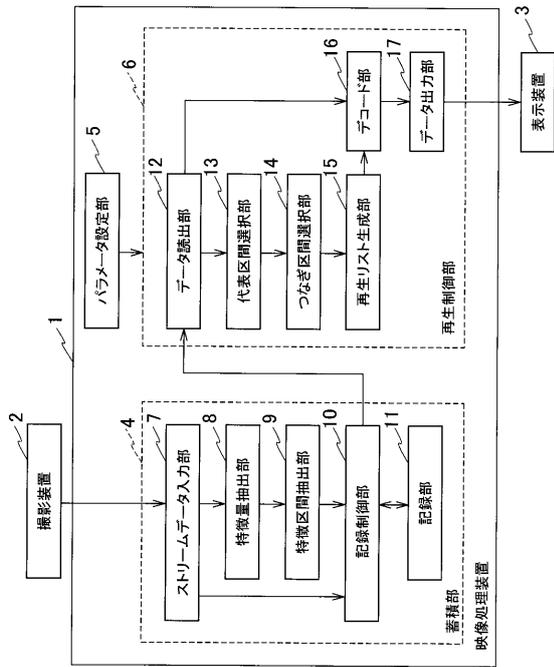
6 4 音響合成部

8 3 表示部

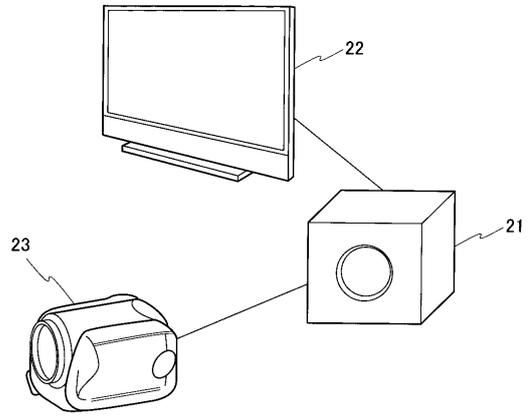
1 4 1 映像検索部

50

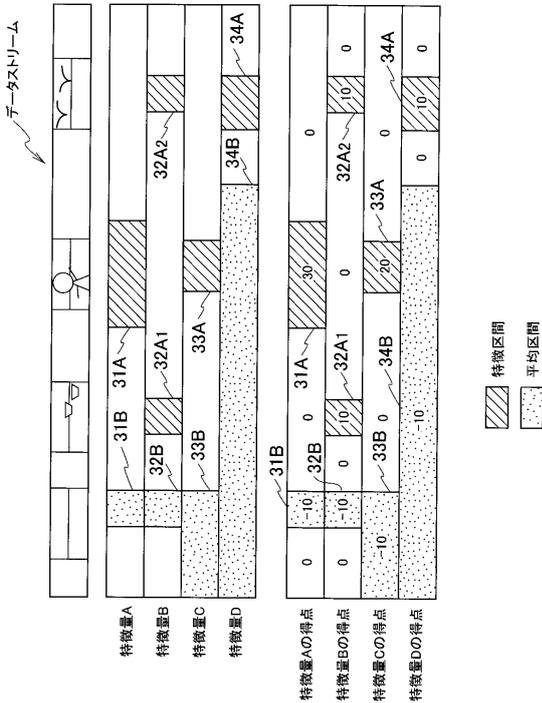
【 図 1 】



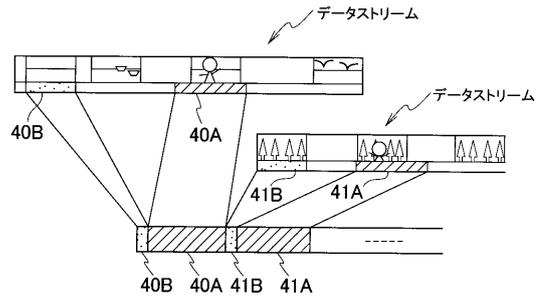
【 図 2 】



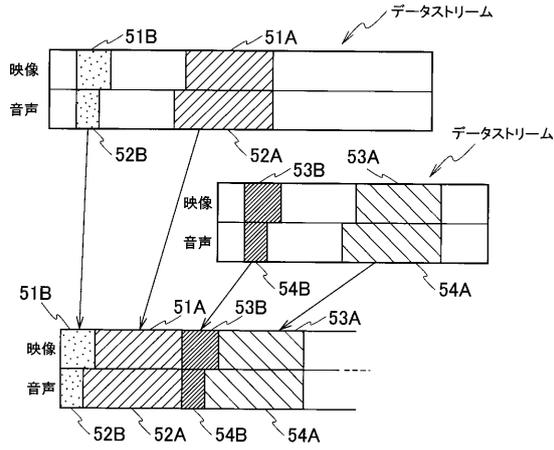
【 図 3 】



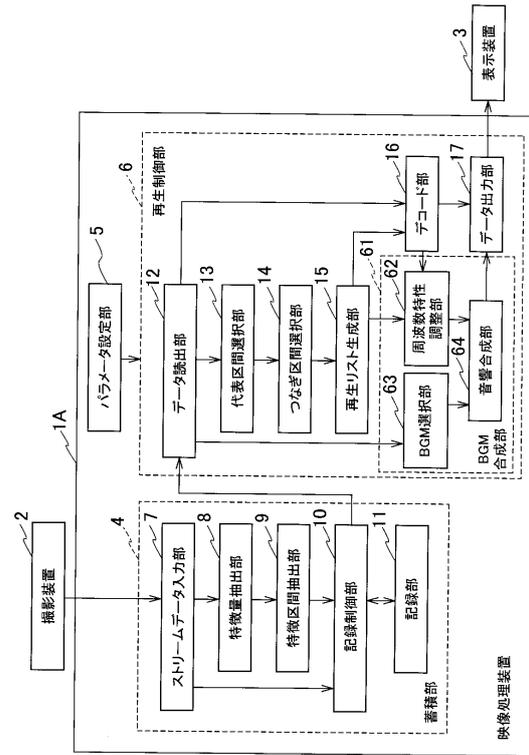
【 図 4 】



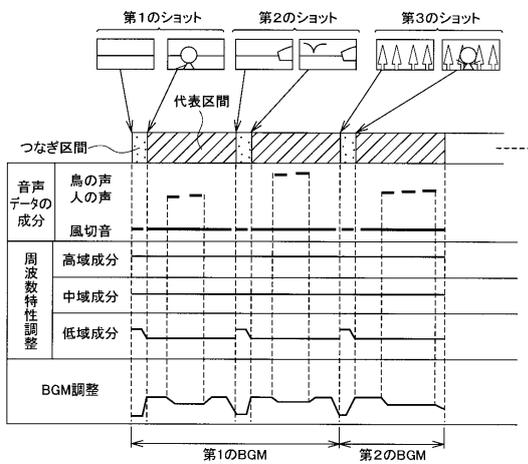
【図5】



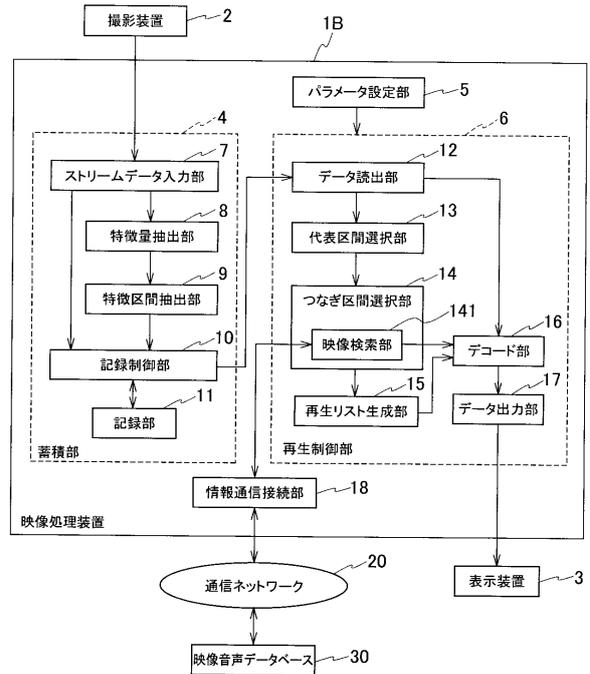
【図6】



【図7】



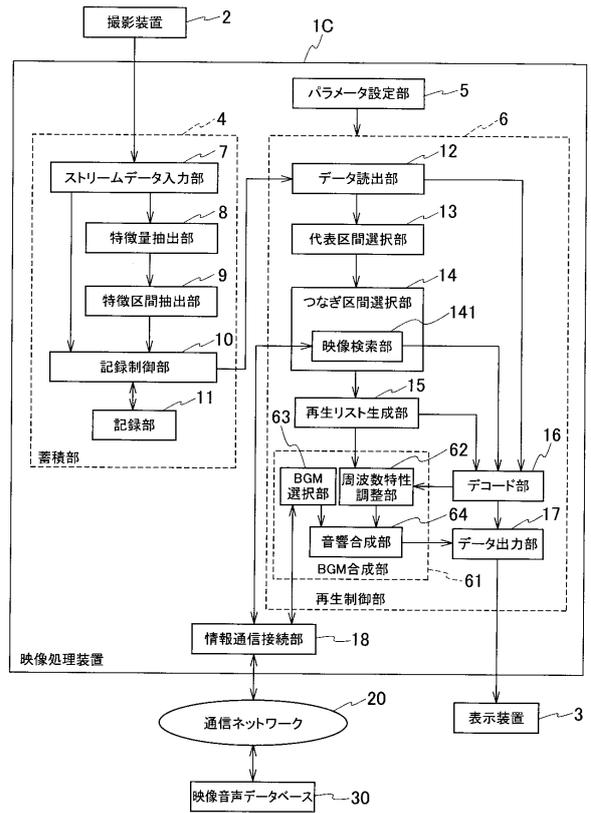
【図8】



【 図 9 】



【 図 1 0 】



【 図 1 1 】

