

(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2008-60622

(P2008-60622A)

(43) 公開日 平成20年3月13日(2008.3.13)

(51) Int.Cl.	F I	テーマコード (参考)
H 0 4 N 5/91 (2006.01)	H 0 4 N 5/91 N	5 C 0 5 3
G 1 1 B 27/00 (2006.01)	G 1 1 B 27/00 D	5 D 0 4 4
G 1 1 B 20/10 (2006.01)	G 1 1 B 20/10 G	5 D 1 1 0
	G 1 1 B 20/10 3 O 1 Z	

審査請求 未請求 請求項の数 14 O L (全 32 頁)

(21) 出願番号	特願2006-231429 (P2006-231429)	(71) 出願人	000002185
(22) 出願日	平成18年8月29日 (2006. 8. 29)		ソニー株式会社
			東京都港区港南1丁目7番1号
		(74) 代理人	100112955
			弁理士 丸島 敏一
		(72) 発明者	辻井 訓
			東京都品川区北品川6丁目7番35号 ソニー株式会社内
		Fターム(参考)	5C053 FA14 GB36 GB37 JA30 LA01 LA11 5D044 AB05 AB07 CC04 DE17 DE31 DE49 EF05 FG18 GK12 HL14 JJ03 5D110 AA13 AA27 AA29 CA05 CA06 CA07 CA10 CJ11 DA11 DC16 DE01 FA02

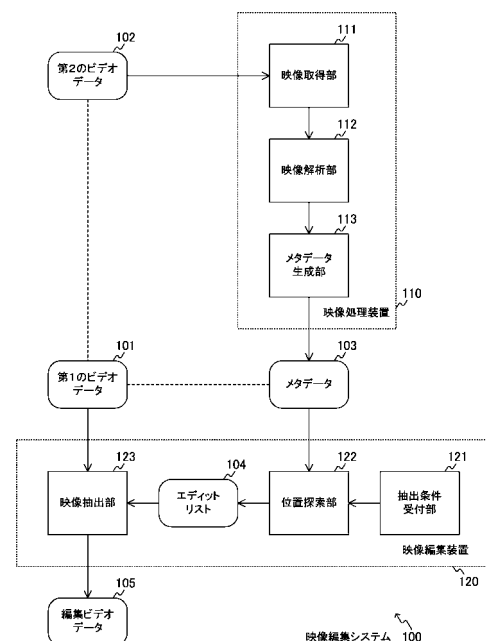
(54) 【発明の名称】 映像編集システム、映像処理装置、映像編集装置、映像処理方法、映像編集方法、プログラムおよびデータ構造

(57) 【要約】

【課題】映像データもしくはその映像データに同期する他の映像データについて、その特徴量を映像データに同期して保持するメタデータを生成し、または、そのメタデータに基づいて映像データを編集する。

【解決手段】第1のビデオデータ101に同期する第2のビデオデータ102から映像処理装置110によってメタデータ103が生成される。メタデータ103は、第1のビデオデータ101に同期し、第2のビデオデータ102の特徴量を時間軸で管理しながら保持する。メタデータ103において抽出条件に合致する位置が映像編集装置120の位置探索部122によって探索され、エディットリスト104が生成される。エディットリスト104に従って、第1のビデオデータ101から映像抽出部123によって映像が抽出され、編集ビデオデータ105が生成される。

【選択図】図1



【特許請求の範囲】**【請求項 1】**

時系列に管理される第 1 の映像データに同期した第 2 の映像データを取得する映像取得手段と、

前記第 2 の映像データにおける特徴量を解析する映像解析手段と、

前記第 1 の映像データと同期して前記特徴量を保持するメタデータを生成するメタデータ生成手段と

を具備することを特徴とする映像処理装置。

【請求項 2】

前記特徴量は、前記第 2 の映像データに含まれる顔の表情であることを特徴とする請求項 1 記載の映像処理装置。 10

【請求項 3】

前記第 1 の映像データを撮像する第 1 の撮像手段をさらに具備し、

前記映像取得手段は、前記第 1 の映像データの撮像と同時に前記第 2 の映像データを撮像する第 2 の撮像手段を含む

ことを特徴とする請求項 1 記載の映像処理装置。

【請求項 4】

前記第 1 の映像データを再生する再生手段をさらに具備し、

前記映像取得手段は、前記再生手段による前記第 1 の映像データの再生と同時に前記第 2 の映像データを撮像する撮像手段を含む 20

ことを特徴とする請求項 1 記載の映像処理装置。

【請求項 5】

前記第 1 の映像データを再生する再生手段をさらに具備し、

前記映像取得手段は、前記再生手段により再生された前記第 1 の映像データを前記第 2 の映像データとして入力する映像入力手段を含む

ことを特徴とする請求項 1 記載の映像処理装置。

【請求項 6】

前記第 1 の映像データおよび前記メタデータは、QuickTime フォーマットにおけるメディアデータアトムの形式で記録されることを特徴とする請求項 1 記載の映像処理装置。 30

【請求項 7】

時系列に管理される第 1 の映像データに同期した第 2 の映像データにおける特徴量を保持するメタデータを取得した後に所定の条件に合致する時系列上の位置を探索してその結果を探索情報として生成する位置探索手段と、

前記探索情報に基づいて前記第 1 の映像データから前記合致する時系列上の位置に対応する部分を抽出する映像抽出手段と

を具備することを特徴とする映像編集装置。

【請求項 8】

前記位置探索手段は、前記メタデータを管理する管理情報を取得して、前記管理情報が前記メタデータに前記特徴量は保持されない旨を示している場合には前記メタデータを取得しないことを特徴とする請求項 7 記載の映像編集装置。 40

【請求項 9】

時系列に管理される第 1 の映像データに同期した第 2 の映像データを取得する映像取得手段と、

前記第 2 の映像データにおける特徴量を解析する映像解析手段と、

前記第 1 の映像データと同期して前記特徴量を保持するメタデータを生成するメタデータ生成手段と、

前記メタデータから所定の条件に合致する時系列上の位置を探索してその結果を探索情報として生成する位置探索手段と、

前記探索情報に基づいて前記第 1 の映像データから前記合致する時系列上の位置に対応 50

する部分を抽出する映像抽出手段と
を具備することを特徴とする映像編集システム。

【請求項 10】

時系列に管理される第 1 の映像データに同期した第 2 の映像データを取得する映像取得手順と、

前記第 2 の映像データにおける特徴量を解析する映像解析手順と、

前記第 1 の映像データと同期して前記特徴量を保持するメタデータを生成するメタデータ生成手順と

を具備することを特徴とする映像処理方法。

【請求項 11】

時系列に管理される第 1 の映像データに同期した第 2 の映像データを取得する映像取得手順と、

前記第 2 の映像データにおける特徴量を解析する映像解析手順と、

前記第 1 の映像データと同期して前記特徴量を保持するメタデータを生成するメタデータ生成手順と

をコンピュータに実行させることを特徴とするプログラム。

【請求項 12】

時系列に管理される第 1 の映像データに同期した第 2 の映像データにおける特徴量を保持するメタデータを取得するメタデータ取得手順と、

前記メタデータにおいて所定の条件に合致する時系列上の位置を探索してその結果を探索情報として生成する位置探索手順と、

前記探索情報に基づいて前記第 1 の映像データから前記合致する時系列上の位置に対応する部分を抽出する映像抽出手順と

を具備することを特徴とする映像編集方法。

【請求項 13】

時系列に管理される第 1 の映像データに同期した第 2 の映像データにおける特徴量を保持するメタデータを取得するメタデータ取得手順と、

前記メタデータにおいて所定の条件に合致する時系列上の位置を探索してその結果を探索情報として生成する位置探索手順と、

前記探索情報に基づいて前記第 1 の映像データから前記合致する時系列上の位置に対応する部分を抽出する映像抽出手順と

をコンピュータに実行させることを特徴とするプログラム。

【請求項 14】

時系列に管理される第 1 の映像データと、

前記第 1 の映像データに同期した第 2 の映像データにおける特徴量を前記第 1 の映像データと同期して保持するメタデータと

を具備するデータ構造であって、

コンピュータが、前記メタデータを取得して、前記メタデータにおいて所定の条件に合致する時系列上の位置を探索してその結果を探索情報として生成して、前記探索情報に基づいて前記第 1 の映像データから前記合致する時系列上の位置に対応する部分を抽出することを特徴とするデータ構造。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、映像編集システムに関し、特に映像データにおける特徴量を示すメタデータを生成する映像編集システム、映像処理装置、または、そのメタデータに基づいて映像データを編集する映像編集装置、これらにおける処理方法ならびに当該方法をコンピュータに実行させるプログラムおよびこれらに用いられるデータ構造に関する。

【背景技術】

【0002】

10

20

30

40

50

記録再生装置における記録形式の一つとして、QuickTimeファイルフォーマットが知られている。このQuickTimeファイルフォーマットは、マルチメディアデータを扱うためのファイルフォーマットであり、映像データ（ビデオデータおよびオーディオデータ）の実データをメディアデータアトム（ムービーデータともいう。）に保持して、その管理情報をムービーアトム（ムービーリソースともいう。）に保持する。これにより、実データに直接手を加えることなく、映像データを"非破壊的に"編集できるようになっている。このQuickTimeファイルフォーマットをベースとしたファイルフォーマットには、ISOベースメディア（ISO Base Media）ファイルフォーマット、そのアプリケーションフォーマットであるMP4（MP4）ファイルフォーマット、MJ2（Motion JPEG2000）ファイルフォーマット、AVC（Advanced Video Coding：MPEG4-part10）ファイルフォーマットなどがある。

【0003】

このような実データ格納部と管理情報格納部に分かれた形式のファイルフォーマットにおいては、オリジナルデータを外部から参照して、再生上の時間軸管理を行うエディットアトムと呼ばれるデータ構造により編集を行う方法が知られている。例えば、そのようなエディットアトムを用いて、記録中の映像に対して非破壊的にマークを付与する映像記録装置が提案されている（例えば、特許文献1参照。）。

【特許文献1】特開2005-303943号公報（図5）

【発明の開示】

【発明が解決しようとする課題】

【0004】

しかしながら、エディットアトムにより構成されるムービーファイルでは、最終的な編集結果が1つだけ保存されるに留まり、他の条件で編集を行うためには最初から処理をやり直さなければならない。例えば、映像データもしくはその映像データに同期する他の映像データにおける何らかの特徴量を解析して、その解析結果に基づいて編集を行う場合、編集条件を変えるたびに特徴量の解析からやり直すことになり、処理効率上の問題がある。

【0005】

そこで、本発明は、映像データもしくはその映像データに同期する他の映像データについて、その特徴量を映像データに同期して保持するメタデータを生成し、または、そのメタデータに基づいて映像データを編集することを目的とする。

【課題を解決するための手段】

【0006】

本発明は、上記課題を解決するためになされたものであり、その第1の側面は、時系列に管理される第1の映像データに同期した第2の映像データを取得する映像取得手段と、上記第2の映像データにおける特徴量を解析する映像解析手段と、上記第1の映像データと同期して上記特徴量を保持するメタデータを生成するメタデータ生成手段とを具備することを特徴とする映像処理装置である。これにより、第1の映像データに同期した第2の映像データの特徴量を、第1の映像データに同期するメタデータに保持させるという作用をもたらす。

【0007】

また、この第1の側面において、上記特徴量は、上記第2の映像データに含まれる顔の表情であってもよい。顔の表情として、笑い、驚き、怒り、眠いといった種別を表現することができる。

【0008】

また、この第1の側面において、上記第1の映像データを撮像する第1の撮像手段をさらに具備し、上記映像取得手段は、上記第1の映像データの撮像と同時に上記第2の映像データを撮像する第2の撮像手段を含んでもよい。これにより、撮像されている第1の映像データに関する映像を第2の映像データとして撮像させるという作用をもたらす。

【0009】

10

20

30

40

50

また、この第１の側面において、上記第１の映像データを再生する再生手段をさらに具備し、上記映像取得手段は、上記再生手段による上記第１の映像データの再生と同時に上記第２の映像データを撮像する撮像手段を含んでもよい。これにより、再生されている第１の映像データに関する映像を第２の映像データとして撮像させるという作用をもたらす。

【００１０】

また、この第１の側面において、上記第１の映像データを再生する再生手段をさらに具備し、上記映像取得手段は、上記再生手段により再生された上記第１の映像データを上記第２の映像データとして入力する映像入力手段を含んでもよい。これにより、再生されている第１の映像データをそのまま第２の映像データとして入力させるという作用をもたらす。

10

【００１１】

なお、この第１の側面において、上記第１の映像データおよび上記メタデータは、QuickTimeフォーマットにおけるメディアデータアトムの形式で記録されてもよい。

【００１２】

また、本発明の第２の側面は、時系列に管理される第１の映像データに同期した第２の映像データにおける特徴量を保持するメタデータを取得した後に所定の条件に合致する時系列上の位置を探索してその結果を探索情報として生成する位置探索手段と、上記探索情報に基づいて上記第１の映像データから上記合致する時系列上の位置に対応する部分を抽出する映像抽出手段とを具備することを特徴とする映像編集装置である。これにより、メタデータに含まれる特徴量が所定の条件に合致する時系列上の位置について、第１の映像データの対応する部分を抽出して、非破壊的に編集させるという作用をもたらす。

20

【００１３】

また、この第２の側面において、上記位置探索手段は、上記メタデータを管理する管理情報を取得して、上記管理情報が上記メタデータに上記特徴量は保持されない旨を示している場合には上記メタデータを取得しないようにしてもよい。これにより、無意味なメタデータへのアクセスを回避させるという作用をもたらす。

【００１４】

また、本発明の第３の側面は、時系列に管理される第１の映像データに同期した第２の映像データを取得する映像取得手段と、上記第２の映像データにおける特徴量を解析する映像解析手段と、上記第１の映像データと同期して上記特徴量を保持するメタデータを生成するメタデータ生成手段と、上記メタデータから所定の条件に合致する時系列上の位置を探索してその結果を探索情報として生成する位置探索手段と、上記探索情報に基づいて上記第１の映像データから上記合致する時系列上の位置に対応する部分を抽出する映像抽出手段とを具備することを特徴とする映像編集システムである。これにより、第２の映像データの特徴量を保持するメタデータを中間状態として、第１の映像データを非破壊的に編集させるという作用をもたらす。

30

【００１５】

また、本発明の第４の側面は、時系列に管理される第１の映像データと、上記第１の映像データに同期した第２の映像データにおける特徴量を上記第１の映像データと同期して保持するメタデータとを具備するデータ構造であって、コンピュータが、上記メタデータを取得して、上記メタデータにおいて所定の条件に合致する時系列上の位置を探索してその結果を探索情報として生成して、上記探索情報に基づいて上記第１の映像データから上記合致する時系列上の位置に対応する部分を抽出することを特徴とするデータ構造である。これにより、第１の映像データの編集のための中間状態としてメタデータを保持させるという作用をもたらす。

40

【発明の効果】

【００１６】

本発明によれば、映像データもしくはその映像データに同期する他の映像データについて、その特徴量を映像データに同期して保持するメタデータを生成し、または、そのメタ

50

データに基づいて映像データを編集することができるという優れた効果を奏し得る。

【発明を実施するための最良の形態】

【0017】

次に本発明の実施の形態について図面を参照して詳細に説明する。

【0018】

図1は、本発明の実施の形態における映像編集システム100の一構成例を示す図である。この映像編集システム100は、メタデータ103に基づいて第1のビデオデータ101を編集した編集ビデオデータ105を出力する映像編集装置120と、第2のビデオデータ102の特徴量を解析してメタデータ103を出力する映像処理装置110とを備えている。

10

【0019】

第1のビデオデータ101および第2のビデオデータ102は、時系列に管理される映像データであり、動画像データに加えて音声データを含んでもよい。第1のビデオデータ101は、映像編集装置120における編集対象となる映像データである。第2のビデオデータ102は、第1のビデオデータ101に同期しており、第1のビデオデータ101の撮像または再生と同時に撮像され得る。また、第1のビデオデータ101および第2のビデオデータ102は、同一の内容であってもよい。

【0020】

メタデータ103は、第1のビデオデータ101に同期しており、第2のビデオデータ102の特徴量を時間軸で管理しながら保持するものである。特徴量としては、後述のように、第2のビデオデータ102に含まれる顔の表情が想定される。

20

【0021】

映像処理装置110は、映像取得部111と、映像解析部112と、メタデータ生成部113とを備えている。また、映像編集装置120は、抽出条件受付部121と、位置探索部122と、映像抽出部123とを備えている。

【0022】

映像取得部111は、第2のビデオデータ102を取得するものである。この映像取得部111は、映像を光学レンズにより撮像するビデオカメラであってもよく、また、電子信号を入力する入力端子などであってもよい。

【0023】

30

映像解析部112は、映像取得部111によって取得された第2のビデオデータ102における特徴量を解析するものである。映像解析部112は、例えば、特徴量として第2のビデオデータ102に含まれる顔の画像を抽出して、その顔の表情を判断する。顔の表情を判断する処理手順は公知の技術を使用することができる。例えば、顔画像中の外眼角点を結ぶ線分と口角点を結ぶ線分との比から笑顔の判断をする技術（例えば、特開2005-266984）や、顔の構成要素毎に基準画像と評価ポイントを設けて平均値を算出して被撮影者の表情を判断する技術（例えば、特開2004-46591）等が提案されている。

【0024】

40

メタデータ生成部113は、映像解析部112によって解析された特徴量を保持するメタデータ103を生成するものである。この生成されたメタデータ103は、第1のビデオデータ101の各時刻に対応して、第2のビデオデータ102の特徴量を保持している。これにより、例えば、第1のビデオデータ101の各時刻において、第2のビデオデータ102にどのような顔の表情が含まれているかを、編集のための中間状態として保持することができる。

【0025】

抽出条件受付部121は、抽出条件の入力を受け付けるものである。例えば、顔の表情を条件とするのであれば、笑顔を抽出するのか、または、驚いた顔を抽出するのか、といった条件を受け付ける。これらの条件は論理積（AND）や論理和（OR）などにより組み合わせ条件にすることができる。

50

【 0 0 2 6 】

位置探索部 1 2 2 は、抽出条件受付部 1 2 1 によって受け付けられた抽出条件によりメタデータ 1 0 3 を探索するものである。これにより、抽出条件に合致する時系列上の位置が得られ、エディットリスト 1 0 4 として保持される。

【 0 0 2 7 】

映像抽出部 1 2 3 は、エディットリスト 1 0 4 に基づいて、第 1 のビデオデータ 1 0 1 から抽出条件に合致する時系列上の位置に対応する部分を抽出して、編集ビデオデータ 1 0 5 に出力するものである。

【 0 0 2 8 】

図 2 は、本発明の実施の形態における映像処理装置 1 1 0 の第 1 の構成例を示す図である。この第 1 の構成例では、記録部 2 1 0 に、撮像部 2 1 1 と、映像加工部 2 1 2 と、映像圧縮部 2 1 3 と、ファイル生成部 2 1 4 と、書込み部 2 1 5 と、撮像部 2 1 6 と、映像解析部 2 1 7 と、記録制御部 2 1 8 と、記録媒体 2 1 9 とが備えられている。

10

【 0 0 2 9 】

撮像部 2 1 1 は、被写体を第 1 のビデオデータ 1 0 1 として撮像するものである。映像加工部 2 1 2 は、撮像部 2 1 1 によって撮像された映像に対してエフェクト処理などの加工を施すものである。映像圧縮部 2 1 3 は、映像加工部 2 1 2 によって加工された映像を圧縮するものである。

【 0 0 3 0 】

撮像部 2 1 6 は、被写体を撮像している撮影者（もしくは、被写体の様子を鑑賞している鑑賞者）の顔を第 2 のビデオデータ 1 0 2 として撮像するものである。映像解析部 2 1 7 は、撮像部 2 1 6 によって撮像された映像を解析するものである。すなわち、映像解析部 2 1 7 は、映像に含まれる撮影者の顔の表情を解析する。この解析結果は、メタデータ 1 0 3 となる。

20

【 0 0 3 1 】

ファイル生成部 2 1 4 は、映像圧縮部 2 1 3 によって圧縮された映像（第 1 のビデオデータ 1 0 1 ）および映像解析部 2 1 7 によって解析された撮影者の顔の表情の解析結果（メタデータ 1 0 3 ）を含むファイルをそれぞれ所定のファイル形式として生成するものである。

【 0 0 3 2 】

書込み部 2 1 5 は、ファイル生成部 2 1 4 によって生成されたファイルを記録媒体 2 1 9 に書き込むものである。記録媒体 2 1 9 としては、ハードディスクなどのディスク状記録媒体やメモリスティックなどの半導体記録媒体を想定することができる。

30

【 0 0 3 3 】

記録制御部 2 1 8 は、記録部 2 1 0 における記録媒体 2 1 9 への記録動作を制御するものである。

【 0 0 3 4 】

この第 1 の構成例では、撮像部 2 1 1 によって被写体を第 1 のビデオデータ 1 0 1 として撮像しながら、同時に撮像部 2 1 6 によって撮影者の顔の表情を第 2 のビデオデータ 1 0 2 として撮像し、その表情を解析してメタデータ 1 0 3 を生成している。

40

【 0 0 3 5 】

図 3 は、本発明の実施の形態における映像処理装置 1 1 0 の第 1 の構成例による使用態様を示す図である。図 3 (a) では、ビデオカメラ装置 5 2 0 の前面のカメラ 5 2 1 によって被写体 5 0 1 が撮像されているのと同時に、ビデオカメラ装置 5 2 0 の操作面のカメラ 5 2 2 によって撮影者 5 0 2 の顔が撮像されている。これにより、ビデオカメラ装置 5 2 0 の内部で撮影者 5 0 2 の顔の表情を解析することによって、メタデータを生成することができる。

【 0 0 3 6 】

また、ビデオカメラ装置 5 2 0 はネットワーク 5 1 0 に接続されてもよい。このネットワーク 5 1 0 によって、ビデオカメラ装置 5 2 0 の前面のカメラ 5 2 1 によって撮像され

50

た被写体 5 0 1 の映像を、図 3 (b) のテレビ装置 5 3 0 や図 3 (c) のコンピュータ装置 5 4 0 に配信することができる。

【 0 0 3 7 】

図 3 (b) のテレビ装置 5 3 0 にはビデオカメラ装置 5 3 1 が接続され、鑑賞者 5 0 3 の顔が撮像されている。これにより、テレビ装置 5 3 0 またはビデオカメラ装置 5 3 1 の内部で鑑賞者 5 0 3 の顔の表情を解析することによって、メタデータを生成することができる。

【 0 0 3 8 】

図 3 (c) のコンピュータ装置 5 4 0 の前面にはカメラ 5 4 1 が設けられ、鑑賞者 5 0 4 の顔が撮像されている。これにより、コンピュータ装置 5 4 0 の内部で鑑賞者 5 0 4 の顔の表情を解析することによって、メタデータを生成することができる。

10

【 0 0 3 9 】

このように、映像処理装置 1 1 0 の第 1 の構成例では、第 1 のビデオデータ 1 0 1 の撮像と同時に第 2 のビデオデータ 1 0 2 を撮像して、この第 2 のビデオデータ 1 0 2 からメタデータ 1 0 3 を生成する。第 2 のビデオデータ 1 0 2 に含まれる顔の表情には、第 1 のビデオデータ 1 0 1 の映像に対する何らかの反応が反映されているものと考えられるため、その特徴量をメタデータ 1 0 3 に中間状態として保存しておいて、第 1 のビデオデータ 1 0 1 の編集に利用しようとするものである。

【 0 0 4 0 】

図 4 は、本発明の実施の形態における映像処理装置 1 1 0 の第 2 の構成例を示す図である。この第 2 の構成例では、記録部 2 1 0 に、撮像部 2 1 1 と、ファイル生成部 2 1 4 と、書込み部 2 1 5 と、映像解析部 2 1 7 と、記録制御部 2 1 8 と、記録媒体 2 1 9 とが備えられている。また、再生部 2 2 0 に、表示部 2 2 1 と、映像加工部 2 2 2 と、映像伸張部 2 2 3 と、ファイル復号部 2 2 4 と、読出し部 2 2 5 と、記録媒体 2 2 9 とが備えられている。

20

【 0 0 4 1 】

記録媒体 2 2 9 は、第 1 のビデオデータ 1 0 1 を所定のファイル形式により記録するものである。読出し部 2 2 5 は、第 1 のビデオデータ 1 0 1 を含むファイルを記録媒体 2 2 9 から読み出すものである。ファイル復号部 2 2 4 は、読出し部 2 2 5 によって読み出されたファイルを復号するものである。映像伸張部 2 2 3 は、ファイル復号部 2 2 4 によって復号されたファイル内の圧縮された映像を伸張するものである。映像加工部 2 2 2 は、映像伸張部 2 2 3 によって伸張された映像に対してエフェクト処理などの加工を施すものである。

30

【 0 0 4 2 】

表示部 2 2 1 は、映像加工部 2 2 2 から出力された映像を表示するものである。これにより、記録媒体 2 2 9 に記録されていた第 1 のビデオデータ 1 0 1 が表示部 2 2 1 に再生表示される。

【 0 0 4 3 】

撮像部 2 1 1 は、被写体を第 2 のビデオデータ 1 0 2 として撮像するものである。映像解析部 2 1 7 は、撮像部 2 1 1 によって撮像された映像を解析するものである。すなわち、映像解析部 2 1 7 は、映像に含まれる鑑賞者の顔の表情を解析する。この解析結果は、メタデータ 1 0 3 となる。

40

【 0 0 4 4 】

ファイル生成部 2 1 4 は、映像解析部 2 1 7 によって解析された鑑賞者の顔の表情の解析結果 (メタデータ 1 0 3) を含むファイルを所定のファイル形式として生成するものである。

【 0 0 4 5 】

書込み部 2 1 5 は、ファイル生成部 2 1 4 によって生成されたファイルを記録媒体 2 1 9 に書き込むものである。記録制御部 2 1 8 は、記録部 2 1 0 における記録媒体 2 1 9 への記録動作を制御するものである。

50

【 0 0 4 6 】

なお、記録媒体 2 1 9 および 2 2 9 としては、ハードディスクなどのディスク状記録媒体やメモリスティックなどの半導体記録媒体を想定することができるが、両者は互いに異なる種類の記録媒体であってもよい。

【 0 0 4 7 】

この第 2 の構成例では、表示部 2 2 1 によって第 1 のビデオデータ 1 0 1 を再生表示しながら、同時に撮像部 2 1 1 によって鑑賞者の顔の表情を第 2 のビデオデータ 1 0 2 として撮像し、その表情を解析してメタデータ 1 0 3 を生成している。

【 0 0 4 8 】

図 5 は、本発明の実施の形態における映像処理装置 1 1 0 の第 2 の構成例による使用態様を示す図である。図 2 (a) では、ビデオカメラ装置 5 2 0 の操作面に再生表示画面が表示されており、その操作面のカメラ 5 2 2 によって鑑賞者 5 0 5 の顔が撮像されている。これにより、ビデオカメラ装置 5 2 0 の内部で鑑賞者 5 0 5 の顔の表情を解析することによって、メタデータを生成することができる。

【 0 0 4 9 】

図 5 (b) のテレビ装置 5 3 0 にはビデオカメラ装置 5 3 1 が接続され、鑑賞者 5 0 3 の顔が撮像されている。これにより、テレビ装置 5 3 0 またはビデオカメラ装置 5 3 1 の内部で鑑賞者 5 0 3 の顔の表情を解析することによって、メタデータを生成することができる。

【 0 0 5 0 】

図 5 (c) のコンピュータ装置 5 4 0 の前面にはカメラ 5 4 1 が設けられ、鑑賞者 5 0 4 の顔が撮像されている。これにより、コンピュータ装置 5 4 0 の内部で鑑賞者 5 0 4 の顔の表情を解析することによって、メタデータを生成することができる。

【 0 0 5 1 】

このように、映像処理装置 1 1 0 の第 2 の構成例では、第 1 のビデオデータ 1 0 1 の再生と同時に第 2 のビデオデータ 1 0 2 を撮像して、この第 2 のビデオデータ 1 0 2 からメタデータ 1 0 3 を生成する。第 2 のビデオデータ 1 0 2 に含まれる顔の表情には、第 1 のビデオデータ 1 0 1 の映像に対する何らかの反応が反映されているものと考えられるため、その特徴量をメタデータ 1 0 3 に中間状態として保存しておいて、第 1 のビデオデータ 1 0 1 の編集に利用しようとするものである。

【 0 0 5 2 】

図 6 は、本発明の実施の形態における映像処理装置 1 1 0 の第 3 の構成例を示す図である。この第 3 の構成例では、記録部 2 1 0 に、映像入力部 2 0 6 と、ファイル生成部 2 1 4 と、書込み部 2 1 5 と、映像解析部 2 1 7 と、記録制御部 2 1 8 と、記録媒体 2 1 9 とが備えられている。また、再生部 2 2 0 に、映像加工部 2 2 2 と、映像伸張部 2 2 3 と、ファイル復号部 2 2 4 と、読出し部 2 2 5 と、記録媒体 2 2 9 とが備えられている。

【 0 0 5 3 】

再生部 2 2 0 の構成は、第 2 の構成例の場合と同様であるが、表示部 2 2 1 は省かれ、映像加工部 2 2 2 の出力が記録部 2 1 0 にそのまま供給されている点異なる。

【 0 0 5 4 】

記録部 2 1 0 の構成も、第 2 の構成例の場合と同様であるが、撮像部 2 1 1 の代わりに、再生部 2 2 0 からの映像を入力する映像入力部 2 0 6 を備える点異なる。

【 0 0 5 5 】

すなわち、この第 3 の構成例では、再生部 2 2 0 によって再生された映像 (第 1 のビデオデータ 1 0 1) をそのまま記録部 2 1 0 の入力映像 (第 2 のビデオデータ 1 0 2) として供給して、それに含まれる顔の表情を解析してメタデータ 1 0 3 を生成している。

【 0 0 5 6 】

図 7 は、本発明の実施の形態における映像処理装置 1 1 0 の一実施例であるカメラ一体型撮像装置の構成例を示す図である。この撮像装置は、撮像部 3 0 1 と、映像加工部 3 3 0 と、映像圧縮部 3 4 1 と、圧縮制御部 3 4 2 と、記録媒体アクセス部 3 5 1 と、ドライ

10

20

30

40

50

ブ制御部 352 と、操作受付部 360 と、表示部 370 と、システム制御部 390 とを備えている。

【0057】

撮像部 301 は、被写体を撮像して映像データとして出力するものである。映像加工部 330 は、撮像部 301 から出力された映像データにエフェクト処理を施すものである。映像圧縮部 341 は、映像加工部 330 によって加工された映像データを圧縮するものである。圧縮制御部 342 は、映像圧縮部 341 における圧縮処理の制御を行うものである。

【0058】

記録媒体アクセス部 351 は、記録媒体 309 に対する書込みや読出しを行うものである。ドライブ制御部 352 は、記録媒体アクセス部 351 による書込みや読出しを制御するものである。

【0059】

操作受付部 360 は、ユーザによる操作入力を受け付けるものであり、各種ボタンや GUI (Graphical User Interface) などが想定される。表示部 370 は、撮像中の映像や再生映像、または、ユーザに対する各種メッセージなどを表示するものである。

【0060】

システム制御部 390 は、撮像装置の全体を制御するものであり、例えば、マイクロプロセッサなどにより実現され得る。このシステム制御部 390 は、操作受付部 360 によって受け付けられた操作入力によって映像の録画の開始、停止や、録画の経過時間情報などを制御するとともに、ユーザに対する表示部 370 における表示を制御する。また、システム制御部 390 は、カメラ制御部 329 や圧縮制御部 342 との間で情報をやり取りして、ドライブ制御部 352 を介して記録媒体 309 に対する書込み制御を行う。

【0061】

また、撮像部 301 は、ズームレンズ 311 と、アイリス (絞り) 312 と、フォーカスレンズ 313 と、フィルタ 314 と、撮像素子 321 と、A/D 変換器 322 と、カメラ信号処理回路 323 と、検波部 324 と、ズーム制御部 325 と、角速度センサ 326 と、カメラ制御部 329 とを備えている。

【0062】

ズームレンズ 311 は、ズーム (拡大) 処理を行うためのレンズである。アイリス 312 は、被写体からの光量を調整するための絞りである。フォーカスレンズ 313 は、被写体に焦点を合わせるためのレンズである。フィルタ 314 は、赤外線を除去するためのフィルタである。

【0063】

撮像素子 321 は、光学レンズ群から供給された光を電気信号に変換する光電変換素子であり、例えば、CCD (Charge Coupled Devices) などにより実現され得る。この撮像素子 321 により、被写体の画像が、例えば RGB (赤、緑、青) の 3 原色に相当する 3 つの映像信号として取り出される。

【0064】

A/D 変換器 322 は、撮像素子 321 から供給されたアナログの電気信号をデジタル信号に変換するものである。カメラ信号処理回路 323 は、A/D 変換器 322 により変換されたデジタル信号に対して、白色の基準を定めるホワイトバランスなどの信号処理を施すものである。

【0065】

検波部 324 は、カメラ信号処理回路 323 によって信号処理の施された映像信号のフィードバックを受けて、各種の検波処理を行うものである。例えば、自動的に被写体に焦点を合わせるためのオートフォーカス (AF: Auto Focus) 検波、自動的に露光を行うためのオートエクスポージャ (AE: Auto Exposure) 検波、自動的にホワイトバランスを行うためのオートホワイトバランス (AWB: Auto White Balance) 検波などを行うものである。

10

20

30

40

50

【 0 0 6 6 】

ズーム制御部 3 2 5 は、ユーザからの操作入力などに従ってズームレンズ 3 1 1 を移動させてズーム処理を制御するものである。角速度センサ 3 2 6 は、撮像装置の角速度を検出するものであり、例えば、ジャイロ스코プなどにより手ぶれの度合いを検出するものである。

【 0 0 6 7 】

カメラ制御部 3 2 9 は、撮像部 3 0 1 の制御を行うものである。例えば、カメラ制御部 3 2 9 は、角速度センサ 3 2 6 において検知された手ぶれに対して手ぶれ補正を行って画質の劣化を低減するように制御を行う。また、カメラ制御部 3 2 9 は、撮像素子 3 2 1 からの映像入力の制御、検波部 3 2 4 における処理の制御、ズーム制御部 3 2 5 における処理の制御などを行う。

10

【 0 0 6 8 】

図 8 は、Quick Time ファイルフォーマットをベースとしたファイル形式（以下、Quick Time ベースファイル形式という。）の構造例を示す図である。このファイル形式では、ファイルの内容が実データ格納部と、その実データを参照するために必要な場所情報などを格納する管理情報格納部とに分かれている。Quick Time ファイルフォーマットでは、実データ格納部はメディアデータアトム（media data atom、タイプ名：'m d a t'）と呼ばれ、管理情報格納部はムービーアトム（movie atom、タイプ名：'m o o v'）と呼ばれる。なお、「アトム（atom）」は「ボックス（box）」と表現されることもある。また、ムービーアトムはムービーリソースと表現されることがあり、メディアデータアトムは単にメディアデータまたはムービーデータと表現されることがある。

20

【 0 0 6 9 】

これらメディアデータアトムおよびムービーアトムは、同一のファイルに含まれていてもよく、別ファイルに分かれていてもよい。例えば、図 8 のように、動画像（V 1 等）や音声（A 1 等）のメディアデータを含むメディアデータアトム 6 1 2 と、それを参照するムービーアトム 6 1 1 とを同一のファイル 6 1 0 に格納するようにしてもよく、また、メディアデータアトム 6 1 2 を参照するムービーアトム 6 2 1 を別のファイル 6 2 0 に格納するようにしてもよい。前者の形式を有するファイルは自己内包型ファイルと呼ばれ、後者の形式を有するファイルは外部参照型ファイルと呼ばれる。そのため、ムービーアトムは、外部参照するメディアデータアトムが含まれる外部ファイルの相対パスまたは絶対パスを示す管理情報を格納できるようになっている。

30

【 0 0 7 0 】

メディアデータアトム 6 1 2 には、例えば M P E G 1 オーディオ（MPEG1 Audio Layer2）に基づく圧縮符号化方式によって符号化されたオーディオデータおよび M P E G 2 ビデオ（MPEG2 Video）規定に従う圧縮符号化方式によって符号化された画像データが格納される。符号化方式はこれらに限定されるものではなく、例えば、ビデオデータであれば Motion J P E G（Motion JPEG）や M J 2（Motion JPEG2000）、M P E G 4（M P 4）や A V C（Advanced Video Coding：MPEG4-part10）、オーディオデータであればドルビー A C 3（Dolby AC3）や A T R A C（Adaptive TRansform Acoustic Coding）などでもよく、また、圧縮符号化が施されていないリニアデータを格納することも可能である。

40

【 0 0 7 1 】

図 9 は、Quick Time ファイルフォーマットにおける階層構造を示す図である。メディアデータアトム（'m d a t'）における実データはサンプル（sample）と呼ばれる最小管理単位に分かれており、このサンプルを任意の個数分集めたものがチャンク（chunk）と呼ばれる。メディアデータアトム（'m d a t'）の管理情報であるムービーアトム（'m o o v'）では、サンプルのサイズや、チャンクの先頭格納場所、各サンプルの表示時間等が格納される。

【 0 0 7 2 】

ムービーアトム（'m o o v'）は、ムービーヘッダアトム（'m v h d'）と、トラック

50

アトム ('t r a k') 等から構成される。

【 0 0 7 3 】

ムービーヘッダアトム ('m v h d') は、ムービーアトムのヘッダ情報を保持する部分であり、ムービー全体の特徴を示すものである。例えば、ムービー全体の期間や時間スケール、作成日等を項目として含む。

【 0 0 7 4 】

トラックアトム ('t r a k') は、サウンド、ビデオ、テキストといった異なるタイプのデータをそれぞれ別のトラックにより格納するものであり、この図では、ビデオのトラックアトムとして、トラックヘッダアトム ('t k h d') と、エディットアトム ('e d t s') と、ユーザデータアトム ('u d t a') と、メディアアトム ('m d i a') とを含んで構成される。また、オーディオのトラックアトムについては省略されているが、ビデオの場合と同様の構成を備えて構成される。

【 0 0 7 5 】

トラックヘッダアトム ('t k h d') は、トラックアトムのヘッダ情報を保持する部分であり、そのトラックの特徴を示すものである。例えば、ビデオのピクセル数やサウンドの音量、作成日等を項目として含む。

【 0 0 7 6 】

エディットアトム ('e d t s') は、トラックの編集情報をエディットリストアトム ('e l s t') として保持するものである。なお、このエディットアトムについては、図 27 により詳述する。

【 0 0 7 7 】

ユーザデータアトム ('u d t a') は、必要に応じてユーザにより定義された任意の情報を含むものである。例えば、ムービーのウィンドウ位置や再生方法、作成情報等を保持することができる。このユーザデータアトムは、ムービーユーザデータをリスト形式により保持する。

【 0 0 7 8 】

メディアアトム ('m d i a') は、そのトラックで実際に用いられる実データに関する情報を格納するものである。すなわち、メディアアトムは、メディア全体に関する情報、メディアデータの取扱いに関する情報、メディアの構成に関する情報等を格納する。実データはサンプル (sample) と呼ばれる最小管理単位に分かれており、このサンプルを任意の個数分集めたものがチャンク (chunk) と呼ばれる。メディアアトムでは、サンプルのサイズや、チャンクの先頭格納場所、各サンプルの表示時間等が格納される。

【 0 0 7 9 】

このメディアアトムは、メディアヘッダアトム ('m d h d') と、メディアハンドラアトム ('h d l r') と、メディア情報アトム ('m i n f') 等から構成される。

【 0 0 8 0 】

メディアヘッダアトム ('m d h d') は、メディアアトムのヘッダ情報を保持する部分であり、メディア全体としての特徴を示すものである。

【 0 0 8 1 】

メディアハンドラアトム ('h d l r') は、メディア毎の取り扱いに関する情報を保持するものである。

【 0 0 8 2 】

メディア情報アトム ('m i n f') は、そのメディアタイプで表現される情報を保持するものである。このメディア情報アトムは、ビデオメディア情報ヘッダアトム ('v m h d') と、データハンドラアトム ('h d l r') と、データ情報アトム ('d i n f') と、サンプルテーブルアトム ('s t b l') 等から構成される。

【 0 0 8 3 】

ビデオメディア情報ヘッダアトム ('v m h d') は、ビデオトラックにおいて、ビデオメディアに関するヘッダ情報を保持するものである。なお、オーディオトラックの場合、サウンドメディアに関するヘッダ情報を保持するサウンドメディア情報ヘッダアトム ('

10

20

30

40

50

s m h d') が、ビデオメディアヘッダアトム (' v m h d') の代わりに含まれる。

【 0 0 8 4 】

データハンドラアトム (' h d l r') は、ビデオメディアの取り扱いに関する情報を保持するものである。

【 0 0 8 5 】

データ情報アトム (' d i n f') は、実際に参照する実データの格納先に関する情報を保持するものである。このデータ情報アトムには、参照する実データの格納方法、格納場所、ファイル名に関する情報を保持するデータリファレンスアトム (' d r e f') が含まれる。

【 0 0 8 6 】

サンプルテーブルアトム (' s t b l') は、そのメディアの実データの最小管理単位であるサンプルに関する情報を保持するものである。このサンプルテーブルアトムは、サンプルディスクリプションアトム (' s t s d') と、時間対サンプルアトム (' s t s s') と、サンプルサイズアトム (' s t s z') と、サンプル対チャンクアトム (' s t s c') と、チャンクオフセットアトム (' s t c o') 等から構成される。

【 0 0 8 7 】

サンプルディスクリプションアトム (' s t s d') は、各サンプルに関する圧縮方式やその特性に関する情報を保持するものである。時間対サンプルアトム (' s t s s') は、各サンプルと時間との関係性を保持するものである。サンプルサイズアトム (' s t s z') は、各サンプルのデータ量を保持するものである。サンプル対チャンクアトム (' s t s c') は、チャンクとそのチャンクを構成するサンプルの関係性を保持するものである。チャンクオフセットアトム (' s t c o') は、ファイル先頭からの各チャンクの先頭位置までのオフセットを保持するものである。

【 0 0 8 8 】

本発明の実施の形態では、第 1 のビデオデータ 1 0 1 を Q u i c k T i m e ベースファイル形式により保持するのみならず、メタデータ 1 0 3 もこの Q u i c k T i m e ベースファイル形式により保持する。これにより、第 2 のビデオデータ 1 0 2 の特徴量を時間軸で管理しながらメタデータ 1 0 3 に保持することができる。

【 0 0 8 9 】

図 1 0 は、本発明の実施の形態におけるファイルの保存形式の一例を示す図である。

【 0 0 9 0 】

図 1 0 (a) は、第 1 のビデオデータ 1 0 1 を含むビデオファイル 6 3 0 の構成例を示す図である。このビデオファイル 6 3 0 は、第 1 のビデオデータ 1 0 1 を有するメディアデータ 6 3 2 と、メディアデータ 6 3 2 を管理するムービーリソース 6 3 1 とを備えている。メディアデータ 6 3 2 は、第 1 のビデオデータ 1 0 1 の各サンプル 6 3 3 を含んでおり、これらはムービーリソース 6 3 1 によって管理される。

【 0 0 9 1 】

図 1 0 (b) は、メタデータ 1 0 3 を含むメタファイル 6 4 0 の構成例を示す図である。このメタファイル 6 4 0 は、メタデータ 1 0 3 を有するメディアデータ 6 4 2 と、メディアデータ 6 4 2 を管理するムービーリソース 6 4 1 とを備えている。メディアデータ 6 4 2 は、メタデータ 1 0 3 の各サンプル 6 4 3 を含んでおり、これらはムービーリソース 6 4 1 によって管理される。また、ムービーリソース 6 4 1 は、メディアデータ 6 3 2 も外部参照する形式で、同様に一つの時間軸によって管理している。

【 0 0 9 2 】

図 1 0 (c) は、ビデオファイル 6 3 0 およびメタファイル 6 4 0 によって実現されるビデオトラック 6 5 0 およびメタトラック 6 6 0 の時間軸上の流れを示す図である。ここでは、簡略化のため、ビデオファイル 6 3 0 の映像信号および音声信号のうち、ビデオトラック 6 5 0 のみを図示している。

【 0 0 9 3 】

ビデオトラック 6 5 0 では、各サンプル 6 5 3 が時間軸上に並んでいる。また、メタト

10

20

30

40

50

トラック 660 では、ビデオトラック 650 のサンプル 653 と同期して、各サンプル 663 が時間軸上に並んでいる。例えば、第 2 のビデオデータ 102 の時刻 t_1 から t_2 の区間において笑顔が特徴量として抽出された場合、その旨がメタトラック 660 の時刻 t_1 から t_2 の区間において記録される。同様に、第 2 のビデオデータ 102 の時刻 t_3 から t_4 の区間において驚いた顔が特徴量として抽出された場合、その旨がメタトラック 660 の時刻 t_3 から t_4 の区間において記録される。

【0094】

すなわち、メタデータ 103 を示すメタトラック 660 は、第 2 のビデオデータ 102 を介して、第 1 のビデオデータ 101 を示すビデオトラック 650 と同期していることになる。

10

【0095】

なお、図中では省略されているが、顔の表情の分類上、無表情である区間においては、無表情である旨を示す情報がメタトラック 660 に記録される。

【0096】

図 11 は、本発明の実施の形態におけるメタトラックの階層構造の一例を示す図である。このメタトラックは、図 9 で説明したビデオトラックと基本的に同様の構成を有している。但し、トラック配下にトラックリファレンスアトム ('t r e f') を有し、メディアアトム配下にトラックインプットマップアトム ('i m a p') を有する点でビデオトラックとは異なっている。

20

【0097】

トラックリファレンスアトム ('t r e f') は、ソーストラック (第 1 のビデオデータ 101) との参照関係を指定するための情報を保持するものである。そのため、トラックリファレンスアトムは、指定対象となるトラックのトラックヘッダアトム ('t k h d') に格納されているトラック固有のトラック ID を指定するトラックリファレンスタイプアトム ('s s r c') を含む。このトラックリファレンスタイプアトムに含まれるトラック ID の数は、ソーストラックの数と一致する。

【0098】

トラックインプットマップアトム ('i m a p') は、ソーストラックに関する情報を保持するものであり、Quick Time における QT アトム構造と呼ばれるデータ構造により構成される。このトラックインプットマップアトムには、QT アトムコンテナ ('s e a n') を最上位アトムとするコンテナによってバッキングされたトラックインプット QT アトム (' i n') が 1 つ以上含まれる。このトラックインプット QT アトムの数は、ソーストラックの数と一致する。

30

【0099】

トラックインプット QT アトム (' i n') は、インプットタイプ QT アトム (' t y') およびデータソースタイプ QT アトム ('d t s t') を保持する。インプットタイプ QT アトム (' t y') は、ソーストラックがビデオメディアであることを指定するものである。また、データソースタイプ QT アトム ('d t s t') は、ソーストラックに対して固有の名称を与えるものである。

40

【0100】

図 12 は、本発明の実施の形態におけるメタトラックを含むムービーアトムの記載例を示す図である。この記載例では、ソーストラック (第 1 のビデオデータ 101) としてビデオトラックアトム (v i d e o) が 1 つだけ含まれている。そのため、メタトラックアトムにおけるトラックインプット QT アトム (' i n') も 1 つだけ含まれている。

【0101】

図 13 は、本発明の実施の形態におけるメタトラックを含むムービーアトムの他の記載例を示す図である。また、図 14 は、図 13 のムービーアトムにおけるメタトラックアトムの記載例を示す図である。この記載例では、ソーストラック (第 1 のビデオデータ 101) としてビデオトラックアトムが 2 つ含まれている (v i d e o 1 および v i d e o 2)。そのため、図 14 に示すように、メタトラックアトムにおけるトラックインプット

50

Q Tアトム (' i n') も2つ含まれている。

【0102】

図15は、本発明の実施の形態におけるメタトラックのサンプルディスクリプションアトム ('s t s d') の記載例を示す図である。この記載例では、メタサンプルディスクリプションエントリがM個 (Mは1以上の整数) 含まれている。このメタサンプルディスクリプションエントリは、特徴量の種類の数と一致する。例えば、特徴量の種類として、無表情の顔と笑顔の2種類の分類をするのであれば、メタサンプルディスクリプションエントリは2つになる。

【0103】

なお、同図において、カッコ内の数字は各フィールドのバイト数を表す。

10

【0104】

メタサンプルディスクリプションエントリは、Q u i c k T i m eにおけるサンプルディスクリプションエントリに対してストリームディスクリプターアトムを拡張追加した構造になっている。サンプルディスクリプションエントリにおけるデータフォーマット (Data Format) フィールドは、本来、エフェクト効果を指定するためのものである。本発明の実施の形態では、このフィールドを拡張のために用いている。これにより、通常のQ u i c k T i m eファイルフォーマットとの間で互換性を維持しながら、拡張を施すことができる。

【0105】

図16は、本発明の実施の形態におけるデータフォーマットフィールドの一例を示す図である。この図に示すように、データフォーマットフィールドは、本来、エフェクト効果を指定するためのものである。同図において、アルファベット小文字で示している種別は、Q u i c k T i m eにおいて定義済のエフェクト種別である。例えば、タイプ名 'b r c o' は、明るさを示すブライトネス (brightness) と画像における黒色および白色の幅を示すコントラスト (contrast) とを変化させる効果を指定するものである。

20

【0106】

一方、アルファベット大文字で示している種別はQ u i c k T i m eにおいて定義されていないエフェクト種別である。本発明の実施の形態では、同図最下欄にあるユーザ定義のメタデータであることを示すタイプ名 'U D E F' をこのデータフォーマットフィールドで指定することによって、メタデータとして独自拡張された意味を有することを示している。

30

【0107】

図17は、本発明の実施の形態におけるストリームディスクリプターアトム ('s t r d') の記載例を示す図である。このストリームディスクリプターアトムは、Q u i c k T i m eにおける他のアトム構造と同様に、サイズ (Size)、タイプ (Type)、バージョン (Version) およびフラグ群 (Flags) の各フィールドを保持している。

【0108】

サイズフィールドは、このサイズフィールドを含むストリームディスクリプターアトム全体の大きさを保持するものである。タイプフィールドは、ストリームディスクリプターアトムのタイプ名として 's t r d' を保持するものである。バージョンフィールドおよびフラグ群フィールドは、将来の拡張用に確保されているものであり、ここでは全てゼロが設定されるものとする。

40

【0109】

ストリームディスクリプターアトム ('s t r d') は、以下に説明するように、データフォーマット (Data Format)、ユーザデファインドメタタイプ (User Defined Meta Type)、パラメータフラグ (Parameter Flag) の3つのフィールドをさらに保持している。

【0110】

データフォーマットフィールドは、図16により説明したメタサンプルディスクリプションエントリのデータフォーマットフィールドと形式上同じものを保持するフィールドであり、本発明の実施の形態ではタイプ名 'U D E F' を示すことになる。

50

【 0 1 1 1 】

ユーザデファインドメタタイプフィールドは、図 1 8 に示すように、2 バイトのオーナー ID (Owner ID) と 2 バイトのメタ ID (Meta ID) とを保持している。オーナー ID は、メーカー毎に割り当てられた ID であり、これにより、各メーカーは、メタ ID によって独自の拡張定義を用いることができるようになる。これらオーナー ID およびメタ ID は、データフォーマットフィールドがタイプ名 'U D E F' を示す場合にのみ有効になるものである。例えば、メタ ID として、図 1 9 に示すように、顔の表情として、笑い (smile)、驚き (surprise)、怒り (angry)、眠い (sleepy) といった種別を表現することができる。

【 0 1 1 2 】

なお、オーナー ID およびメタ ID の 2 つのフィールドに分けて詳細メタデータ種別を定義するのは、メーカー内で閉じて、重複なくメタデータ種別の管理運用を行うためである。仮に、このような区別を設けないと、新規に定義したいメーカー間で名称が重複し、もしくは、申請順番などの管理が煩雑になるおそれがある。そこで、メタサンプルディスクリプションエントリ側のデータフォーマットフィールドにて指定する独自定義メタデータ種別には、'U D E F' として独自定義メタデータ大別であることだけを指示し、詳細な独自拡張定義メタデータ種別はオーナー ID およびメタ ID の両フィールドを組み合わせることによって、どのメーカーが定義したどのようなメタデータ種別かを指示するものである。

【 0 1 1 3 】

パラメータフラグフィールドは、メタデータによる効果が有効であるか否かを示すフィールドである。例えば、図 2 0 に示すように、1 6 ビットのうちの 1 ビットを用いて、メタデータによる効果が「有効」であるか「無効」であるかを示す。これにより、ムービーリソース ('m o o v') におけるパラメータフラグフィールドが「無効」を示している場合には、メディアデータ ('m d a t') をアクセスするまでもなく、メタデータによる効果がないものと判断することができ、処理の負荷を低減することができる。

【 0 1 1 4 】

図 2 1 は、本発明の実施の形態におけるメタトラックのメディアデータ (メタデータ 1 0 3) のサンプルの記載例を示す図である。メタトラックのサンプルは、ビデオトラックのサンプルと同様に、サンプル毎にメディアデータアトムに格納される。

【 0 1 1 5 】

ここでは、第 2 のビデオデータ 1 0 2 の対応するサンプルにおいて抽出された顔の数 (face_number) と、それぞれの顔についてその表情の度合いを示すことができるようになっている。例えば、笑顔度合い、驚き度合い、怒り度合い、眠さ度合いをそれぞれ示すことができる。

【 0 1 1 6 】

図 2 2 は、本発明の実施の形態におけるメタデータを含むファイルの階層構造の一例を示す図である。この例では、ソーストラック (第 1 のビデオデータ 1 0 1) が 1 つ (ソーストラック 1) だけであることが想定されている。このソーストラック 1 のムービーリソース ('m o o v') のトラックヘッダアトム ('t k h d') には、そのトラック ID である「# 1」が保持されている。また、メタトラック (メタデータ 1 0 3) のムービーリソースのトラックヘッダアトムには、そのトラック ID である「# 2」が保持されている。

【 0 1 1 7 】

メタトラックのムービーリソースでは、トラックリファレンスアトム ('t r e f') のトラックリファレンスタイプアトム ('s s r c') に、ソーストラック 1 のトラック ID 「# 1」が保持されている。

【 0 1 1 8 】

また、メディアアトム ('m d i a') のトラックインプットマップアトム ('i m a p') には、インプットタイプ Q T アトム (' t y') としてビデオメディアを表す 'v i d e' が設定され、データソースタイプ Q T アトム ('d t s t') としてソーストラック

10

20

30

40

50

1 の名称 's r c A' が設定されている。

【 0 1 1 9 】

また、メタトラックのムービーリソースにおいて、サンプルディスクリプションアトム ('s t s d') のメタIDにより、2つの種別「m e t a _ _ t y p e 1」および「m e t a _ _ t y p e 2」が定義されている。

【 0 1 2 0 】

この例では、メタトラックのメディアデータ ('m d a t') において、メタサンプルが4つ設けられている。ソーストラックは1つだけであり、全て同じソーストラック1を示している。また、メタサンプル#1および#3が「m e t a _ _ t y p e 1」を示し、メタサンプル#2および#4が「m e t a _ _ t y p e 2」を示している。

10

【 0 1 2 1 】

図23は、図22の例におけるソーストラック(第1のビデオデータ101)とメタトラック(メタデータ103)との関係を示す図である。この例では、ソーストラックは1つだけであり、全て同じソーストラック1('s r c A')を対象としている。

【 0 1 2 2 】

ここで、「m e t a _ _ t y p e 2」を特定区間として抽出する場合には、ソーストラックからこの特定区間のみが抽出されることになり、それ以外の区間は不要な区間として扱われる。

【 0 1 2 3 】

図24は、本発明の実施の形態におけるメタデータを含むファイルの階層構造の他の例を示す図である。この例では、ソーストラック(第1のビデオデータ101)が2つ(ソーストラック1およびソーストラック2)存在することが想定されている。ソーストラック1のムービーリソース('m o o v')のトラックヘッダアトム('t k h d')には、そのトラックIDである「#1」が保持されている。ソーストラック2のムービーリソースのトラックヘッダアトム('t k h d')には、そのトラックIDである「#2」が保持されている。また、メタトラック(メタデータ103)のムービーリソースのトラックヘッダアトム('t k h d')には、そのトラックIDである「#3」が保持されている。

20

【 0 1 2 4 】

メタトラックのムービーリソースでは、トラックリファレンスアトム('t r e f')のトラックリファレンスタイプアトム('s s r c')に、ソーストラック1のトラックID「#1」およびソーストラック2のトラックID「#2」がそれぞれ保持されている。

30

【 0 1 2 5 】

また、メディアアトム('m d i a')のトラックインプットマップアトム('i m a p')には、2つのトラックインプットQTアトム('i n')が含まれており、1つ目のトラックインプットQTアトムには、インプットタイプQTアトム('t y')としてビデオメディアを表す'v i d e'が設定され、データソースタイプQTアトム('d t s t')としてソーストラック1の名称's r c A'が設定されている。2つ目のトラックインプットQTアトムには、インプットタイプQTアトムとしてビデオメディアを表す'v i d e'が設定され、データソースタイプQTアトムとしてソーストラック2の名称's r c B'が設定されている。

40

【 0 1 2 6 】

また、メタトラックのムービーリソースにおいて、サンプルディスクリプションアトム ('s t s d') のメタIDにより、2つの種別「m e t a _ _ t y p e 1」および「m e t a _ _ t y p e 2」が定義されている。

【 0 1 2 7 】

この例では、メタトラックのメディアデータ ('m d a t') において、メタサンプルが4つ設けられている。ソーストラックは2つ存在しており、メタサンプル#1および#2が's r c A'を参照し、メタサンプル#3および#4が's r c B'を参照している。また、メタサンプル#1および#3が「m e t a _ _ t y p e 1」を示し、メタサンプル#2および#4が「m e t a _ _ t y p e 2」を示している。

50

【 0 1 2 8 】

図 2 5 は、図 2 4 の例におけるソーストラック（第 1 のビデオデータ 1 0 1）とメタトラック（メタデータ 1 0 3）との関係を示す図である。この例では、ソーストラックは 2 つ存在しており、メタサンプル # 1 および # 2 が 's r c A' を参照し、メタサンプル # 3 および # 4 が 's r c B' を参照している。

【 0 1 2 9 】

ここで、「m e t a _ t y p e 2」を特定区間として抽出する場合には、ソーストラックからこの特定区間のみが抽出されることになり、それ以外の区間は不要な区間として扱われる。

【 0 1 3 0 】

図 2 6 は、本発明の実施の形態におけるソーストラック（第 1 のビデオデータ 1 0 1）およびメタトラック（メタデータ 1 0 3）と編集トラック（編集ビデオデータ 1 0 5）との関係例を示す図である。この例では、ソーストラックは、時刻 t s 0 から始まり、時刻 t s 6 に終了している。また、メタトラックはソーストラックに同期しており、時刻 t s 1 から時刻 t s 2 の区間に笑顔を検出した旨を示し、時刻 t s 3 から時刻 t s 4 の区間に驚いた顔を検出した旨を示し、時刻 t s 5 から時刻 t s 6 の区間に笑顔を検出した旨を示している。

【 0 1 3 1 】

また、編集していないムービーファイルにおいては、ソーストラック時間の時間軸はメディア時間の時間軸と一対一対応していることが多いため、ここでは、時刻 t s 1 = 時刻 t m 1、時刻 t s 2 = 時刻 t m 2、時刻 t s 3 = 時刻 t m 3、時刻 t s 4 = 時刻 t m 4、時刻 t s 5 = 時刻 t m 5、時刻 t s 6 = 時刻 t m 6 としている。

【 0 1 3 2 】

ここで、メタトラックにおいて笑顔または驚いた顔を示す区間を抽出条件としてソーストラックを編集することを想定すると、その出力として図のような編集トラックが生成される。すなわち、ソーストラック時間の時刻 t s 1 から時刻 t s 2 の区間のソーストラックの部分が、編集トラックにおける編集トラック時間の時刻 0 から時刻 t e 1 の区間になり、ソーストラック時間の時刻 t s 3 から時刻 t s 4 の区間のソーストラックの部分が、編集トラックにおける編集トラック時間の時刻 t e 1 から時刻 t e 2 の区間になり、ソーストラック時間の時刻 t s 5 から時刻 t s 6 の区間のソーストラックの部分が、編集トラックにおける編集トラック時間の時刻 t e 2 から時刻 t e 3 の区間になる。

【 0 1 3 3 】

図 2 7 は、Q u i c k T i m e ファイルフォーマットにおけるエディットアトム（'e d t s'）の記載例を示す図である。このエディットアトムは、図 1 7 により説明したストリームディスクリプターアトム（'s t r d'）と同様に、サイズ、タイプ、バージョンおよびフラグ群の各フィールドを保持している。このエディットアトムでは、タイプ名として 'e d t s' が保持されている。

【 0 1 3 4 】

このエディットアトムは、さらにエディットリストアトム（'e l s t'）を保持する。このエディットリストアトムは、エディットアトムと同様に、サイズ、タイプ、バージョンおよびフラグ群の各フィールドを保持している。このエディットリストアトムでは、タイプ名として 'e l s t' が保持されている。エディットリストアトムは、さらに N 個（N は 1 以上の整数）のエディットリストエントリ（Edit List Entry）と、その数（Number of Entries）とを含んでいる。

【 0 1 3 5 】

エディットリストエントリの各々は、セグメント期間（Segment duration）と、メディア時間（Media time）と、メディアレート（Media rate）とを備えている。

【 0 1 3 6 】

このエディットアトム（'e d t s'）を模式的に表したものが図 2 8 である。図 2 8（a）のように、エディットアトム 6 8 0 は、エディットアトム 6 8 0 の大きさを示すサイ

10

20

30

40

50

ズ 6 8 1 と、エディットアトムであることを示すタイプ 6 8 2 と、エディットアトムのバージョン 6 8 3 と、未使用のフラグ群 6 9 4 と、エディットリストアトム 5 7 0 とから構成される。

【 0 1 3 7 】

エディットリストアトム 6 9 0 は、エディットリストアトム 6 9 0 の大きさを示すサイズ 6 9 1 と、エディットリストアトムであることを示すタイプ 6 9 2 と、エディットリストアトムのバージョン 6 9 3 と、未使用のフラグ群 6 9 4 と、エディットリストテーブル 6 9 6 と、エディットリストテーブル 6 9 6 のエントリ数 6 9 5 とから構成される。

【 0 1 3 8 】

エディットリストテーブル 6 9 6 はエントリ数 6 9 5 により示される数のエントリから構成される。図 2 8 (b) のようにエディットリストテーブル 6 9 6 の各エントリは、セグメント期間 6 9 7 と、メディア時間 6 9 8 と、メディアレート 6 9 9 とから構成される。セグメント期間 6 9 7 は、対応する編集単位の期間を示す。メディア時間 6 9 8 は、対応する編集単位のメディアデータアトムにおける開始時刻を示す。なお、このメディア時間 6 9 8 が「 - 1 」を示している場合には、対応する編集単位はメディアデータアトムに存在しないことを意味する。メディアレート 6 9 9 は、再生の際の時間比率を示すものであり、メディアデータアトムにおける時間軸と編集後の時間軸とで再生時間が変わらない場合には「 1 . 0 」を示すことになる。

【 0 1 3 9 】

図 2 9 は、図 2 6 の例におけるエディットリストアトム (' e l s t ') の内容例を示す図である。図 2 9 (a) は編集前のエディットリストアトムの内容である。編集前の状態では、エントリは 1 つだけ存在する。セグメント期間は全期間の $t s 6$ を示す。メディア時間は先頭時刻の 0 を示す。また、メディアレート 6 9 9 は、「 1 . 0 」である。

【 0 1 4 0 】

図 2 9 (b) は編集後のエディットリストアトムの内容である。編集後は、3 つのエントリが生成される。

【 0 1 4 1 】

第 1 のエントリでは、最初の笑顔の期間 $t e 1 (= t m 2 - t m 1)$ を示す。メディア時間は最初の笑顔の先頭時刻 $t m 1$ を示す。また、メディアレート 6 9 9 は、「 1 . 0 」である。

【 0 1 4 2 】

第 2 のエントリでは、驚いた顔の期間 $t e 2 - t e 1 (= t m 4 - t m 3)$ を示す。メディア時間は驚いた顔の先頭時刻 $t m 3$ を示す。また、メディアレート 6 9 9 は、「 1 . 0 」である。

【 0 1 4 3 】

第 3 のエントリでは、2 回目の笑顔の期間 $t e 3 - t e 2 (= t m 6 - t m 5)$ を示す。メディア時間は 2 回目の笑顔の先頭時刻 $t m 5$ を示す。また、メディアレート 6 9 9 は、「 1 . 0 」である。

【 0 1 4 4 】

図 3 0 は、本発明の実施の形態におけるソーストラック (第 1 のビデオデータ 1 0 1) およびメタトラック (メタデータ 1 0 3) と編集トラック (編集ビデオデータ 1 0 5) との他の関係例を示す図である。この例では、ソーストラックおよびメタトラックは図 2 6 の例と同様の関係になっている。

【 0 1 4 5 】

ここで、メタトラックにおいて笑顔を示す区間を抽出条件としてソーストラックを編集することを想定すると、その出力として図のような編集トラックが生成される。すなわち、ソーストラック時間の時刻 $t s 1$ から時刻 $t s 2$ の区間のソーストラックの部分が、編集トラックにおける編集トラック時間の時刻 0 から時刻 $t e 1$ の区間になり、ソーストラック時間の時刻 $t s 5$ から時刻 $t s 6$ の区間のソーストラックの部分が、編集トラックにおける編集トラック時間の時刻 $t e 1$ から時刻 $t e 2$ の区間になる。

10

20

30

40

50

【 0 1 4 6 】

図 3 1 は、図 3 0 の例におけるエディットリストアトム (' e l s t ') の内容例を示す図である。この図は編集後のエディットリストアトムの内容である。編集後は、2つのエントリが生成される。

【 0 1 4 7 】

第 1 のエントリでは、最初の笑顔の期間 $t e 1 (= t m 2 - t m 1)$ を示す。メディア時間は最初の笑顔の先頭時刻 $t m 1$ を示す。また、メディアレート 6 9 9 は、「 1 . 0 」である。

【 0 1 4 8 】

第 2 のエントリでは、2 回目の笑顔の期間 $t e 2 - t e 1 (= t m 6 - t m 5)$ を示す。メディア時間は 2 回目の笑顔の先頭時刻 $t m 5$ を示す。また、メディアレート 6 9 9 は、「 1 . 0 」である。

【 0 1 4 9 】

このようにして生成されたエディットリストアトム (エディットリスト 1 0 4) は映像抽出部 1 2 3 に供給され、この映像抽出部 1 2 3 においてソーストラック (第 1 のビデオデータ 1 0 1) から編集トラック (編集ビデオデータ 1 0 5) が生成される。すなわち、元のソーストラックを破壊することのない非破壊型編集を実現することができる。

【 0 1 5 0 】

次に本発明の実施の形態における映像編集システムの動作について図面を参照して説明する。

【 0 1 5 1 】

図 3 2 は、本発明の実施の形態における映像処理装置 1 1 0 による処理手順の一例を示す図である。まず、映像取得部 1 1 1 によって、第 2 のビデオデータ 1 0 2 の映像が取得される (ステップ S 9 1 1) 。この第 2 のビデオデータ 1 0 2 は、第 1 のビデオデータ 1 0 1 に同期しており、図 2 の例のように第 1 のビデオデータ 1 0 1 の撮像と同時に撮像されたものでもよく、図 4 の例のように第 1 のビデオデータ 1 0 1 の再生と同時に撮像されたものでもよく、また、図 6 の例のように第 1 のビデオデータ 1 0 1 と同一の内容であってもよい。

【 0 1 5 2 】

ステップ S 9 1 1 で映像が取得されると、取得された第 2 のビデオデータ 1 0 2 における特徴量が映像解析部 1 1 2 によって解析される (ステップ S 9 1 2) 。例えば、特徴量として第 2 のビデオデータ 1 0 2 に含まれる顔の画像が抽出されて、その顔の表情が判断される。

【 0 1 5 3 】

そして、ステップ S 9 1 2 において解析された特徴量を時間軸により管理するメタデータ 1 0 3 がメタデータ生成部 1 1 3 によって生成される (ステップ S 9 1 3) 。この生成されたメタデータ 1 0 3 は、第 1 のビデオデータ 1 0 1 の各時刻に対応して、第 2 のビデオデータ 1 0 2 の特徴量を保持するものである。

【 0 1 5 4 】

これらステップ S 9 1 1 乃至 S 9 1 3 の処理手順は、第 2 のビデオデータ 1 0 2 の映像が全て処理されるまで繰り返される (ステップ S 9 1 4) 。

【 0 1 5 5 】

図 3 3 は、本発明の実施の形態における映像編集装置 1 2 0 による処理手順の一例を示す図である。まず、抽出条件受付部 1 2 1 によって、抽出条件の入力が受け付けられる (ステップ S 9 2 1) 。そして、その抽出条件によってメタデータ 1 0 3 における位置が探索される (ステップ S 9 3 0) 。これにより、抽出条件に合致する時系列上の位置が得られ、エディットリスト 1 0 4 として保持される。

【 0 1 5 6 】

このエディットリスト 1 0 4 に基づいて、第 1 のビデオデータ 1 0 1 から抽出条件に合致する時系列上の位置に対応する部分が映像抽出部 1 2 3 によって抽出され、編集ビデオ

10

20

30

40

50

データ 105 として出力される (ステップ S 923)。

【0157】

図 34 は、図 33 の位置探索処理 (ステップ S 930) における処理手順の一例を示す図である。まず、メタデータのムービーリソース (図 11 のメタトラック) が取得される (ステップ S 931)。そして、このムービーリソースのサンプルディスクリプションアトム ('s t s d') に含まれるパラメータフラグ (図 20 参照) によりエフェクトの有効性、すなわち特徴量の記録の有効性が判断される (ステップ S 932)。

【0158】

ステップ S 932 において「有効」とであると判断された場合には、メタデータの対応するメディアデータのサンプルが取得される (ステップ S 933)。その結果、メタデータのサンプルの示す特徴量が抽出条件と一致していれば (ステップ S 934)、その該当する区間 (セグメント) がエディットリストアトム (エディットリスト 104) のエントリとして登録される (ステップ S 935)。

【0159】

一方、ステップ S 932 において「無効」とであると判断された場合には、メタデータのメディアデータは取得されることなく、そのサンプルに関する処理は終了する。

【0160】

これらステップ S 931 乃至 S 935 の処理手順は、メタデータ 103 のサンプルが全て処理されるまで繰り返される (ステップ S 936)。

【0161】

このように、本発明の実施の形態によれば、第 1 のビデオデータ 101 に同期する第 2 のビデオデータ 102 から映像処理装置 110 によってメタデータ 103 が生成される。このメタデータ 103 は、第 1 のビデオデータ 101 に同期しており、第 2 のビデオデータ 102 の特徴量を時間軸で管理しながら保持するものである。また、本発明の実施の形態によれば、メタデータ 103 において抽出条件に合致する位置が映像編集装置 120 の位置探索部 122 によって探索され、エディットリスト 104 が生成される。このエディットリスト 104 に従って、第 1 のビデオデータ 101 から映像抽出部 123 によって映像が抽出され、編集ビデオデータ 105 が生成される。すなわち、本発明の実施の形態によれば、メタデータ 103 を中間状態として、第 1 のビデオデータ 101 を非破壊的に編集することができる。

【0162】

なお、本発明の実施の形態は本発明を具現化するための一例を示したものであり、以下に示すように特許請求の範囲における発明特定事項とそれぞれ対応関係を有するが、これに限定されるものではなく本発明の要旨を逸脱しない範囲において種々の変形を施すことができる。

【0163】

すなわち、請求項 1 において、映像取得手段は例えば映像取得部 111、撮像部 211、216、または、映像入力部 206 に対応する。また、映像解析手段は例えば映像解析部 112 または 217 に対応する。また、メタデータ生成手段は例えばメタデータ生成部 113 またはファイル生成部 214 に対応する。

【0164】

また、請求項 3 において、第 1 の撮像手段は例えば撮像部 211 に対応する。また、第 2 の撮像手段は例えば撮像部 216 に対応する。

【0165】

また、請求項 4 において、再生手段は例えば再生部 220 に対応する。また、撮像手段は例えば撮像部 211 に対応する。

【0166】

また、請求項 5 において、再生手段は例えば再生部 220 に対応する。また、映像入力手段は例えば映像入力部 206 に対応する。

【0167】

10

20

30

40

50

また、請求項 7 において、位置探索手段は例えば位置探索部 1 2 2 に対応する。また、映像抽出手段は例えば映像抽出部 1 2 3 に対応する。

【0 1 6 8】

また、請求項 9 において、映像取得手段は例えば映像取得部 1 1 1、撮像部 2 1 1、2 1 6、または、映像入力部 2 0 6 に対応する。また、映像解析手段は例えば映像解析部 1 1 2 または 2 1 7 に対応する。また、メタデータ生成手段は例えばメタデータ生成部 1 1 3 またはファイル生成部 2 1 4 に対応する。また、位置探索手段は例えば位置探索部 1 2 2 に対応する。また、映像抽出手段は例えば映像抽出部 1 2 3 に対応する。

【0 1 6 9】

また、請求項 1 0 および 1 1 において、映像取得手順は例えばステップ S 9 1 1 に対応する。また、映像解析手順は例えばステップ S 9 1 2 に対応する。また、メタデータ生成手順は例えばステップ S 9 1 3 に対応する。

【0 1 7 0】

また、請求項 1 2 および 1 3 において、メタデータ取得手順は例えばステップ S 9 3 3 に対応する。また、位置探索手順は例えばステップ S 9 3 4 に対応する。また、映像抽出手順は例えばステップ S 9 2 3 に対応する。

【0 1 7 1】

また、請求項 1 4 において、第 1 の映像データは例えば第 1 のビデオデータ 1 0 1 に対応する。また、第 2 の映像データは例えば第 2 のビデオデータ 1 0 2 に対応する。また、メタデータは例えばメタデータ 1 0 3 に対応する。

【0 1 7 2】

なお、本発明の実施の形態において説明した処理手順は、これら一連の手順を有する方法として捉えてもよく、また、これら一連の手順をコンピュータに実行させるためのプログラム乃至そのプログラムを記憶する記録媒体として捉えてもよい。

【図面の簡単な説明】

【0 1 7 3】

【図 1】本発明の実施の形態における映像編集システム 1 0 0 の一構成例を示す図である。

【図 2】本発明の実施の形態における映像処理装置 1 1 0 の第 1 の構成例を示す図である。

【図 3】本発明の実施の形態における映像処理装置 1 1 0 の第 1 の構成例による使用態様を示す図である。

【図 4】本発明の実施の形態における映像処理装置 1 1 0 の第 2 の構成例を示す図である。

【図 5】本発明の実施の形態における映像処理装置 1 1 0 の第 2 の構成例による使用態様を示す図である。

【図 6】本発明の実施の形態における映像処理装置 1 1 0 の第 3 の構成例を示す図である。

【図 7】本発明の実施の形態における映像処理装置 1 1 0 の一実施例であるカメラ一体型撮像装置の構成例を示す図である。

【図 8】Quick Time ベースファイル形式の構造例を示す図である。

【図 9】Quick Time ファイルフォーマットにおける階層構造を示す図である。

【図 1 0】本発明の実施の形態におけるファイルの保存形式の一例を示す図である。

【図 1 1】本発明の実施の形態におけるメタトラックの階層構造の一例を示す図である。

【図 1 2】本発明の実施の形態におけるメタトラックを含むムービーアトムの記載例を示す図である。

【図 1 3】本発明の実施の形態におけるメタトラックを含むムービーアトムの他の記載例を示す図である。

【図 1 4】図 1 3 のムービーアトムにおけるメタトラックアトムの記載例を示す図である。

10

20

30

40

50

【図 1 5】本発明の実施の形態におけるメタトラックのサンプルディスクリプションアトム('s t s d')の記載例を示す図である。

【図 1 6】本発明の実施の形態におけるデータフォーマットフィールドの一例を示す図である。

【図 1 7】本発明の実施の形態におけるストリームディスクリプターアトム('s t r d')の記載例を示す図である。

【図 1 8】本発明の実施の形態におけるユーザデファインドメタタイプフィールドの記載例を示す図である。

【図 1 9】本発明の実施の形態におけるメタ ID のビットフィールド構成例を示す図である。

10

【図 2 0】本発明の実施の形態におけるパラメータフラグフィールドのビットフィールド構成例を示す図である。

【図 2 1】本発明の実施の形態におけるメタトラックのメディアデータのサンプルの記載例を示す図である。

【図 2 2】本発明の実施の形態におけるメタデータを含むファイルの階層構造の一例を示す図である。

【図 2 3】図 2 2 の例におけるソーストラックとメタトラックとの関係を示す図である。

【図 2 4】本発明の実施の形態におけるメタデータを含むファイルの階層構造の他の例を示す図である。

【図 2 5】図 2 4 の例におけるソーストラックとメタトラックとの関係を示す図である。

20

【図 2 6】本発明の実施の形態におけるソーストラックおよびメタトラックと編集トラックとの関係例を示す図である。

【図 2 7】Q u i c k T i m e ファイルフォーマットにおけるエディットアトム('e d t s')の記載例を示す図である。

【図 2 8】Q u i c k T i m e ファイルフォーマットにおけるエディットアトムを模式的に表した図である。

【図 2 9】図 2 6 の例におけるエディットリストアトム('e l s t')の内容例を示す図である。

【図 3 0】本発明の実施の形態におけるソーストラックおよびメタトラックと編集トラックとの他の関係例を示す図である。

30

【図 3 1】図 3 0 の例におけるエディットリストアトムの内容例を示す図である。

【図 3 2】本発明の実施の形態における映像処理装置 1 1 0 による処理手順の一例を示す図である。

【図 3 3】本発明の実施の形態における映像編集装置 1 2 0 による処理手順の一例を示す図である。

【図 3 4】図 3 3 の位置探索処理(ステップ S 9 3 0)における処理手順の一例を示す図である。

【符号の説明】

【0 1 7 4】

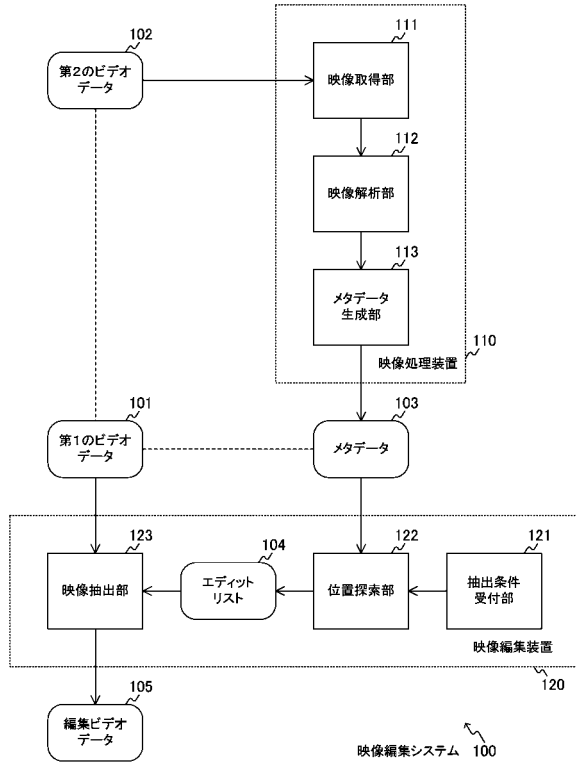
- 1 0 0 映像編集システム
- 1 0 1 第 1 のビデオデータ
- 1 0 2 第 2 のビデオデータ
- 1 0 3 メタデータ
- 1 0 4 エディットリスト
- 1 0 5 編集ビデオデータ
- 1 1 0 映像処理装置
- 1 1 1 映像取得部
- 1 1 2 映像解析部
- 1 1 3 メタデータ生成部
- 1 2 0 映像編集装置

40

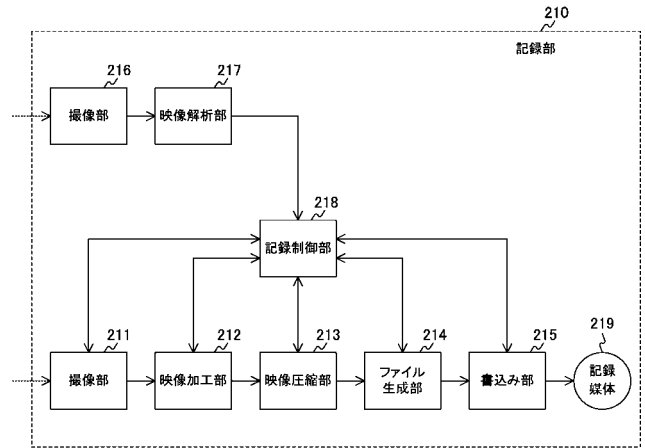
50

1 2 1	抽出条件受付部	
1 2 2	位置探索部	
1 2 3	映像抽出部	
2 0 6	映像入力部	
2 1 0	記録部	
2 1 1	撮像部	
2 1 2	映像加工部	
2 1 3	映像圧縮部	
2 1 4	ファイル生成部	
2 1 5	書込み部	10
2 1 6	撮像部	
2 1 7	映像解析部	
2 1 8	記録制御部	
2 1 9	記録媒体	
2 2 0	再生部	
2 2 1	表示部	
2 2 2	映像加工部	
2 2 3	映像伸張部	
2 2 4	ファイル復号部	
2 2 5	読出し部	20
2 2 9	記録媒体	
3 0 1	撮像部	
3 0 9	記録媒体	
3 1 1	ズームレンズ	
3 1 2	アイリス	
3 1 3	フォーカスレンズ	
3 1 4	フィルタ	
3 2 1	撮像素子	
3 2 2	A / D 変換器	
3 2 3	カメラ信号処理回路	30
3 2 4	検波部	
3 2 5	ズーム制御部	
3 2 6	角速度センサ	
3 2 9	カメラ制御部	
3 3 0	映像加工部	
3 4 1	映像圧縮部	
3 4 2	圧縮制御部	
3 5 1	記録媒体アクセス部	
3 5 2	ドライブ制御部	
3 6 0	操作受付部	40
3 7 0	表示部	
3 9 0	システム制御部	
5 0 1	被写体	
5 0 2	撮影者	
5 0 3 ~ 5 0 5	鑑賞者	
5 1 0	ネットワーク	
5 2 0、5 3 1	ビデオカメラ装置	
5 2 1、5 2 2、5 4 1	カメラ	
5 3 0	テレビ装置	
5 4 0	コンピュータ装置	50

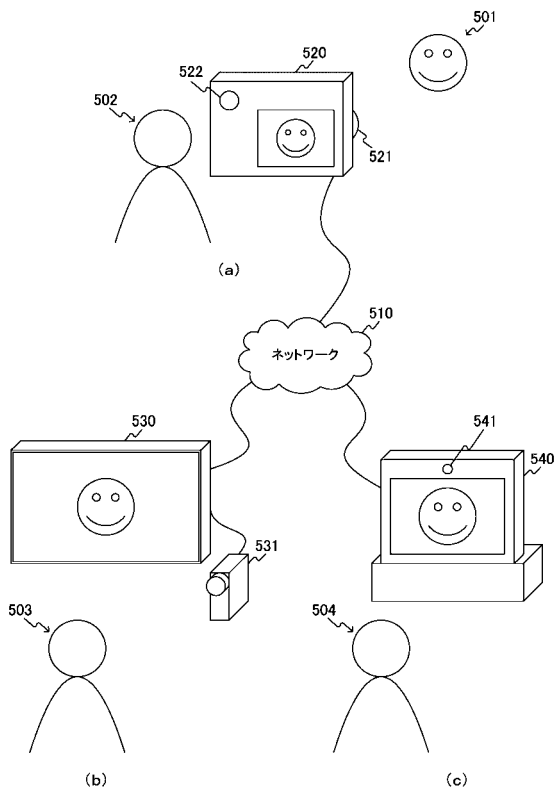
【図 1】



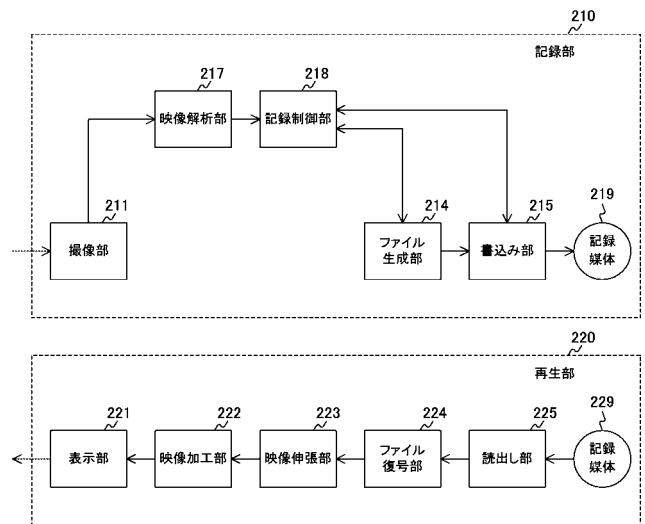
【図 2】



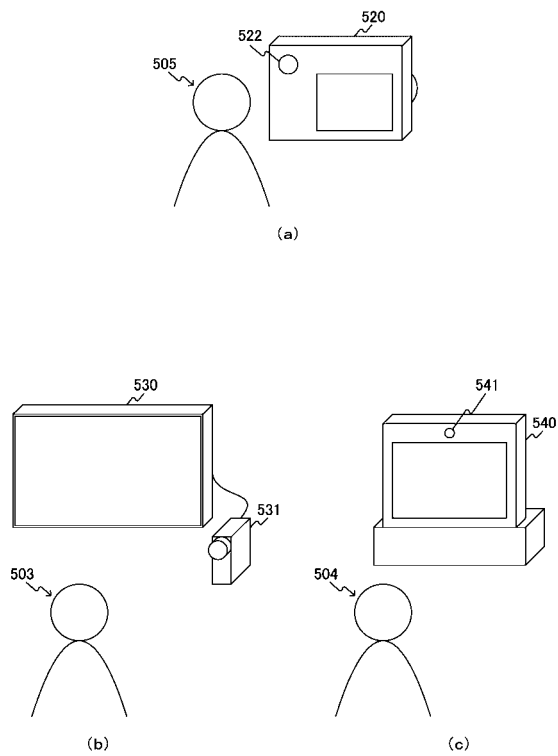
【図 3】



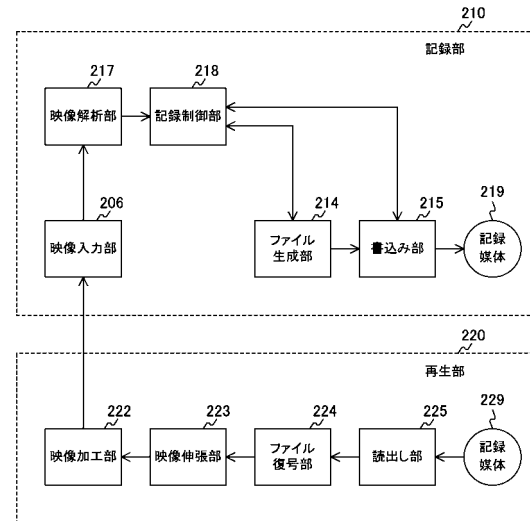
【図 4】



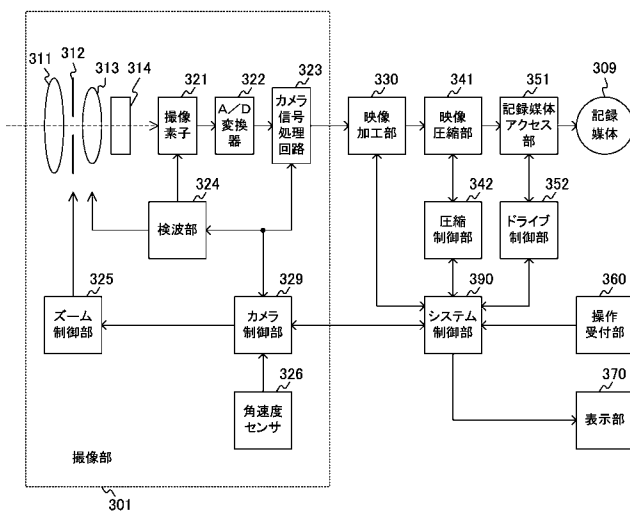
【図 5】



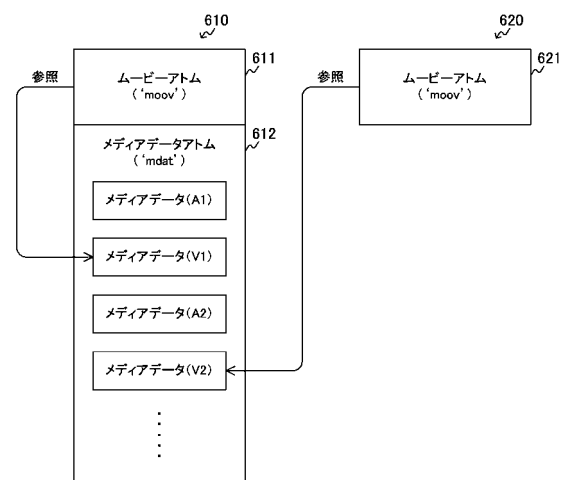
【図 6】



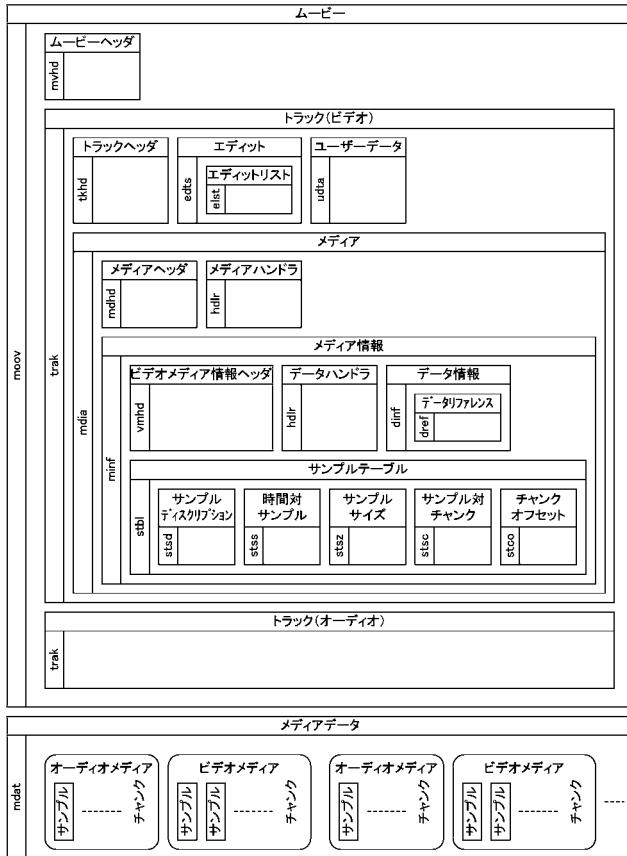
【図 7】



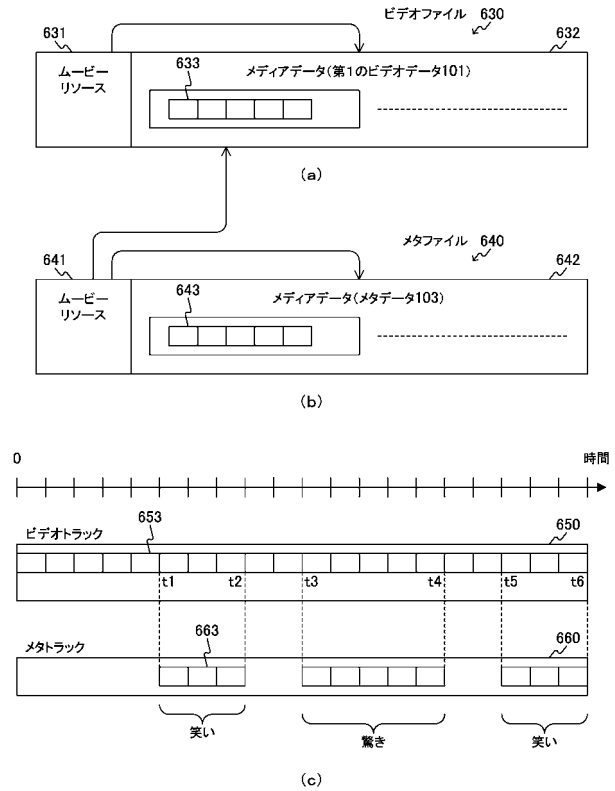
【図 8】



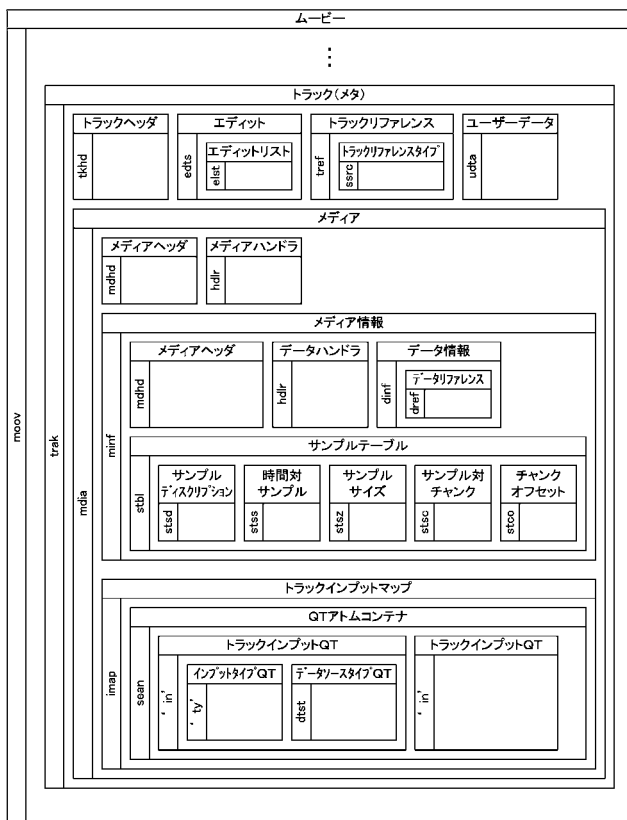
【図 9】



【図 10】



【図 11】



【図 12】

Atom	Type
Movie atom {	
Movie Header atom {	'mvhd'
Track atom (video) {	'trak'
Track Header atom {	'tkhd'
Edit atom {	'edts'
Edit List atom {	'elst'
Media atom {	'mdia'
Media Header atom {	'mdhd'
Media Handler Reference atom {	'hdlr'
Media Information atom {	'minf'
Video Media Information Header atom {	'vmhd'
Data Handler Reference atom {	'hdlr'
Data Information atom {	'dinf'
Data Reference atom {	'dref'
Sample Table atom {	'stbl'
Sample Description atom {	'std'
Time-to-Sample atom {	'stts'
Sample Size atom {	'stsz'
Sample-to-Chunk atom {	'stsc'
Chunk Offset atom {	'stco'
} }	
User data atom {	'udta'
} }	
Track atom (meta) {	'trak'
Track Header atom {	'tkhd'
Edit atom {	'edts'
Edit List atom {	'elst'
Track Reference atom {	'trf'
Track Reference Type atom {	'trst'
Media atom {	'mdia'
Media Header atom {	'mdhd'
Media Handler Reference atom {	'hdlr'
Media Information atom {	'minf'
Video Media Information Header atom {	'vmhd'
Data Handler Reference atom {	'hdlr'
Data Information atom {	'dinf'
Data Reference atom {	'dref'
Sample Table atom {	'stbl'
Sample Description atom {	'std'
Time-to-Sample atom {	'stts'
Sample Size atom {	'stsz'
Sample-to-Chunk atom {	'stsc'
Chunk Offset atom {	'stco'
} }	
} }	
Track Input Map atom {	'imap'
QT atom container {	'sean'
Track Input QT atom {	'in'
Input Type QT atom {	'ty'
Data Source Type QT atom {	'dst'
} }	
} }	
User data atom {	'udta'
} }	
} }	
Movie Data atom {	'mdat'

【図 13】

Atom	Type
Movie atom {	'moov'
Movie Header atom	'mvhd'
Track atom (video 1) {	'trak'
Track Header atom	'tkhd'
Edit atom {	'edts'
Edit List atom	'elst'
Media atom {	'mdia'
Media Header atom	'mdhd'
Media Handler Reference atom	'hdlr'
Media Information atom {	'minf'
Video Media Information Header atom	'vmhd'
Data Handler Reference atom	'hdlr'
Data Information atom {	'dinf'
Data Reference atom	'dref'
Sample Table atom {	'stbl'
Sample Description atom	'stds'
Time-to-Sample atom	'stts'
Sample Size atom	'stsz'
Sample-to-Chunk atom	'stsc'
Chunk Offset atom	'stco'
}	
}	
User data atom	'udta'
Track atom (video 2) {	'trak'
Track Header atom	'tkhd'
Edit atom {	'edts'
Edit List atom	'elst'
Media atom {	'mdia'
Media Header atom	'mdhd'
Media Handler Reference atom	'hdlr'
Media Information atom {	'minf'
Video Media Information Header atom	'vmhd'
Data Handler Reference atom	'hdlr'
Data Information atom {	'dinf'
Data Reference atom	'dref'
Sample Table atom {	'stbl'
Sample Description atom	'stds'
Time-to-Sample atom	'stts'
Sample Size atom	'stsz'
Sample-to-Chunk atom	'stsc'
Chunk Offset atom	'stco'
}	
}	
User data atom	'udta'
Track atom (meta) {	'trak'
...	
Movie Data atom	'mdat'

【図 14】

Atom	Type
Track atom (meta) {	'trak'
Track Header atom	'tkhd'
Edit atom {	'edts'
Edit List atom	'elst'
Track Reference atom {	'tref'
Track Reference Type atom	'srct'
Media atom {	'mdia'
Media Header atom	'mdhd'
Media Handler Reference atom	'hdlr'
Media Information atom {	'minf'
Video Media Information Header atom	'vmhd'
Data Handler Reference atom	'hdlr'
Data Information atom {	'dinf'
Data Reference atom	'dref'
Sample Table atom {	'stbl'
Sample Description atom	'stds'
Time-to-Sample atom	'stts'
Sample Size atom	'stsz'
Sample-to-Chunk atom	'stsc'
Chunk Offset atom	'stco'
}	
Track Input Map atom {	'imap'
QAtom container {	'sean'
Track Input QAtom {	'in'
Input Type QAtom	'ty'
Data Source Type QAtom	'dtst'
}	
Track Input QAtom {	'in'
Input Type QAtom	'ty'
Data Source Type QAtom	'dtst'
}	
}	
User data atom	'udta'

【図 15】

Sample Description atom {
(4) Size
(4) Type(="stds")
(1) Version
(3) Flags
(4) Number of Entries
(112) Meta Sample Description entry#1
:
(112) Meta Sample Description entry#M
}
Meta Sample Description entry {
(4) Size
(4) Data Format
(6) Reserved
(2) Data Reference Index
(2) Version
(2) Revision Level
(4) Vendor
(4) Temporal Quality
(4) Spatial Quality
(2) Width
(2) Height
(4) Horizontal Resolution
(4) Vertical Resolution
(4) Data Size
(2) Frame Count
(32) Compressor Name
(2) Depth
(2) Color Table ID
/* Data Format extension atom */
(26) Stream Descriptor atom
}

【図 16】

タイプ名	エフェクト名	内容
'blur'	Blur	Amount of blurring 7 levels
'solr'	Color Style	Solarize amount, Posterize amount may be specified
'tin'	Color Tint	Black and white, X-Ray, Sepia, Cobalt
'edge'	Edge Detection	Edge thickness 7 levels, Colorize on edge detection
'embs'	Emboss	Amount of emboss 7 levels
'hslb'	HSL Color Balance	Hue, saturation, lightness may be specified
'rgbb'	RGB Color Balance	R/G/B may be specified
'shrp'	Sharpen	Amount of sharpness 7 levels
'YPST'	Posterization	Posterize
'MOSA'	Mosaic	Mosaic
'NEGA'	RGB Reverse (Negative)	The positive-negative reverse effect
'brco'	Brightness and Contrast	Brightness and contrast adjustment
'ckey'	Chroma Key	Color Replacement
'dslt'	Cross Fade	Fade level may be specified
'push'	Push	Direction may be specified
'slid'	Slide	Angle may be specified
'RDOT'	Random Dot	Replaces the first source with second source by dot
'blnd'	Alpha Compositor	Straight, reverse alpha blending
'zoom'	Zoom	Zoom position, distance may be specified
'smp1'	SMPTE Wipe	Selectable from 34 types of effects from up, down, left, right, diagonal, zigzag shapes. SMPTE-ID range: 1-99
'smp2'	SMPTE Iris	Selectable from 26 types of effects from polygon, oval, star, cats eye, keyhole shapes. SMPTE-ID range: 101-199
'smp3'	SMPTE Radial	Selectable from 39 types of radial effects from Rotating direction, Rotating center. SMPTE-ID range: 201-299
'smp4'	SMPTE Matrix	Selectable from 34 types of matrix effects from horizontal, vertical, clockwise, spiral. SMPTE-ID range: 301-399
'UDEF'	User Defined Effect Meta	User-original video effect meta data

【 図 1 7 】

Stream Descriptor atom {
(4) Size
(4) Type(≡' strd')
(1) Version
(3) Flags
/* Data Format specific data */
(4) Data Format
(4) User Defined Meta Type
(2) Parameter Flag
}

【 図 1 8 】

User Defined Meta Type {
(2) Owner ID
(2) Meta ID
}

【 図 1 9 】

Meta ID	
ビット	内容
15:5	(リザーブ)
4	眠い (sleepy)
3	怒り (angry)
2	驚き (surprise)
1	笑い (smile)
0	無表情

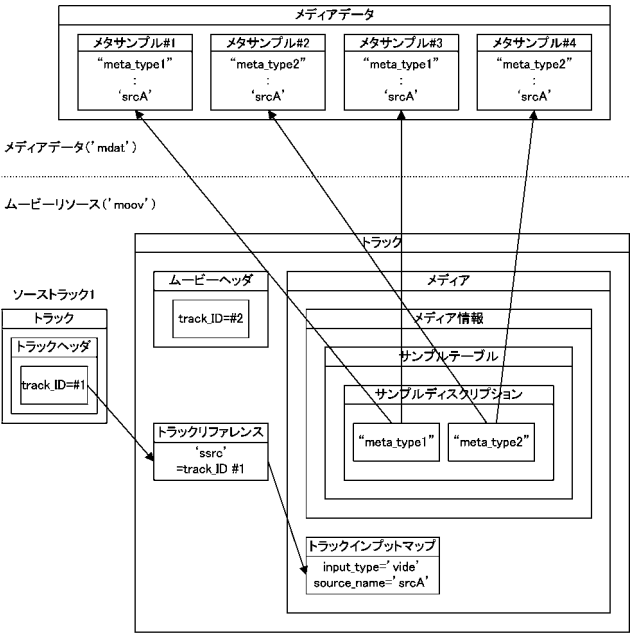
【 図 2 0 】

Parameter Flag			
ビット	フラグ名称	値	内容
15:1	(リザーブ)	0	—
0	エフェクト有効性	1	有効 (アクティブエフェクト)
		0	無効 (スルエフェクト)

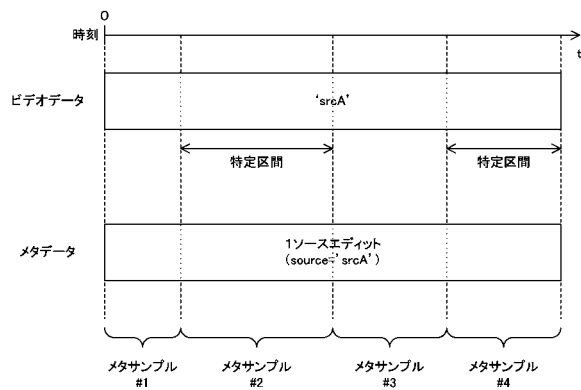
【 図 2 1 】

1sample: {
face_number
#face1 {
{smile / surprise / angry / sleepy}
}
#face2 {
{smile / surprise / angry / sleepy}
}
⋮
#faceN {
{smile / surprise / angry / sleepy}
}
}

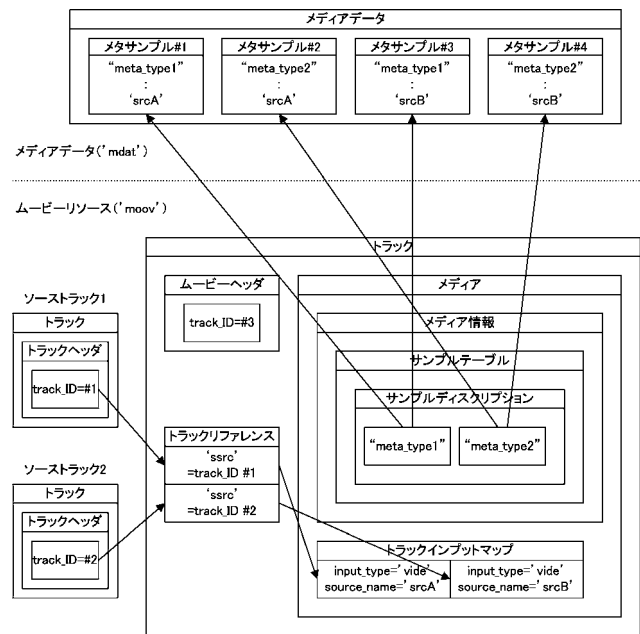
【 図 2 2 】



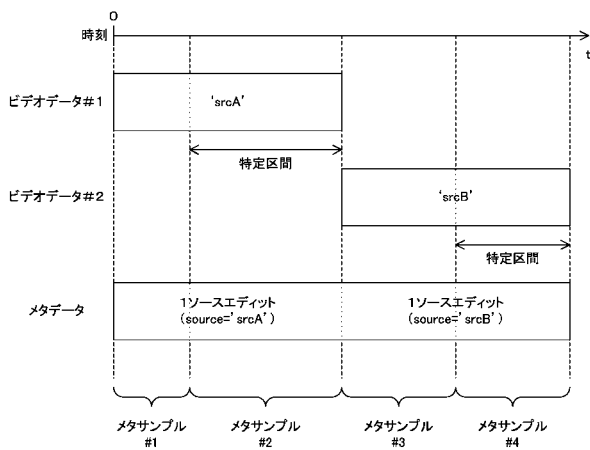
【図 23】



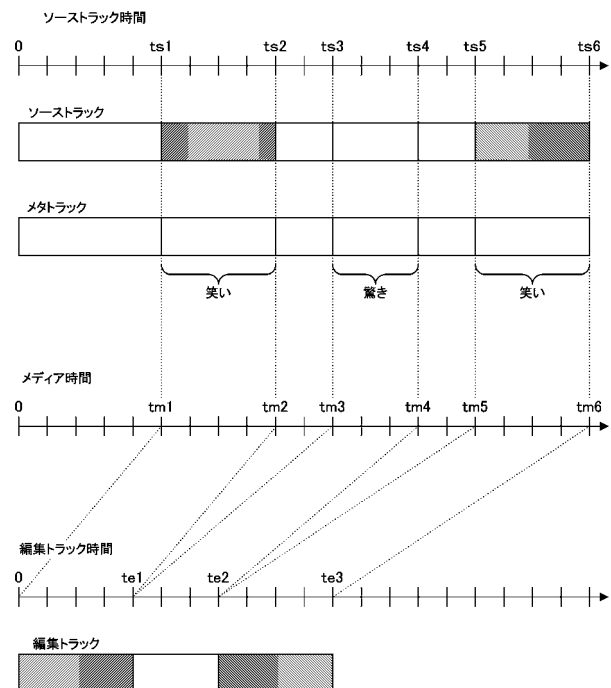
【図 24】



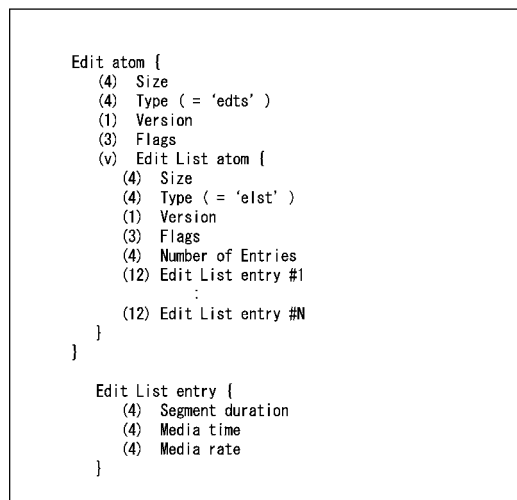
【図 25】



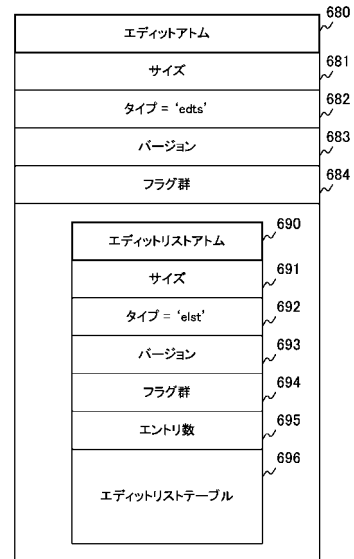
【図 26】



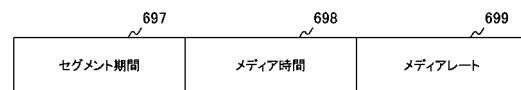
【図 27】



【図 28】



(a)



(b)

【図 29】

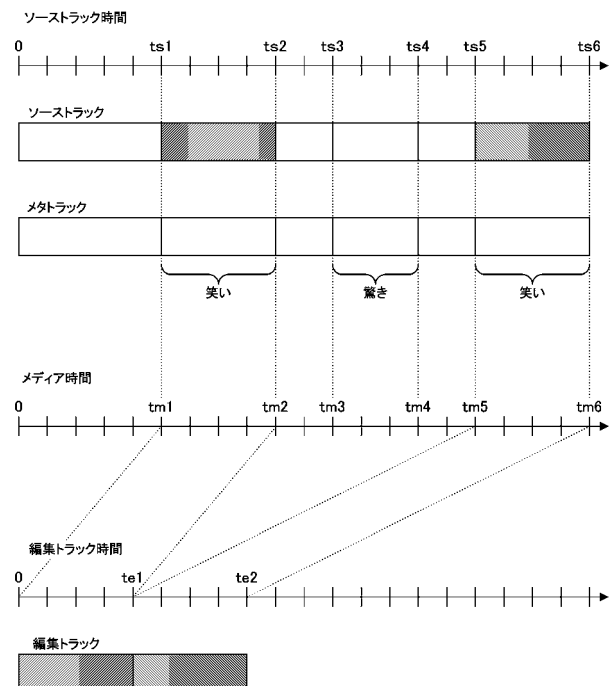
エディットリスト	セグメント期間	メディア時間	メディアレート
エントリ #1	ts6	0	1.0

(a)

エディットリスト	セグメント期間	メディア時間	メディアレート
エントリ #1	te1	tm1	1.0
エントリ #2	te2-te1	tm3	1.0
エントリ #3	te3-te2	tm5	1.0

(b)

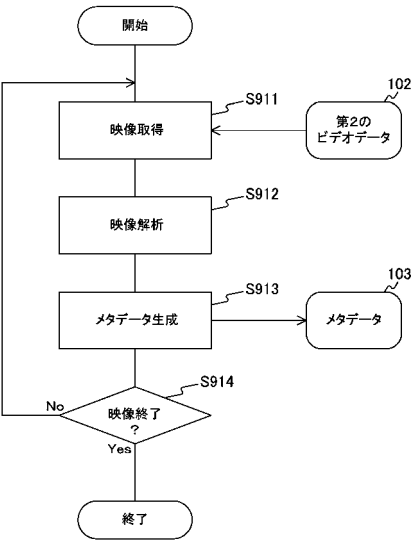
【図 30】



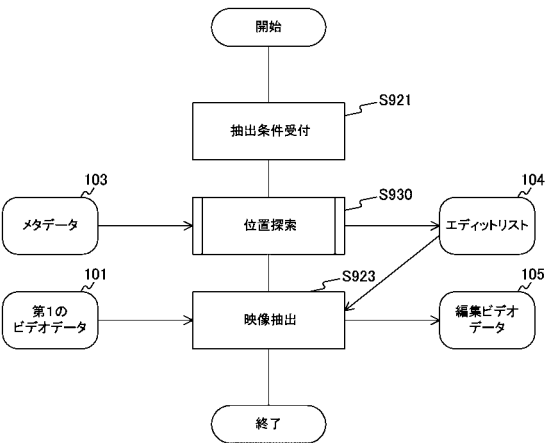
【図 3 1】

エディットリスト	セグメント期間	メディア時間	メディアレート
エントリ#1	te1	tm1	1.0
エントリ#2	te2-te1	tm5	1.0

【図 3 2】



【図 3 3】



【図 3 4】

