



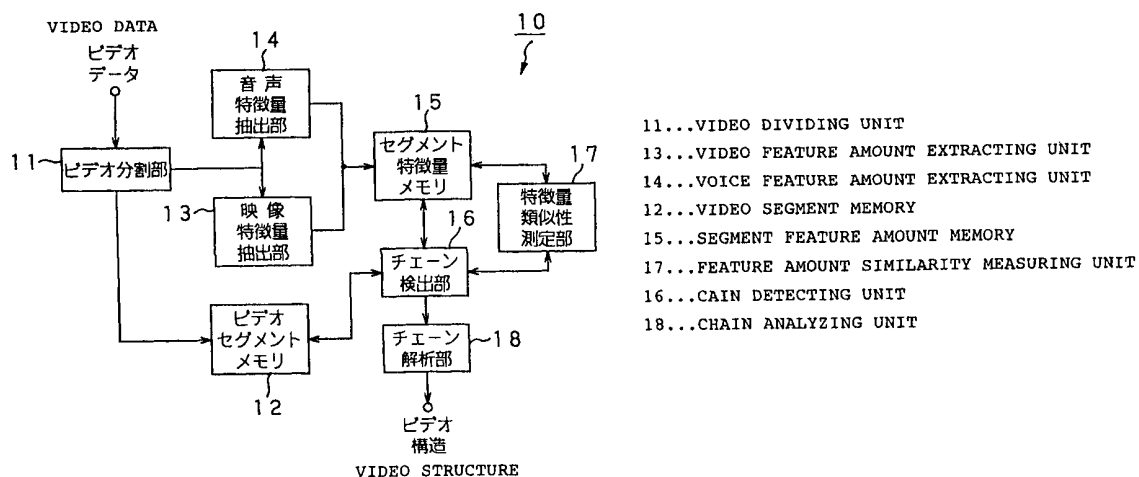
PCT

特許協力条約に基づいて公開された国際出願

(51) 国際特許分類7 H04N 7/32, 5/92, G06T 7/00	A1	(11) 国際公開番号 WO00/45603 (43) 国際公開日 2000年8月3日(03.08.00)
(21) 国際出願番号 PCT/JP00/00422 (22) 国際出願日 2000年1月27日(27.01.00) (30) 優先権データ 特願平11/23069 1999年1月29日(29.01.99) JP (71) 出願人 (米国を除くすべての指定国について) ソニー株式会社(SONY CORPORATION)[JP/JP] 〒141-0001 東京都品川区北品川6丁目7番35号 Tokyo, (JP) (72) 発明者; および (75) 発明者/出願人 (米国についてのみ) ウォーカー トビー(WALKER, Toby)[CA/JP] 〒141-0001 東京都品川区北品川6丁目7番35号 ソニー株式会社内 Tokyo, (JP) (74) 代理人 小池 晃, 外(KOIKE, Akira et al.) 〒105-0001 東京都港区虎ノ門二丁目6番4号 第11森ビル Tokyo, (JP)		(81) 指定国 US, 欧州特許 (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE) 添付公開書類 国際調査報告書

(54)Title: SIGNAL PROCESSING METHOD AND VIDEO/VOICE PROCESSING DEVICE

(54)発明の名称 信号処理方法及び映像音声処理装置



(57) Abstract

A video/voice processing device (10), which is provided with a chain detection unit (16) which uses feature amounts extracted from video segments and/or voice segments divided from an input video data stream and a measuring reference for measuring, using the feature amounts, similarities between pairs of video segments and/or voice segments calculated for each feature amount, so as to detect, out of the video segments and/or voice segments, similar chains comprising a plurality of mutually-similar video segments and/or voice segments; and which is provided with a chain analyzing unit (18) which performs analysis using the similar chains and determines a local video structure and/or a general video structure of the video for outputting.

(57)要約

映像音声処理装置 10 は、入力したビデオデータのストリームから分割された映像セグメント及び／又は音声セグメントから抽出された特徴量と、この特徴量を用いて、各特徴量毎に計算された、映像セグメント及び／又は音声セグメントの対の間の類似性を測定する測定基準とを用いて、映像セグメント及び／又は音声セグメントのうち、互いに類似する複数の映像及び／又は音声セグメントから構成される類似チェーンを検出するチェーン検出部 16 と、類似チェーンを用いて解析し、ビデオの局所的ビデオ構造及び／又は大局的ビデオ構造を決定して出力するチェーン解析部 18 とを備える。

PCTに基づいて公開される国際出願のパンフレット第一頁に掲載されたPCT加盟国を同定するために使用されるコード(参考情報)

AE アラブ首長国連邦	DM ドミニカ	KZ カザフスタン	RU ロシア
AG アンティグア・バーブーダ	DZ アルジェリア	LC セントルシア	SD スーダン
AL アルバニア	EE エストニア	LI リヒテンシュタイン	SE スウェーデン
AM アルメニア	ES スペイン	LK スリ・ランカ	SG シンガポール
AT オーストリア	FI フィンランド	LR リベリア	SI スロヴェニア
AU オーストラリア	FR フランス	LS レソト	SK スロヴァキア
AZ アゼルバイジャン	GA ガボン	LT リトアニア	SL シェラ・レオネ
BA ボスニア・ヘルツェゴビナ	GB 英国	LU ルクセンブルグ	SN セネガル
BB バルバドス	GD グレナダ	LV ラトヴィア	SZ スワジランド
BE ベルギー	GE グルジア	MA モロッコ	TD チャード
BF ブルキナ・ファソ	GH ガーナ	MC モナコ	TG トーゴ
BG ブルガリア	GM ガンビア	MD モルドヴァ	TJ タジキスタン
BJ ベナン	GN ギニア	MG マダガスカル	TM トルクメニスタン
BR ブラジル	GR ギリシャ	MK マケドニア旧ユーゴスラヴィア	TR トルコ
BY ベラルーシ	GW ギニア・ビサウ		TT トリニダード・トバゴ
CA カナダ	HR クロアチア	ML マリ	TZ タンザニア
CF 中央アフリカ	HU ハンガリー	MN モンゴル	UA ウクライナ
CG コンゴ	ID インドネシア	MR モーリタニア	UG ウガンダ
CH スイス	IE アイルランド	MW マラウイ	US 米国
CI コートジボアール	IL イスラエル	MX メキシコ	UZ ウズベキスタン
CM カメルーン	IN インド	MZ モザンビーク	VN ヴェトナム
CN 中国	IS アイスランド	NE ニジェール	YU ユーゴスラヴィア
CR コスタ・リカ	IT イタリア	NL オランダ	ZA 南アフリカ共和国
CU キューバ	JP 日本	NO ノールウェー	ZW ジンバブエ
CY キプロス	KE ケニア	NZ ニュー・ジージーランド	
CZ チェッコ	KG キルギスタン	PL ポーランド	
DE ドイツ	KP 北朝鮮	PT ポルトガル	
DK デンマーク	KR 韓国	RO ルーマニア	

明細書

信号処理方法及び映像音声処理装置

技術分野

本発明は、信号の基礎となる意味構造を反映するパターンを検出して解析する信号処理方法及びビデオ信号の基礎となる意味構造を反映する映像及び／又は音声のパターンを検出して解析する映像音声処理装置に関する。

背景技術

例えばビデオデータに録画されたテレビ番組といった大量の異なる映像データにより構成される映像アプリケーションの中から、興味のある部分等の所望の部分を探して再生したい場合がある。

このように、所望の映像内容を抽出するための一般的な技術としては、アプリケーションの主要場面を描いた一連の映像を並べて作成されたパネルであるストーリーボードがある。このストーリーボードは、ビデオデータをいわゆるショットに分解し、各ショットにおいて代表される映像を表示したものである。このような映像抽出技術は、そのほとんどが、例えば “G. Ahanger and T.D.C. Little, A survey of technologies for parsing and indexing digital video, J. of Visual Communication and Image Representation 7: 28-4, 1996” に記載されているように、ビデオデータからショット

を自動的に検出して抽出するものである。

ところで、例えば代表的な30分のテレビ番組中には、数百ものショットが含まれている。そのため、上述した従来の映像抽出技術においては、ユーザは、抽出された膨大な数のショットを並べたストーリーボードを調べる必要があり、このようなストーリーボードを理解する際、ユーザに大きな負担を強いる必要があった。また、従来の映像抽出技術においては、例えば話し手の変化に応じて交互に2者を撮影した会話場面におけるショットは、冗長のものが多いという問題があった。このように、ショットは、ビデオ構造を抽出する対象としては階層が低すぎて無駄な情報量が多く、このようなショットを抽出する従来の映像抽出技術は、ユーザにとって利便のよいものとはいえなかった。

また、他の映像抽出技術としては、例えば“A. Merlino, D. Mor ey and M. Maybury, Broadcast news navigation using story segmentation, Proc. of ACM Multimedia 97, 1997”や特開平10-136297号公報に記載されているように、ニュースやフットボールゲームといった特定の内容ジャンルに関する非常に特殊な知識を用いるものがある。しかしながら、この従来の映像抽出技術は、目的のジャンルに関しては良好な結果を得ることができるものの他のジャンルには全く役に立たず、さらにジャンルに限定される結果、容易に一般化することができないという問題があった。

さらに、他の映像抽出技術としては、例えばU. S. Patent #5,708,767号公報に記載されているように、いわゆるストーリーユニットを抽出するものがある。しかしながら、この従来の映像抽出技術は、完全に自動化されたものではなく、どのショットが同じ内容を示す

ものであるかを決定するために、ユーザの介入が必要であった。また、この従来の映像抽出技術は、処理に要する計算が複雑であるとともに、適用対象として映像情報のみに限定されるといった問題もあった。

さらにまた、他の映像抽出技術としては、例えば特開平 9 - 2 1 4 8 7 9 号公報に記載されているように、ショット検出と無音部分検出とを組み合わせることによりショットを識別するものがある。しかしながら、この従来の映像抽出技術は、無音部分がショット境界に対応した場合のみに限定されたものであった。

また、他の映像抽出技術としては、例えば“H. Aoki, S. Shimotsuji and O. Hori, A shot classification method to select effective key-frames for video browsing, IPSJ Human Interface SIG Notes, 7:43-50, 1996”や特開平 9 - 9 3 5 8 8 号公報に記載されているように、ストーリーボードにおける表示の冗長を低減するために、反復された類似ショットを検出するものがある。しかしながら、この従来の映像抽出技術は、映像情報のみに適用できるものであり、音声情報に適用できるものではなかった。

さらに、これらのような映像抽出技術は、いわゆる局所的ビデオ構造や、特殊な知識に基づく大局的ビデオ構造しか検出することができなかった。

発明の開示

本発明は、このような実情に鑑みてなされたものであり、上述した従来の映像抽出技術の問題を解決し、種々のビデオデータにおけ

る高いレベルのビデオ構造を抽出する信号処理方法及び映像音声処理装置を提供することを目的とするものである。

上述した目的を達成する本発明にかかる信号処理方法は、供給された信号の内容の意味構造を反映するパターンを検出して解析する信号処理方法であって、信号を構成する連続したフレームのひと続きから形成されるセグメントから、その特徴を表す少なくとも1つ以上の特徴量を抽出する特徴量抽出工程と、特徴量を用いて、特徴量のそれぞれ毎に、セグメントの対の間の類似性を測定する測定基準を算出して、この測定基準によりセグメントの対の間の類似性を測定する類似性測定工程と、特徴量と測定基準とを用いて、セグメントのうち、互いに類似する複数のセグメントから構成される類似チェーンを検出する検出工程とを備えることを特徴としている。

このような本発明にかかる信号処理方法は、信号において類似したセグメントの基本的な構造パターンを検出する。

また、上述した目的を達成する本発明にかかる映像音声処理装置は、供給されたビデオ信号の内容の意味構造を反映する映像及び／又は音声のパターンを検出して解析する映像音声処理装置であって、ビデオ信号を構成する連続した映像及び／又は音声フレームのひと続きから形成される映像及び／又は音声セグメントから、その特徴を表す少なくとも1つ以上の特徴量を抽出する特徴量抽出手段と、特徴量を用いて、特徴量のそれぞれ毎に、映像及び／又は音声セグメントの対の間の類似性を測定する測定基準を算出して、この測定基準により映像及び／又は音声セグメントの対の間の類似性を測定する類似性測定手段と、特徴量と測定基準とを用いて、映像及び／又は音声セグメントのうち、互いに類似する複数の映像及び／又は

音声セグメントから構成される類似チェーンを検出する検出手段とを備えることを特徴としている。

このような本発明にかかる映像音声処理装置は、ビデオ信号において類似した映像及び／又は音声セグメントの基本的な構造パターンを決定して出力する。

図面の簡単な説明

図 1 は、本発明において適用するビデオデータの構成を説明する図であって、モデル化したビデオデータの構造を説明する図である。

図 2 は、局所的ビデオ構造を抽出する類似チェーンを説明する図である。

図 3 は、大局的ビデオ構造を抽出する類似チェーンを説明する図である。

図 4 は、本発明の実施の形態として示す映像音声処理装置の構成を説明するブロック図である。

図 5 は、同映像音声処理装置において、ビデオ構造を検出して解析する際の一連の工程を説明するフローチャートである。

図 6 は、同映像音声処理装置における動的特徴量サンプリング処理を説明する図である。

図 7 は、基本類似チェーンを説明する図である。

図 8 は、リンク類似チェーンを説明する図である。

図 9 は、周期的チェーンを説明する図である。

図 10 は、同映像音声処理装置において、バッチクラスタリング技術を用いて基本類似チェーンを検出する際の一連の工程を説明す

るフローチャートである。

図 1 1 は、非類似性閾値を説明する図である。

図 1 2 は、同映像音声処理装置において、基本類似チェーンのチェーンフィルタリングを行う際の一連の工程を説明するフローチャートである。

図 1 3 は、同映像音声処理装置において、逐次クラスタリング技術を用いて基本類似チェーンを検出する際の一連の工程を説明するフローチャートである。

図 1 4 は、同映像音声処理装置において、リンク類似チェーンを検出する際の一連の工程を説明するフローチャートである。

図 1 5 は、同映像音声処理装置において、周期的チェーンを検出する際の一連の工程を説明するフローチャートである。

図 1 6 は、同映像音声処理装置において、チェーンを用いてシーンを検出する際の一連の工程を説明するフローチャートである。

図 1 7 は、同映像音声処理装置において、チェーンを用いてニュース項目を検出する際の一連の工程を説明するフローチャートである。

図 1 8 は、同映像音声処理装置において、チェーンを用いてスポーツ放送におけるプレイを検出する際の一連の工程を説明するフローチャートである。

図 1 9 は、同映像音声処理装置において、チェーンを用いて周期検出とシーン検出とを組み合わせたトピック検出を行う際の一連の工程を説明するフローチャートである。

発明を実施するための最良の形態

以下、本発明を適用した具体的な実施の形態について図面を参照しながら詳細に説明する。

本発明を適用した実施の形態は、録画されたビデオデータから所望の内容を自動的に探し出して抽出する映像音声処理装置である。特に、この映像音声処理装置は、ビデオデータの基礎となる意味構造を反映する映像及び／又は音声の構造パターンを検出及び解析するものであり、この解析を行うために、類似チェーン（以下、必要に応じてチェーンと略記する。）という概念を導入したものである。この映像音声処理装置の具体的な説明を行う前に、ここではまず本発明において対象とするビデオデータに関する説明を行う。

本発明において対象とするビデオデータについては、図1に示すようにモデル化し、フレーム、セグメント、類似チェーンという構造を有するものとする。すなわち、ビデオデータは、最下位層において、一連のフレームにより構成される。また、ビデオデータは、フレームの1つ上の階層として、連続するフレームのひと続きから形成されるセグメントにより構成される。さらに、ビデオデータは、互いに特定の種類の類似パターンを有する一連のセグメントを類似チェーンとして構成する。

このビデオデータは、映像及び音声の両方の情報を含む。すなわち、このビデオデータにおいてフレームには、単一の静止画像である映像フレームと、一般に数十～数百ミリ秒／長といった短時間においてサンプルされた音声情報を表す音声フレームとが含まれる。

また、セグメントは、単一のカメラにより連続的に撮影された映

像フレームのひと続きから構成され、一般にはショットと呼ばれる。そして、セグメントには、映像セグメントと音声セグメントとが含まれ、ビデオ構造における基本単位となる。これらのセグメントの中で、特に音声セグメントについては、多くの定義が可能であり、例として次に示すようなものが考えられる。まず、音声セグメントは、一般によく知られている方法により検出されたビデオデータ中の無音期間により境界を定められて形成される場合がある。また、音声セグメントは、“D. Kimber and L. Wilcox, Acoustic Segmentation for Audio Browsers, Xerox Parc Technical Report”に記載されているように、例えば、音声、音楽、ノイズ、無音等のように少数のカテゴリに分類された音声フレームのひと続きから形成される場合もある。さらに、音声セグメントは、“S. Pfeiffer, S. Fischer and E. Wolfgang, Automatic Audio Content Analysis, Proceeding of ACM Multimedia 96, Nov. 1996, pp21-30”に記載されているように、2枚の連続する音声フレーム間の或る特徴における大きな変化を音声カット点として検出し、この音声カット点に基づいて決定される場合もある。

このようなビデオデータにおいて類似チェーンとは、互いに類似し、時間的に順序付けられた複数のセグメントであって、その構造パターンは、当該チェーンに含まれる類似セグメント間の関係及びチェーンの構造として満たすべき制約条件によって、幾つかの種類に分類される。形式的には、類似チェーンとは、当該類似チェーンが含むセグメントを $S_{i1}, \dots, S_{ik}$ で表したとき、全てのセグメントに関して $j = 1, \dots, k-1 : i_j < i_{j+1}$ が成り立つ一連のセグメントである。ここで、インデックス i_j は、そのセグメ

ントの元のビデオデータ内におけるセグメント番号を表し、 i への添え字 j は、そのセグメントが当該類似チェーン内において、時間軸上で j 番目に位置していることを表す。なお、類似チェーンには、時間的に不連続なセグメントが含まれるため、チェーンの要素間に時間的ギャップが存在することもある。換言すれば、セグメント $S_{i,j}$, $S_{i,j+1}$ は、元のビデオデータ内において、必ずしも連続しているとは限らない。

類似チェーンを用いることによって、ビデオデータにおいて、後述する局所的ビデオ構造と大局的ビデオ構造との両方に関する有力な手がかりを得ることができる。一般にビデオデータには、視聴者がその概要を知覚的に把握できる手掛かりが存在する。この手掛かりとして最も単純且つ重要なものは、類似する映像セグメント又は音声セグメントの構造パターンであり、この構造パターンこそ類似チェーンにより獲得すべき情報である。

このような類似チェーンとしては、後に詳述するように、基本類似チェーン、リンク類似チェーン、局所チェーン、周期的チェーンがあり、これらは、ビデオデータ解析において最も重要で基本的なものである。

ここで、基本類似チェーンとは、当該基本類似チェーンが含む全てのセグメントが互いに類似したものである。ただし、その構造パターンにおける制約はない。このような基本類似チェーンは、一般に、セグメントをグループ化するためのグルーピングアルゴリズム又はクラスタリングアルゴリズムを用いて得ることができる。また、リンク類似チェーンとは、そのチェーン内において隣接するセグメントが互いに類似したものである。さらに、局所チェーンとは、隣

接するセグメントの各対において、セグメント間の時間間隔が所定の時間よりも小さいものである。そして、周期的チェーンとは、各セグメントが、それよりも m 番目後方のセグメントと類似したものである。すなわち、周期的チェーンは、 m 個のセグメントが近似的に繰り返されることで構成される。

そして、このような類似チェーンは、以下に示すように、ビデオデータにおける例えばシーンといった局所的ビデオ構造や、例えばニュース項目といった大局的ビデオ構造を抽出するのに用いることができる。

ここで、シーンとは、ビデオデータを、その意味内容に基づいて、より高いレベルで記述するために、映像セグメント（ショット）検出或いは音声セグメント検出により得られたセグメントを、例えばセグメント内の知覚的アクティビティ量といったセグメントの特徴を表す特徴量を用いて意味のあるまとまりにグループ化したものである。シーンは、主観的なものであり、ビデオデータの内容或いはジャンルに依存するが、ここでは、その特徴量が互いに類似性を示す映像セグメント又は音声セグメントの反復パターンをグループ化したものとする。

さて、上述した局所的ビデオ構造を抽出する類似チェーンの具体例として、図2に示すように、2人の話し手が互いに会話している場面において、映像セグメントが、話し手に応じて交互に現れる場合を考える。このような反復パターンを有するビデオデータにおいて、各映像セグメントは、A成分及びB成分の各成分毎に、2つの交差するチェーンにより構成される。そのため、一般に、このような交差する局所チェーンは、関連する映像セグメントのグループ或

いはシーンを検出するのに用いることができる。

また、上述した大局的ビデオ構造を抽出する類似チェーンの具体例として、図3に示すように、固定構造を有するニュース番組を考える。このようなビデオデータにおいては、まず、各ニュース項目毎にニュースキャスターが項目を紹介するセグメントが出現し、それに続いて、例えば現地から特派員がリポートするセグメントが出現する。このような固定構造を有するビデオデータにおいては、繰り返し出現するニュースキャスターの映像セグメントは、大局的チェーンを構成する。ここで、ニュースキャスターのセグメントは、各ニュース項目の開始部を示すため、大局的チェーンを用いることによって、ニュース項目を自動的に検出することができる。すなわち、大局的チェーンを用いることによって、同図において、トピックA、B、C、D、・・・といった複数のニュース項目から構成されるビデオデータの中から、各トピックを検出することができる。

本発明を適用した実施の形態として図4に示す映像音声処理装置10は、上述したビデオデータにおけるセグメントの特徴量を用いてセグメント間の類似性を測定し、上述した類似チェーンを自動的に検出するものであり、映像セグメント及び音声セグメントの両方に適用できるものである。そして、映像音声処理装置10は、類似チェーンを解析することによって、ビデオデータから、局所的ビデオ構造であるシーンや、大局的ビデオ構造であるトピック等の高レベルの構造を抽出・再構成することができる。

映像音声処理装置10は、同図に示すように、入力したビデオデータのストリームを映像、音声又はこれらの両方のセグメントに分割するビデオ分割部11と、ビデオデータの分割情報を記憶するビ

デオセグメントメモリ 12 と、各映像セグメントにおける特徴量を抽出する特徴量抽出手段である映像特徴量抽出部 13 と、各音声セグメントにおける特徴量を抽出する特徴量抽出手段である音声特徴量抽出部 14 と、映像セグメント及び音声セグメントの特徴量を記憶するセグメント特徴量メモリ 15 と、映像セグメント及び音声セグメントをチェーンにまとめる検出手段であるチェーン検出部 16 と、2つのセグメント間の類似性を測定する類似性測定手段である特徴量類似性測定部 17 と、種々のビデオ構造を検出する解析手段であるチェーン解析部 18 とを備える。

ビデオ分割部 11 は、例えば、MPEG1 (Moving Picture Experts Group phase 1) やMPEG2 (Moving Picture Experts Group phase 2)、或いはいわゆるDV (Digital Video) のような圧縮ビデオデータフォーマットを含む種々のデジタル化されたフォーマットにおける映像データと音声データとからなるビデオデータのストリームを入力し、このビデオデータを映像、音声又はこれらの両方のセグメントに分割するものである。このビデオ分割部 11 は、入力したビデオデータが圧縮フォーマットであった場合、この圧縮ビデオデータを完全伸張することなく直接処理することができる。ビデオ分割部 11 は、入力したビデオデータを処理し、映像セグメントと音声セグメントとに分割する。また、ビデオ分割部 11 は、入力したビデオデータを分割した結果である分割情報を後段のデオセグメントメモリ 12 に供給する。さらに、ビデオ分割部 11 は、映像セグメントと音声セグメントとに応じて、分割情報を後段の映像特徴量抽出部 13 及び音声特徴量抽出部 14 に供給する。

デオセグメントメモリ 12 は、ビデオ分割部 11 から供給され

たビデオデータの分割情報を記憶する。また、ビデオセグメントメモリ 12 は、後述するチェーン検出部 16 からの問い合わせに応じて、分割情報をチェーン検出部 16 に供給する。

映像特徴量抽出部 13 は、ビデオ分割部 11 によりビデオデータを分割して得た各映像セグメント毎の特徴量を抽出する。映像特徴量抽出部 13 は、圧縮映像データを完全伸張することなく直接処理することができる。映像特徴量抽出部 13 は、抽出した各映像セグメントの特徴量を後段のセグメント特徴量メモリ 15 に供給する。

音声特徴量抽出部 14 は、ビデオ分割部 11 によりビデオデータを分割して得た各音声セグメント毎の特徴量を抽出する。音声特徴量抽出部 14 は、圧縮音声データを完全伸張することなく直接処理することができる。音声特徴量抽出部 14 は、抽出した各音声セグメントの特徴量を後段のセグメント特徴量メモリ 15 に供給する。

セグメント特徴量メモリ 15 は、映像特徴量抽出部 13 及び音声特徴量抽出部 14 からそれぞれ供給された映像セグメント及び音声セグメントの特徴量を記憶する。セグメント特徴量メモリ 15 は、後述する特徴量類似性測定部 17 からの問い合わせに応じて、記憶している特徴量やセグメントを特徴量類似性測定部 17 に供給する。

チェーン検出部 16 は、ビデオセグメントメモリ 12 に保持された分割情報と、1 対のセグメント間の類似性とを用いて、映像セグメント及び音声セグメントをそれぞれチェーンにまとめる。チェーン検出部 16 は、グループ内の各セグメントから開始して、セグメント群の中から類似しているセグメントの反復パターンを検出し、このようなセグメントをチェーンにまとめていく。このチェーン検出部 16 は、チェーンの初期候補をまとめた後、第 2 のフィルタリ

ング段階を用いてチェーンの最終セットを決定する。そして、チェーン検出部 16 は、検出したチェーンを後段のチェーン解析部 18 に供給する。

特徴量類似性測定部 17 は、2つのセグメント間の類似性を測定する。特徴量類似性測定部 17 は、或るセグメントに関する特徴量を検索するようにセグメント特徴量メモリ 15 に問いかける。

チェーン解析部 18 は、チェーン検出部 16 により検出されたチェーン構造を解析し、種々の局所的ビデオ構造及び大局的ビデオ構造を検出する。このチェーン解析部 18 は、後述するように、その細部を特定のアプリケーションに合わせて調整することができる。

このような映像音声処理装置 10 は、類似チェーンを用いて図 5 に概略を示すような一連の処理を行うことによって、ビデオ構造を検出する。

まず、映像音声処理装置 10 は、同図に示すように、ステップ S1 において、ビデオ分割を行う。すなわち、映像音声処理装置 10 は、ビデオ分割部 11 に入力されたビデオデータを映像セグメント又は音声セグメントのいずれか、或いは可能であればその両方に分割する。映像音声処理装置 10 は、適用するビデオ分割方法に特に前提要件を設けない。例えば、映像音声処理装置 10 は、“G. Ahanger and T.D.C. Little, A survey of technologies for parsing and indexing digital video, J. of Visual Communication and Image Representation 7:28-4, 1996” に記載されているような方法によりビデオ分割を行う。このようなビデオ分割の方法は、当該技術分野ではよく知られたものであり、映像音声処理装置 10 は、いかなるビデオ分割方法も適用できるものとする。

続いて、映像音声処理装置 10 は、ステップ S 2 において、特徴量の抽出を行う。すなわち、映像音声処理装置 10 は、映像特徴量抽出部 13 や音声特徴量抽出部 14 によって、そのセグメントの特徴を表す特徴量を計算する。映像音声処理装置 10 においては、例えば、各セグメントの時間長、カラーヒストグラムやテクスチャフィーチャといった映像特徴量や、周波数解析結果、レベル、ピッチといった音声特徴量や、アクティビティ測定結果等が、適用可能な特徴量として計算される。勿論、映像音声処理装置 10 は、適用可能な特徴量としてこれらに限定されるものではない。

続いて、映像音声処理装置 10 は、ステップ S 3 において、特徴量を用いたセグメントの類似性測定を行う。すなわち、映像音声処理装置 10 は、特徴量類似性測定部 17 により非類似性測定を行い、その測定基準によって、2つのセグメントがどの程度類似しているかを測定する。映像音声処理装置 10 は、先のステップ S 2 において抽出した特徴量を用いて、非類似性測定基準を計算する。

続いて、映像音声処理装置 10 は、ステップ S 4 において、チェーンの検出を行う。すなわち、映像音声処理装置 10 は、先のステップ S 3 において計算した非類似性測定基準と、先のステップ S 2 において抽出した特徴量とを用いて、類似したセグメントのチェーンを検出する。

そして、映像音声処理装置 10 は、ステップ S 5 において、チェーンの解析を行う。すなわち、映像音声処理装置 10 は、先のステップ S 4 において検出したチェーンを用いて、ビデオデータの局所的ビデオ構造及び／又は大局的ビデオ構造を決定して出力する。

このような一連の処理を経ることによって、映像音声処理装置 1

0 は、ビデオデータからビデオ構造を検出することができる。したがって、ユーザは、この結果を用いることによって、ビデオデータの内容の索引付けや要約を行ったり、ビデオデータ中の興味のあるポイントに迅速にアクセスしたりすることが可能となる。

以下、同図に示した映像音声処理装置 10 における処理を各工程毎により詳細に説明していく。

まず、ステップ S 1 におけるビデオ分割について説明する。映像音声処理装置 10 は、ビデオ分割部 11 に入力されたビデオデータを映像セグメント又は音声セグメントのいずれか、或いは可能であればその両方に分割するが、このビデオデータにおけるセグメントの境界を自動的に検出するための技術は多くのものがあり、当該映像音声処理装置 10 において、このビデオ分割方法に特別な前提要件を設けないことは上述した通りである。一方、映像音声処理装置 10 において、後の工程によるチェーン検出の精度は、本質的に、基礎となるビデオ分割の精度に依存する。

つぎに、ステップ S 2 における特徴量抽出について説明する。特徴量とは、セグメントの特徴を表すとともに、異なるセグメント間の類似性を測定するためのデータを供給するセグメントの属性である。映像音声処理装置 10 は、映像特徴量抽出部 13 や音声特徴量抽出部 14 により各セグメントの特徴量を計算し、セグメントの特徴を表す。映像音声処理装置 10 は、いかなる特徴量の具体的詳細にも依存するものではないが、当該映像音声処理装置 10 において用いて効果的であると考えられる特徴量としては、例えば以下に示す映像特徴量、音声特徴量、映像音声共通特徴量のようなものがある。映像音声処理装置 10 において適用可能となるこれらの特徴量

の必要条件は、非類似性の測定が可能であることである。さらに、これらの特徴量は、映像音声処理装置 10 が効率化のために特徴量抽出と上述したビデオ分割とを同時に行うことを可能とする必要がある。以下に説明する特徴量は、これらの必要条件を満たすものである。

特徴量としては、まず映像に関するものが挙げられる。以下では、これを映像特徴量と称することにする。映像セグメントは、連続する映像フレームにより構成されるため、映像セグメントから適切な映像フレームを抽出することによって、その映像セグメントの描写内容を、抽出した映像フレームで代表して表現することが可能である。すなわち、映像セグメントの類似性は、適切に抽出された映像フレームの類似性で代替可能である。このことから、映像特徴量は、映像音声処理装置 10 で用いることができる重要な特徴量の 1 つである。この場合の映像特徴量は、単独では静的な情報しか表せないが、映像音声処理装置 10 は、後述するような方法を適用することによって、この映像特徴量に基づく映像セグメントの動的な特徴を抽出することもできる。

映像音声処理装置 10 において、映像における色は、2 つの映像が類似しているかを判断する際の重要な材料となる。カラーヒストグラムを用いて映像の類似性を判断することは、例えば “G. Ahanger and T.D.C. Little, A survey of technologies for parsing and indexing digital video, J. of Visual Communication and Image Representation 7:28-4, 1996” に記載されているように、よく知られている。ここで、カラーヒストグラムとは、例えば HSV や RGB 等の 3 次元色空間を n 個の領域に分割し、映像における画素

の、各領域での出現頻度の相対的割合を計算したものである。そして、得られた情報からは、 n 次元ベクトルが与えられる。圧縮されたビデオデータに関しても、例えばU.S. Patent #5,708,767号公報に記載されているように、カラーヒストグラムを、圧縮データから直接抽出することができる。

映像音声処理装置10では、セグメントを構成する映像におけるもともとのYUV色空間を、色チャンネル当たり2ビットでサンプルして構成した、 $2^{2 \times 3} = 64$ 次元のヒストグラムベクトルを用いている。

このようなヒストグラムは、映像の全体的な色調を表すが、これには時間情報が含まれていない。そこで、映像音声処理装置10においては、もう1つの映像特徴量として、映像相関を計算する。映像音声処理装置10におけるチェーン検出において、複数の類似セグメントが互いに交差した構造は、それがまとまった1つのチェーン構造であることを示す有力な指標となる。例えば会話場面において、カメラの位置は、2人の話し手の間を交互に移動するが、カメラは通常、同一の話し手を再度撮影するときには、ほぼ同じ位置に戻る。このような場合における構造を検出するためには、グレイスケールの縮小映像に基づく相関がセグメントの類似性の良好な指標となることを見出したことから、映像音声処理装置10では、元の映像を $M \times N$ の大きさのグレイスケール映像へ間引き縮小し、これを用いて映像相関を計算する。ここで、 M と N は、両方とも小さな値で十分であり、例えば 8×8 である。すなわち、これらの縮小グレイスケール映像は、 MN 次元の特徴量ベクトルとして解釈される。

さらに上述した映像特徴量とは異なる特徴量としては、音声に関

するものが挙げられる。以下では、この特徴量を音声特徴量と称することにする。音声特徴量とは、音声セグメントの内容を表すことができる特徴量であり、映像音声処理装置 10 は、この音声特徴量として、周波数解析、ピッチ、レベル等を用いることができる。これらの音声特徴量は、種々の文献により知られているものである。

まず、映像音声処理装置 10 は、フーリエ変換等の周波数解析を行うことによって、単一の音声フレームにおける周波数情報の分布を決定することができる。映像音声処理装置 10 は、例えば、1 つの音声セグメントにわたる周波数情報の分布を表すために、FFT (Fast Fourier Transform; 高速フーリエ変換) 成分、周波数ヒストグラム、パワースペクトル、その他の特徴量を用いることができる。

また、映像音声処理装置 10 は、平均ピッチや最大ピッチ等のピッチや、平均ラウドネスや最大ラウドネス等の音声レベルもまた、音声セグメントを表す有効な音声特徴量として用いることができる。

さらに、映像音声処理装置 10 は、ケプストラム特徴量として、ケプストラム係数とその 1 次及び 2 次微分係数とを含み、FFT スペクトル又は LPC (Linear Predictive Coding; 線形予測符号化) 等から得られたケプストラムスペクトル係数を用いることもできる。

さらに他の特徴量としては、映像音声共通特徴量が挙げられる。これは、映像特徴量でもなく音声特徴量でもないが、映像音声処理装置 10 において、チェーン内のセグメントの特徴を表すのに有用な情報を与えるものである。映像音声処理装置 10 は、この映像音声共通特徴量として、アクティビティを用いる。

アクティビティとは、セグメントの内容がどの程度動的或いは静的であるように感じられるかを表す指標である。例えば、視覚的に動的である場合、アクティビティは、カメラが対象物に沿って迅速に移動する度合い若しくは撮影されているオブジェクトが迅速に変化する度合いを表す。

このアクティビティは、カラーヒストグラムのような特徴量のフレーム間非類似性の平均値を測定することによって、間接的に計算される。ここで、フレーム i とフレーム j との間で測定された特徴量 F に対する非類似性測定基準を $d_F(i, j)$ と定義すると、映像アクティビティ V_F は、次式 (1) のように定義される。

$$V_F = \frac{\sum_{i=b}^{f-1} d_F(i, i+1)}{f-b} \quad \dots (1)$$

式 (1) において、 b と f は、それぞれ、1 セグメントにおける最初と最後のフレームのフレーム番号である。映像音声処理装置 10 は、具体的には、例えば上述したヒストグラムを用いて、映像アクティビティ V_F を計算することができる。

ところで、上述した映像特徴量を始めとする特徴量は、基本的にはセグメントの静的情報を表すものであることは上述した通りであるが、セグメントの特徴を正確に表すためには、動的情報をも考慮する必要がある。そこで、映像音声処理装置 10 は、以下に示すような特徴量のサンプリング方法により動的情報を表すこととする。

映像音声処理装置 10 は、例えば図 6 に示すように、1 セグメン

ト内の異なる時点から 1 以上の静的な特徴量を抽出する。このとき、映像音声処理装置 10 は、特徴量の抽出数を、そのセグメント表現における忠実度の最大化とデータ冗長度の最小化とのバランスをとることにより決定する。例えば、セグメント内の或る 1 画像が当該セグメントのキーフレームとして指定可能な場合には、そのキーフレームから計算されたヒストグラムが、抽出すべきサンプリング特徴量となる。

ところで、或るサンプルが常に所定の時点、例えばセグメント内の最後の時点において選択される場合を考える。この場合、黒フレームへ変化 (fade) していく任意の 2 つのセグメントについては、サンプルが同一の黒フレームとなるため、同一の特徴量が得られる結果になる恐れがある。すなわち、これらのセグメントの映像内容がいかなるものであれ、選択した 2 つのフレームは、極めて類似していると判断されてしまう。このような問題は、サンプルが良好な代表値でないために発生するものである。

そこで、映像音声処理装置 10 は、このように固定点で特徴量を抽出するのではなく、セグメント全体における統計的な代表値を抽出することとする。ここでは、一般的な特徴量のサンプリング方法を 2 つの場合、すなわち、(1) 特徴量を実数の n 次元ベクトルとして表すことができる場合と、(2) 非類似性測定基準しか利用できない場合とについて説明する。なお、(1) には、ヒストグラムやパワースペクトル等、最もよく知られている映像特徴量及び音声特徴量が含まれる。

(1) においては、サンプル数は、事前に k と決められており、映像音声処理装置 10 は、“L. Kaufman and P.J. Rousseeuw, Fin

ding Groups in Data:An Introduction to Cluster Analysis, John-Wiley and sons, 1990”に記載されてよく知られているk平均値クラスタリング法(k-means-clustering method)を用いて、セグメント全体についての特徴量をk個の異なるグループに自動的に分割する。そして、映像音声処理装置10は、サンプル値として、k個の各グループから、グループの重心値(centroid)又はこの重心値に近いサンプルを選択する。映像音声処理装置10におけるこの処理の複雑度は、サンプル数に関して単に直線的に増加するにとどまる。

一方、(2)においては、映像音声処理装置10は、“L. Kaufman and P.J. Rousseeuw, Finding Groups in Data:An Introduction to Cluster Analysis, John-Wiley and sons, 1990”に記載されているk-メドイドアルゴリズム法(k-medoids algorithm method)を用いて、k個のグループを形成する。そして、映像音声処理装置10は、サンプル値として、k個のグループ毎に、上述したグループのメドイド(medoid)を用いる。

なお、映像音声処理装置10においては、抽出された動的特徴を表す特徴量についての非類似性測定基準を構成する方法は、その基礎となる静的な特徴量の非類似性測定基準に基づくが、これについては後述する。

このようにして、映像音声処理装置10は、静的な特徴量を複数抽出し、これらの複数の静的な特徴量を用いることによって、動的特徴を表すことができる。

以上のように、映像音声処理装置10は、種々の特徴量を抽出することができる。これらの各特徴量は、一般に、単一ではセグメン

トの特徴を表すのに不十分であることが多い。そこで、映像音声処理装置 10 は、これらの各種特徴量を組み合わせることで、互いに補完し合う特徴量の組を選択することができる。例えば、映像音声処理装置 10 は、上述したカラーヒストグラムと映像相関とを組み合わせることによって、各特徴量が有する情報よりも多くの情報を得ることができる。

つぎに、図 5 中ステップ S 3 における特徴量を用いたセグメントの類似性測定について説明する。映像音声処理装置 10 は、2 つの特徴量について、それがどの程度非類似であるかを測定する実数値を計算する関数である非類似性測定基準を用いて、特徴量類似性測定部 17 によりセグメントの類似性測定を行う。この非類似性測定基準は、その値が小さい場合は 2 つの特徴量が類似していることを示し、値が大きい場合は非類似であることを示す。ここでは、特徴量 F に関する 2 つのセグメント S_1 、 S_2 の非類似性を計算する関数を非類似性測定基準 $d_F(S_1, S_2)$ と定義する。このような関数は、以下の式 (2) で与えられる関係を満足する。

$$\begin{aligned} d_F(S_1, S_2) &= 0 && (S_1 = S_2 \text{ のとき}) \\ d_F(S_1, S_2) &\geq 0 && (\text{全ての } S_1, S_2 \text{ について}) \\ d_F(S_1, S_2) &= d_F(S_2, S_1) && (\text{全ての } S_1, S_2 \text{ について}) \end{aligned} \quad \dots (2)$$

ところで、非類似性測定基準の中には、或る特定の特徴量にのみ適用可能なものがあるが、“G. Ahanger and T.D.C. Little, A su

urvey of technologies for parsing and indexing digital video, J. of Visual Communication and Image Representation 7:28-4, 1996” や “L. Kaufman and P.J. Rousseeuw, Finding Groups in Data: An Introduction to Cluster Analysis, John-Wiley and sons, 1990” に記載されているように、一般には、多くの非類似性測定基準は、 n 次元空間における点として表される特徴量についての類似性を測定するのに適用可能である。その具体例は、ユークリッド距離、内積、 L_1 距離等である。ここで、特に L_1 距離が、ヒストグラムや映像相関等の特徴量を含む種々の特徴量に対して有効に作用することから、映像音声処理装置 10 は、 L_1 距離を導入する。ここで、2つの n 次元ベクトルを A , B とした場合、 A , B 間の L_1 距離 $d_{L_1}(A, B)$ は、次式 (3) で与えられる。

$$d_{L_1}(A, B) = \sum_{i=1}^n |A_i - B_i| \quad \dots (3)$$

ここで、下付文字 i は、 n 次元ベクトル A , B のそれぞれの i 番目の要素を示すものである。

また、映像音声処理装置 10 は、上述したように、動的特徴を表す特徴量として、セグメントにおける様々な時点での静的な特徴量を抽出する。そして、映像音声処理装置 10 は、抽出された2つの動的特徴量の間の類似性を決定するために、その非類似性測定基準として、その基礎となる静的特徴量の間の非類似性測定基準を用いる。これらの動的特徴量の非類似性測定基準は、多くの場合、各動的特徴量から選択された最も類似した静的特徴量の対の非類似性値

を用いて決定されるのが最良である。この場合、2つの抽出された動的特徴量 SF_1 , SF_2 の間の非類似性測定基準は、次式 (4) のように定義される。

$$d_s(SF_1, SF_2) = \min_{F_1 \in SF_1, F_2 \in SF_2} d_F(F_1, F_2) \quad \dots (4)$$

ここで、上式 (4) における関数 $d_F(F_1, F_2)$ は、その基礎となる静的特徴量 F についての非類似性測定基準を示す。なお、場合によっては、特徴量の非類似性の最小値をとる代わりに、最大値又は平均値をとってもよい。

ところで、映像音声処理装置 10 は、セグメントの類似性を決定する上で、単一の特徴量だけでは不十分であり、同一セグメントに関する多数の特徴量からの情報を組み合わせることを必要とする場合も多い。この1つの方法として、映像音声処理装置 10 は、種々の特徴量に基づく非類似性を、それぞれの特徴量の重み付き組み合わせとして計算する。すなわち、映像音声処理装置 10 は、 k 個の特徴量 $F_1, F_2, \dots, F_k$ が存在する場合、次式 (5) に表される組み合わせた特徴量に関する非類似性測定基準 $d_F(S_1, S_2)$ を用いる。

$$d_F(S_1, S_2) = \sum_{i=1}^k w_i d_{F_i}(S_1, S_2) \quad \dots (5)$$

ここで、 $\{w_i\}$ は、 $\sum_i w_i = 1$ となる重み係数である。

以上のように、映像音声処理装置 10 は、図 5 中ステップ S 2 において抽出された特徴量を用いて非類似性測定基準を計算し、当該セグメント間の類似性を測定することができる。

つぎに、図 5 中ステップ S 4 におけるチェーン検出について説明する。映像音声処理装置 10 は、非類似性測定基準と抽出した特徴量とを用いて、類似セグメント間のつながりを表す類似チェーンを検出する。ここでは、まず、幾つかのタイプの類似チェーンを定義し、各タイプの類似チェーンを検出するためのアルゴリズムについて具体的に説明する。

ところで、以下に定義される類似チェーンのタイプは、それぞれ互いに独立したものであるため、映像音声処理装置 10 においては、1 つのチェーンが複数のタイプに属することが可能である。ここでは、このようなチェーンを、定義したタイプ名を組み合わせで称することにする。例えば、局所均一リンクチェーンは、後述するように、局所的であって均一でありリンク類似チェーンのことを示す。

さて、類似チェーンのタイプは、当該類似チェーンが含む類似セグメント間の関係に制約を有するものと、当該類似チェーンの構造に制約を有するものとに大別される。なお、以下の定義において、チェーン C とは、一連のセグメント $S_{i1}, \dots, S_{im}$ を表すこととする。ここで、インデックス i_k は、そのセグメントの、元のビデオデータ内におけるセグメント番号を表し、また i への添え字 k は、そのセグメントが当該類似チェーン内において、時間軸上で k 番目に位置していることを表す。また、これらの一連のセグメントは、常に時間軸上において順序付けられているものとし、全ての k

$= 1, \dots, m-1$ について $i_k < i_{k+1}$ である。さらに、 $|C|$ は、チェーンの長さを表し、 C^{start} 及び C^{end} は、それぞれ、ビデオデータにおけるチェーン C の開始時刻及び終了時刻を表すものとする。より正確には、チェーン C の開始時刻は、チェーン C における最初のセグメントの開始時刻であり、チェーン C の終了時刻は、チェーン C における最後のセグメントの終了時刻である。さらにまた、或るセグメントを A とした場合、その類似セグメントを、 A' , A'' , A''' , $\dots$ で表す。最後に、2つのセグメントが類似しているとは、それらの非類似性測定基準が、後述する非類似性閾値よりも小さい状態であることとし、これを $similar(S_1, S_2)$ で表す。

当該類似チェーンが含む類似セグメント間の関係に制約を有する類似チェーンとしては、基本類似チェーン、リンク類似チェーン、周期的チェーンがある。

まず、基本類似チェーンであるが、これは、図7に示すように、全てのセグメントが互いに類似したチェーン C である。なお、基本類似チェーンに構造的制約はない。この基本類似チェーンは、多くの場合、類似セグメントをグループ化するためのグルーピングアルゴリズム又はクラスタリングアルゴリズムの結果として得られるものである。

一方、リンク類似チェーンとは、図8に示すように、隣接するセグメントが互いに類似したチェーン C である。すなわち、リンク類似チェーンでは、全ての $k = 1, \dots, |C| - 1$ について、 $similar(S_k, S_{k+1})$ である。このリンク類似チェーンは、上述した類似セグメントの定義から、 A' , A'' , A''' , $\dots$

・ ・ と記述することができる。

さらに、周期的チェーンとは、図 9 に示すように、各セグメントが、その後方 m 番目のセグメントと類似したチェーン C_{cyclic} である。すなわち、周期的チェーンでは、全ての $k = 1, \dots, |C_{cyclic}| - 1$ について、 $similar(S_k, S_{k+1})$ である。換言すれば、周期的チェーンは、 m 個の一連のセグメントの近似的な繰り返しとして構成される。これより、周期的チェーンは、 $S_1, S_2, \dots, S_m, S_1', S_2', \dots, S_m', S_1'', S_2'', \dots, S_m'', \wedge, S_1''', S_2''', \dots, S_m'''$ と記述することができる。

一方、構造的制約を有する類似チェーンとしては、局所チェーン、均一チェーンがある。

ここで、局所チェーンとは、上述したように、隣接するセグメントの各対において、セグメント間の時間間隔が所定の時間よりも小さいチェーン C である。すなわち、局所チェーンでは、チェーン内の 2 つのセグメント間において許容される時間間隔の最大値を gap と表すと、全ての $k = 1, \dots, |C| - 1$ について、隣接するセグメント $S_{i_k}, S_{i_{k+1}}$ に対して、 $i_{k+1} - i_k \leq gap$ である。

また、チェーン内のセグメントがほぼ等しい時間間隔で現れる場合、これは重要なビデオ構造の有力な指標となりうるが、このようなチェーン C を均一チェーンと定義する。ここで、チェーン C の均一性 $uniformity(C)$ を、次式 (6) に示すように、等間隔時間からの時間間隔のずれの平均値を、そのチェーンの長さで規格化したものとして定義する。

$$uniformity(C) = \frac{\sum_{i=1}^{|C|-1} \left| (S_{i+1}^{start} - S_i^{start}) - \frac{(C^{end} - C^{start})}{|C|} \right|}{|C| \cdot |C^{end} - C^{start}|} \quad \dots (6)$$

上式(6)で示されるチェーンCの均一性 *uniformity* (C) は、0 から 1 の範囲の値をとり、その値が小さい場合、セグメントの時間間隔分布が均一な分布に近いことを示す。この均一性 *uniformity* (C) の値が所定の均一性閾値よりも小さい場合、チェーンCを均一チェーンとみなす。

以下、映像音声処理装置10において、このような各種チェーンのそれぞれを検出するための処理について説明する。

映像音声処理装置10は、上述した基本類似チェーンを検出するために、バッチクラスタリング技術或いは逐次クラスタリング技術を用いる。

バッチクラスタリング技術とは、チェーンを一括して検出する技術である。ただし、この技術を適用するためには、チェーン検出を行う前に、全てのビデオ分割を終了しておく必要がある。一方の逐次クラスタリング技術は、チェーンを逐次的に検出していく技術であり、もしビデオ分割及び特徴量抽出のまた逐次的に行われるならば、ビデオデータを再生しつつ逐次的にビデオ解析を行うことが可能となる。さらには、もし映像音声処理装置10に十分な計算能力があるならば、この逐次的チェーン検出を実時間、換言すれば、ビ

デオデータを取込又は記録すると同時にチェーンを検出していくことができる。しかしながら、逐次的なビデオ解析は、その精度に問題を生じることがある。すなわち、逐次的な方法の場合、最適チェーン構造を決定するための大局的な情報がなく、さらにはセグメントの入力順序に敏感であるため、低品質の結果を生じることがある。

映像音声処理装置 10 は、バッチクラスタリング技術を用いる場合には、図 10 に示すように、2つの工程を経ることによって、基本類似チェーンを検出する。

まず、映像音声処理装置 10 は、ステップ S 11 において、候補チェーンの検出を行う。すなわち、映像音声処理装置 10 は、ビデオデータにおける類似セグメントを検出し、クラスタとしてまとめる。これにより得られたセグメントのクラスタ群は、基本類似チェーンを検出する上での初期候補となる。

映像音声処理装置 10 は、類似チェーンの初期候補を求める際、任意のクラスタリング技術を用いることができるが、ここでは、“L. Kaufman and P. J. Rousseeuw, Finding Groups in Data: An Introduction to Cluster Analysis, John-Wiley and sons, 1990”に記載されている階層的クラスタリング方法 (hierarchical clustering method) を用いることにする。このアルゴリズムは、まず、最も類似した 2つのセグメントを 1つの対としてまとめることにより始まり、クラスタ間の類似性測定基準を用いて、各段階で最も類似したクラスタの対を次々とまとめていく。このアルゴリズムにおいて、2つのクラスタ C_1 , C_2 間の非類似性測定基準 $d_c(C_1, C_2)$ を、次式 (7) に示すように、それぞれのクラスタに含まれる 2つのセグメント間の最小非類似性として定義する。

$$d_c(C_1, C_2) = \min_{s_1 \in C_1, s_2 \in C_2} d_s(s_1, s_2) \quad \dots (7)$$

なお、映像音声処理装置 10 においては、必要に応じて、上式 (7) で示される最小関数の代わりに、最大関数又は平均関数を用いてもよい。

ところで、この階層的クラスタリング法は、仮に何らの制約のない場合、ビデオデータに含まれる全てのセグメントを単一のグループにまとめてしまう。そこで、映像音声処理装置 10 は、図 11 に示すように、非類似性閾値 δ_{sim} を導入し、この非類似性閾値 δ_{sim} との比較によって、或るセグメントが他方のセグメントと類似であるか否かを判断する。ここで、非類似性閾値 δ_{sim} とは、同図に示すように、2 つのセグメントがどの程度類似している場合に同一のチェーンに属するものとみなすかを決定する閾値である。そして、映像音声処理装置 10 は、全クラス対の非類似性がこの非類似性閾値 δ_{sim} を超えない範囲において、セグメントをクラスにまとめていく。

なお、映像音声処理装置 10 は、非類似性閾値 δ_{sim} をユーザにより設定するようにしてもよく、自動的に決定してもよい。ただし、非類似性閾値 δ_{sim} として固定値を用いる場合には、その最適値は、ビデオデータの内容に依存することとなる。例えば、変化に富んだ映像内容を有するビデオデータの場合、非類似性閾値 δ_{sim} は、高い値に設定される必要がある。一方、変化が少ない映像内容を有するビデオデータの場合、非類似性閾値 δ_{sim} は、低い値に設定される必

要がある。ここで一般に、非類似性閾値 δ_{sim} が高い場合には、検出されるクラスタ数は少なくなり、非類似性閾値 δ_{sim} が低い場合には、検出されるクラスタ数は多くなるという性質がある。

これより、映像音声処理装置 10 においては、適切な非類似性閾値 δ_{sim} を決定することが、その性能を左右する上で重要となる。そのため、映像音声処理装置 10 においては、非類似性閾値 δ_{sim} をユーザにより設定する場合には、上述したことを考慮した上で設定する必要がある。一方、映像音声処理装置 10 は、以下に示す方法により、有効な非類似性閾値 δ_{sim} を自動的に決定することもできる。

例えば、その 1 つの方法として、映像音声処理装置 10 は、

(n) (n - 1) / 2 個のセグメント対の間の非類似性の分布における平均値やメジアン（中央値）といった統計量を用いて、非類似性閾値 δ_{sim} を得ることができる。いま、全てのセグメント対における非類似性の平均値とその標準偏差をそれぞれ μ , σ とした場合、非類似性閾値 δ_{sim} は、 $a\mu + b\sigma$ の形式で表すことができる。ここで、 a 及び b は定数であり、それぞれ、0.5 及び 0.1 に設定することが良好な結果を与えることを見出している。

実用上においては、映像音声処理装置 10 は、全てのセグメント対について、それらの間の非類似性を求める必要はなく、その平均値 μ 及び標準偏差 σ が真値に十分近い結果を与えるに足りるセグメント対を、全セグメント対集合からランダムに選択し、その非類似性を求めればよい。映像音声処理装置 10 は、このようにして得られた平均値 μ 及び標準偏差 σ を用いることによって、適切な非類似性閾値 δ_{sim} を自動的に得ることができる。すなわち、映像音声処理装置 10 は、例えば、セグメント対の全数を n 、任意の小さい定数

をCとした場合、C_nで与えられる数のセグメント対の非類似性を抽出することによって、適切な非類似性閾値 δ_{sim} を自動的に決定することができる。

映像音声処理装置10は、これまでに示したようにセグメントのクラスタリングを行った後、各クラスタにて、当該各クラスタに含まれるセグメントを並べ替えることによって、基本類似チェーンの初期候補を得ることができる。

ところで、図10中ステップS11において検出したチェーン候補は、その多くが、実際のビデオ構造とは無関係のものである。これより、映像音声処理装置10は、どのチェーン候補がビデオ構造の骨格をなす重要なチェーンであるか、或いは、ビデオ構造に関連するチェーンであるかを決定する必要がある。そのため、映像音声処理装置10は、ステップS12において、チェーンの品質を示す数的基準に対応する品質測定基準を用いたチェーンフィルタリングを行う。すなわち、映像音声処理装置10は、ビデオ構造解析におけるチェーン候補の重要性及び関連性を測定し、所定の品質測定基準閾値を上回るチェーン候補のみをチェーン検出の結果として出力する。ここで、フィルタリングで使用される関連性測定関数として最も単純な例は、チェーン候補が受け入れられるか否かを示すブール関数であるが、映像音声処理装置10は、必要に応じて、より複雑な関連性測定関数を用いてもよい。

ところで、映像音声処理装置10においては、チェーン品質測定基準として、チェーン長、チェーン密度、チェーン強度等が用いられる。

まず、チェーン長であるが、これは、1つのチェーンが保有する

セグメントの数と定義される。ここで、映像音声処理装置 10 が、このチェーン長を、そのチェーン品質測定基準として用いることができるのは、一般にチェーン長が小さい場合であり、それは通常ノイズとしてみなすことが可能であることに依る。例えば、或るチェーンが単一セグメントしか有していない場合、それは何らの情報を有していない。すなわち、チェーン長に基づく品質測定基準では、その制約として、チェーンが保有すべきセグメント数の最小値が与えられることとなる。

次に、チェーン密度であるが、これは、或るチェーンが保有する全セグメント数と、そのチェーンが占めるビデオデータの部分領域における全セグメント数との比として定義される。これは、チェーンが限られた時間領域内に集中して存在する方が好ましい場合があることに依る。この場合、映像音声処理装置 10 は、このチェーン密度を、そのチェーン品質測定基準として用いればよい。

最後に、チェーン強度であるが、これは、チェーン内の各セグメントが互いにどの程度類似しているかを示す指標であり、当該セグメントが互いに類似しているほど、そのチェーンは高い強度を有しているとみなす。なお、映像音声処理装置 10 において、このチェーン強度を測定する方法については、以下に示すチェーン内類似性測定法や、全ての可能なセグメント対の間の非類似性の平均値をとる方法、或いは、全ての可能なセグメント対の間の非類似性の最大値をとる方法を含め、多数存在する。

一例として、映像音声処理装置 10 が、チェーン内類似性測定法によりチェーン強度を測定する場合を示す。ここで、チェーン内類似性測定法とは、チェーンを構成するセグメントの類似性を、それ

それぞれのセグメントと、そのチェーンが含む最も代表的なセグメントとの非類似性の平均値として表す方法である。典型的なセグメントの例としては、チェーンの重心 (centroid) セグメントが挙げられる。いま、チェーン C における重心セグメントを $S_{centroid}$ とすると、この重心セグメント $S_{centroid}$ は、次式 (8) で定義される。

$$S_{centroid} = \underset{S_A \in C}{\operatorname{argmin}} \frac{1}{|C|} \sum_{S_B \in C} d_F(S_A, S_B) \quad \dots (8)$$

ここで、上式 (8) における argmin は、評価対象の式の値を最小とする入力 $S_A \in C$ を選択することを表す。

これより、チェーン強度を $d_{centroid}$ とすると、このチェーン強度 $d_{centroid}$ は、次式 (9) のように表される。

$$d_{centroid} = \frac{1}{|C|} \sum_{S \in C} d_F(S, S_{centroid}) \quad \dots (9)$$

さて、映像音声処理装置 10 は、上述したチェーン品質測定基準を用いて、具体的に図 12 に示すような一連の処理によりチェーンフィルタリングを行う。

まず、映像音声処理装置 10 は、ステップ S21 において、チェーンリスト C_{list} を候補チェーンで初期化するとともに、フィルタリングチェーンリスト $C_{filtered}$ を空状態にする。

続いて、映像音声処理装置 10 は、ステップ S22 において、チ

チェーンリストC_{list}が空状態であるか否かを判別する。

ここで、チェーンリストC_{list}が空状態であった場合には、映像音声処理装置10は、対象とする候補チェーンが存在しないことから、一連の処理を終了する。

一方、チェーンリストC_{list}が空状態でない場合には、映像音声処理装置10は、ステップS23において、或るチェーンCをチェーンリストC_{list}の最初の要素とし、チェーンCをチェーンリストC_{list}から除去する。

続いて、映像音声処理装置10は、ステップS24において、チェーンCに関するチェーン品質測定基準を計算する。

そして、映像音声処理装置10は、ステップS25において、このチェーン品質測定基準が品質測定基準閾値よりも大きいかな否かを判別する。

ここで、チェーン品質測定基準が品質測定基準閾値よりも小さい場合には、映像音声処理装置10は、ステップS22へと処理を移行し、再び別のチェーンに関する処理を行う。

一方、チェーン品質測定基準が品質測定基準閾値よりも大きい場合には、映像音声処理装置10は、ステップS26において、フィルタリングチェーンリストC_{filtered}にチェーンCを追加する。

そして、映像音声処理装置10は、ステップS27において、チェーンリストC_{list}が空状態であるか否かを判別する。

ここで、チェーンリストC_{list}が空状態であった場合には、映像音声処理装置10は、対象とする候補チェーンが存在しないことから、一連の処理を終了する。

一方、チェーンリストC_{list}が空状態でない場合には、映像音声

処理装置 10 は、ステップ S 23 へと処理を移行する。このようにして、映像音声処理装置 10 は、チェーンリスト C_{list} が空状態となるまで処理を繰り返す。

このような一連の処理によって、映像音声処理装置 10 は、チェーンフィルタリングを行い、どのチェーンが、ビデオ構造の骨格をなす重要なチェーンであるか、或いは、ビデオ構造に関連するチェーンであるかを決定することができる。

以上のように、映像音声処理装置 10 は、このようなバッチクラスタリング技術を用いて、基本類似チェーンを検出することができる。

ところで、映像音声処理装置 10 は、バッチクラスタリング技術とは別の方法として、上述した逐次クラスタリング技術を用いて、基本類似チェーンを検出することもできる。すなわち、映像音声処理装置 10 は、ビデオデータにおけるセグメントを、その入力順にしたがって 1 つずつ処理して、チェーン候補リストを繰り返し更新していく。映像音声処理装置 10 は、この場合にも、バッチクラスタリング技術と同様に、チェーン検出の主たる工程を 2 段階に分けて行う。すなわち、映像音声処理装置 10 は、まず、逐次クラスタリングアルゴリズムを用いて、類似セグメントのクラスタを検出する。次に、映像音声処理装置 10 は、バッチクラスタリング技術と同様のチェーン品質測定基準を用いて、検出されたクラスタをフィルタリングしていく。ここで、映像音声処理装置 10 は、逐次クラスタリング技術を用いた場合のフィルタリング処理として、チェーンのフィルタリングが早い段階で進められる点において、バッチクラスタリング技術の場合と異なる。

さて、逐次クラスタリング技術においては、セグメントのクラスタリングを行う際に、逐次クラスタリングアルゴリズムを用いる。ところで、一般に、ほとんどの逐次クラスタリングは、局所最適に行われる。すなわち、逐次クラスタリングアルゴリズムでは、新たなセグメントが入力される度に、そのセグメントを既存のクラスタに割り当てるか、或いは、そのセグメントのみを含む新たなクラスタを生成するかを局所的に判断している。一方、より精巧な逐次クラスタリングアルゴリズムとしては、セグメントの入力順序にともなうバイアス効果を防ぐため、新たなセグメントが入力される度に、クラスタ分割そのものを更新するものもある。このようなアルゴリズムについては、“J. Roure and L. Talavera, Robust incremental clustering with bad instance orderings: a new strategy, In Proceedings of the Sixth Iberoamerican Conference on Artificial Intelligence, IBERAMIA-98. Pages 136-147. Lisbon, Portugal. Helder Coelho ed., LNAI vol. 1484. Springer Verlag, 1998”の記載を参照することができる。

映像音声処理装置10は、逐次クラスタリングアルゴリズムの一例として、図13に示すような処理を行う。ここでは、セグメントに分割されたビデオデータが、セグメント S_1 、 $\dots$ 、 S_n を有しているものとする。なお、ここでは、チェーン解析の工程も含めた一連の処理について説明する。

まず、映像音声処理装置10は、同図に示すように、ステップS31において、チェーンリスト C_{list} を空状態に初期化し、ステップS32において、セグメント番号 i を1に設定する。

次に、映像音声処理装置10は、ステップS33において、セグ

メント番号 i が総セグメント数 n よりも小さいか否かを判別する。

ここで、セグメント番号 i が総セグメント数 n よりも大きい場合には、映像音声処理装置 10 は、対象とするセグメントが存在しないため、一連の処理を終了する。

一方、セグメント番号 i が総セグメント数 n よりも小さい場合には、映像音声処理装置 10 は、ステップ S 3 4 において、セグメント S_i 、すなわちここではセグメント S_i を取り込み、ステップ S 3 5 において、チェーンリスト C_{list} が空状態であるか否かを判別する。

ここで、チェーンリスト C_{list} が空状態である場合には、映像音声処理装置 10 は、ステップ S 4 2 へと処理を移行する。

一方、チェーンリスト C_{list} が空状態でない場合には、映像音声処理装置 10 は、ステップ S 3 6 において、セグメント S_i に対する非類似性が最小であるチェーン C_{min} を求める。ここで、チェーン C_{min} は、次式 (10) のように定義される。

$$C_{min} = \underset{C \in C_{list}}{\operatorname{argmin}} d_{sc}(C, S_i) \quad \dots (10)$$

上式 (10) において、 $d_{sc}(C, S)$ は、チェーン C とセグメント S との間の非類似性測定基準を表し、次式 (11) で与えられる。

$$d_{sc}(C, S) = \min_{s_i \in C} d_F(S, s_i) \quad \dots (11)$$

これは、バッチクラスタリング技術において定義した類似性測定基準である上式(7)において、その第2引数を、当該セグメントのみを含んだクラスタとしたものと等価である。以下では、チェーン C_{min} とセグメント S_i との間の最小非類似性 $d_{sc}(C_{min}, S_i)$ を、単に d_{min} として表すこととする。

次に、映像音声処理装置10は、ステップS37において、バッチクラスタリング技術の場合において説明したような非類似性閾値 δ_{sim} を用い、最小非類似性 d_{min} が非類似性閾値 δ_{sim} よりも小さいか否かを判別する。

ここで、最小非類似性 d_{min} が非類似性閾値 δ_{sim} よりも大きい場合には、映像音声処理装置10は、ステップS42の処理へと移行し、唯一の要素セグメントとして当該セグメント S_i のみを有する新たなチェーン C_{new} を生成し、ステップS43において、新たなチェーン C_{new} をチェーンリスト C_{list} に追加して、ステップS39の処理へと移行する。

一方、最小非類似性 d_{min} が非類似性閾値 δ_{sim} よりも小さい場合には、映像音声処理装置10は、ステップS38において、チェーン C_{min} に当該セグメント S_i を追加する。すなわち、映像音声処理装置10は、 $C_{min} \leftarrow C_{min} \cup S_i$ とする。

そして、映像音声処理装置10は、ステップS39において、チェーンをフィルタリングする。すなわち、映像音声処理装置10は、上述したように、各要素チェーン $C \in C_{list}$ について、チェーン C の品質を測定して、品質測定基準閾値を上回る品質測定基準を有するチェーンのみを選択し、これをチェーンリスト $C_{filtered}$ に追加する。

さらに、映像音声処理装置 10 は、ステップ S 40 において、逐次的にチェーンを解析する。すなわち、映像音声処理装置 10 は、その時点でのフィルタリングされたチェーンリスト $C_{filtered}$ を解析モジュールに通す。

そして、映像音声処理装置 10 は、ステップ S 41 において、セグメント番号 i に 1 を加算し、ステップ S 33 の処理へと移行する。

このようにして、映像音声処理装置 10 は、セグメント番号 i が総セグメント数 n よりも大きくなるまで、以上の一連の処理を繰り返す。セグメント番号 i が総セグメント数 n よりも大きくなった際のチェーンリスト C_{list} の各要素チェーンを、基本類似チェーンとして検出する。

なお、同図に示す一連の処理は、入力されたビデオデータに含まれる総セグメント数 n が既知であることを前提としている。しかしながら、一般には、総セグメント数 n が前もって与えられていない場合も多い。その場合、逐次クラスタリングアルゴリズムは、同図中ステップ S 33 において、セグメントの入力が引き続きあるか否かによって、処理の続行或いは終了を判別すればよい。

このような一連の処理によって、映像音声処理装置 10 は、逐次クラスタリング技術を用いた基本類似チェーンの検出を行うことができる。

つぎに、上述したリンク類似チェーンを検出する処理について説明する。映像音声処理装置 10 におけるリンク類似チェーンの検出は、基本類似チェーン検出の特殊なケースとして考えることができる。映像音声処理装置 10 は、逐次クラスタリングアルゴリズムを用いたリンク類似チェーン検出方法として、図 14 に示すような処

理を行う。ここでは、セグメントに分割されたビデオデータが、セグメント S_1 、 $\dots$ 、 S_n を有しているものとする。なお、ここでは、チェーン解析の工程も含めた一連の処理を説明する。

映像音声処理装置 10 は、同図に示すように、ステップ S 5 1 において、チェーンリスト C_{list} を空状態に初期化し、ステップ S 5 2 において、セグメント番号 i を 1 に設定する。

次に、映像音声処理装置 10 は、ステップ S 5 3 において、セグメント番号 i が総セグメント数 n よりも小さいか否かを判別する。

ここで、セグメント番号 i が総セグメント数 n よりも大きい場合には、映像音声処理装置 10 は、対象とするセグメントが存在しないため、一連の処理を終了する。

一方、セグメント番号 i が総セグメント数 n よりも小さい場合には、映像音声処理装置 10 は、ステップ S 5 4 において、セグメント S_i 、すなわちここではセグメント S_1 を取り込み、ステップ S 5 5 において、セグメント S_1 に対する非類似性が最小であるチェーン C_{min} を求める。ここで、チェーン C_{min} は、次式 (12) のように定義される。

$$C_{min} = \underset{C \in C_{list}}{argmin} d_{sc}(C, S_i) \quad \dots (12)$$

上式 (12) において、 $d_{sc}(C, S)$ は、やはりチェーン C とセグメント S との間の非類似性測定基準を表すが、リンク類似チェーン検出においては、この非類似性測定基準 $d_{sc}(C, S)$ は、次式 (13) で与えられる。

$$d_{sc} = (C, S) d_F(S_{|C|}, S_i) \quad \dots (13)$$

すなわち、非類似性測定基準 $d_{sc}(C, S)$ は、基本類似チェーンの検出の際に用いた非類似性測定基準である上式 (11) とは異なり、当該セグメントと、チェーン C における最後の要素セグメントとの間の非類似性として与えられる。

次に、映像音声処理装置 10 は、ステップ S 56 において、上述したような非類似性閾値 δ_{sim} を用い、最小非類似性 d_{min} が非類似性閾値 δ_{sim} よりも小さいか否かを判別する。

ここで、最小非類似性 d_{min} が非類似性閾値 δ_{sim} よりも大きい場合には、映像音声処理装置 10 は、ステップ S 61 の処理へと移行し、唯一の要素セグメントとして当該セグメント S_i のみを有する新たなチェーン C_{new} を生成し、ステップ S 62 において、新たなチェーン C_{new} をチェーンリスト C_{list} に追加して、ステップ S 58 の処理へと移行する。

一方、最小非類似性 d_{min} が非類似性閾値 δ_{sim} よりも小さい場合には、映像音声処理装置 10 は、ステップ S 57 において、チェーン C_{min} の末端に当該セグメント S_i を追加する。すなわち、映像音声処理装置 10 は、 $C_{min} \leftarrow C_{min}, S_i$ とする。

そして、映像音声処理装置 10 は、ステップ S 58 において、チェーンをフィルタリングする。すなわち、映像音声処理装置 10 は、上述したように、各要素チェーン $C \in C_{list}$ について、チェーン C の品質を測定して、品質測定基準閾値を上回る品質測定基準を有す

るチェーンのみを選択し、これをチェーンリスト *C_{filtered}* に追加する。なお、映像音声処理装置 10 は、この工程を省略することもできる。

さらに、映像音声処理装置 10 は、ステップ S 59 において、逐次的にチェーンを解析する。すなわち、映像音声処理装置 10 は、その時点でのフィルタリングされたチェーンリスト *C_{filtered}* を解析モジュールに通す。

そして、映像音声処理装置 10 は、ステップ S 60 において、セグメント番号 *i* に 1 を加算し、ステップ S 53 の処理へと移行する。

このようにして、映像音声処理装置 10 は、セグメント番号 *i* が総セグメント数 *n* よりも大きくなるまで、以上の一連の処理を繰り返す。セグメント番号 *i* が総セグメント数 *n* よりも大きくなった際のチェーンリスト *C_{list}* の各要素チェーンを、リンク類似チェーンとして検出する。

このような一連の処理によって、映像音声処理装置 10 は、このような逐次クラスタリング技術を用いて、リンク類似チェーンを検出することができる。

なお、同図に示す一連の処理は、入力されたビデオデータに含まれる総セグメント数 *n* が既知であることを前提としている。しかしながら、一般には、総セグメント数 *n* が前もって与えられていない場合も多い。その場合、逐次クラスタリングアルゴリズムは、同図中ステップ S 53 において、セグメントの入力が引き続きあるか否かによって、処理の続行或いは終了を判別すればよい。

つぎに、上述した周期的チェーンを検出する処理について説明する。周期的チェーン *C_{cyclic}* は、*k* 個の異なる基本類似チェーン又

はリンク類似チェーンがまとまったもの $\{C_1, \dots, C_k\}$ とみなすことができる。以下、周期的チェーン C_{cyclic} 内のセグメントを、 $S_1, \dots, S_n$ と記述し、また $C(S_i)$ は、セグメント S_i の出現元のチェーン番号 $1, \dots, k$ を示すこととする。これより、 C_{cyclic} が周期的チェーンであるならば、 $C(S_1), C(S_2), \dots, C(S_n)$ なる一連のチェーン番号の並びは、 $i_1, \dots, i_k, i_1, \dots, i_k, \dots, i_1, \dots, i_k$ という形式で記述されることとなる。ここで、その1周期分 $i_1, \dots, i_k$ は、チェーン番号 $1, \dots, k$ の順列、換言すれば、重複しない任意の並びである。なお、以下では、1周期内に含まれるセグメントの数が1つである周期的チェーン $i_1, i_1, \dots, i_1$ を基本周期チェーンと称することとする。

ところで、通常、ビデオデータにおける周期的構造は、各周期が完全に一致したものではなく近似的なものであるため、映像音声処理装置10は、図15に示すような一連の処理によって、ビデオデータ内の近似的な周期的チェーンを探す。ここで、映像音声処理装置10は、必要に応じて、その元となる基本周期チェーンが均一でなければならないという制約条件を追加することができる。ここでは、この制約条件のもとに行われる処理について説明する。

まず、映像音声処理装置10は、同図に示すように、ステップS71及びステップS72において、ビデオデータに含まれる基本周期チェーンを検出し、それに基づいて初期チェーンリストを生成し、さらに初期チェーンリストに含まれる基本周期チェーンの全てが均一チェーンの制約条件を満たすように、初期チェーンリストを更新する。

すなわち、映像音声処理装置 10 は、ステップ S 7 1 において、上述した基本類似チェーン又はリンク類似チェーンを検出するアルゴリズムを用いて、初期チェーンリスト C_{list} を求める。

そして、映像音声処理装置 10 は、ステップ S 7 2 において、初期チェーンリストに含まれる各チェーン C について、その均一性を確認し、チェーン C が均一でない場合には、このチェーン C を、その時間的間隔が最大となるような複数の均一サブチェーンに分割する。続いて、映像音声処理装置 10 は、得られた均一サブチェーンを、上述した基本類似チェーン又はリンク類似チェーンを検出するアルゴリズムにおいて説明したようなチェーン品質測定基準を用いてフィルタリングし、選択された均一サブチェーンを初期チェーンリスト C_{list} に追加する。

次に、映像音声処理装置 10 は、ステップ S 7 3 において、チェーンリスト C_{list} の中から、時間的に重複して交差する 1 対のチェーン、すなわち、 $\exists C_1, C_2 \mid [C_1^{start}, C_1^{end}] \cap [C_2^{start}, C_2^{end}]$ なるチェーン C_1, C_2 を求める。

そして、映像音声処理装置 10 は、ステップ S 7 4 において、このような重複しているチェーン C_1, C_2 が存在するか否かを判別する。

ここで、重複しているチェーン C_1, C_2 が存在しない場合には、映像音声処理装置 10 は、チェーンリスト C_{list} が既に複数の周期的チェーンを含んでいるものとして、一連の処理を終了する。

一方、重複しているチェーン C_1, C_2 が存在する場合には、映像音声処理装置 10 は、ステップ S 7 5 乃至ステップ S 7 8 において、2 つのチェーン C_1, C_2 がまとまった 1 つの周期的チェーンを構成

するか否かを決定するため、その２つの周期的チェーンを合わせた周期的チェーンにおいて、各周期の間の整合性を評価する。

すなわち、映像音声処理装置１０は、ステップＳ７５において、２つのチェーン C_1 、 C_2 を合わせて、新たな周期的チェーン C_M を形成する。ここで、チェーン C_M におけるセグメントを S_1 、 S_2 、 $\dots$ 、 $S_{|C_M|}$ と表すこととする。

続いて、映像音声処理装置１０は、ステップＳ７６において、セグメント S_1 の出現元のチェーン番号 $C(S_1)$ を C とし、チェーン番号の並び $C(S_1)$ 、 $C(S_2)$ 、 $\dots$ 、 $C(S_{|C_M|})$ において C の発生毎に、すなわち、セグメント S_1 と同じチェーンに属するセグメントが出現する直前を境に、チェーン C_M をサブチェーン C_M^1 、 C_M^2 、 $\dots$ 、 C_M^k に分解する。この結果、映像音声処理装置１０は、次式（１４）に示すようなサブチェーンのリストを得る。

$$\begin{aligned} C_M^1 &= S_1, \dots, S_{i_1}, \\ C_M^2 &= S_{i_1+1}, \dots, S_{i_2}, \\ &\vdots \\ C_M^k &= S_{i_{k-1}+1}, \dots, S_{i_k}, \end{aligned} \quad \dots \quad (14)$$

この操作から明らかなように、上式（１４）では、全ての C_M^j について、 $C(S_{i_{j-1}+1}) = C(S_1)$ が成り立つ。

続いて、映像音声処理装置１０は、ステップＳ７７において、最も出現頻度の高いサブチェーン C_M^{cycle} を見つける。すなわち、映像音声処理装置１０は、次式（１５）に示すような処理を行う。

$$C_M^{cycle} = \underset{C_M^k}{argmax} \left| \left\{ C_M^i \mid C_M^i = C_M^k, i \in \{1, \dots, k\} \right\} \right| \quad \dots (15)$$

そして、映像音声処理装置 10 は、ステップ S 7 8 において、最も出現頻度の高いサブチェーン C_M^{cycle} が、元のチェーン C_M の 1 周期となりうるか否かを評価する。すなわち、映像音声処理装置 10 は、整合係数 $mesh$ を、次式 (16) で示すように、ステップ S 7 6 にて求めた C_M^{cycle} の出現頻度のサブチェーン総数に対する比で定義し、続くステップ S 7 9 において、この整合係数が所定の閾値を越えるか否かを判別する。

$$mesh = \frac{\left| \left\{ C_M^i \mid C_M^i = C_M^{cycle}, i \in \{1, \dots, k\} \right\} \right|}{k} \quad \dots (16)$$

ここで、整合係数が閾値を越えていない場合には、映像音声処理装置 10 は、ステップ S 7 3 の処理へと移行し、他の重複しているチェーンを求めて同様の処理を繰り返す。

一方、整合係数が閾値を越えている場合には、映像音声処理装置 10 は、ステップ S 8 0 において、チェーン C_1 , C_2 をチェーンリスト C_{list} から除去して、ステップ S 8 1 において、チェーン C_M を

チェーンリスト C_{list} に追加し、ステップ S 7 3 の処理へと移行する。

映像音声処理装置 1 0 は、チェーンリスト C_{list} に含まれる全ての周期的チェーンについて重複しているチェーンが存在しなくなるまでこのような一連の処理を繰り返すことによって、最終的な周期的チェーンを含むチェーンリスト C_{list} を得ることができる。

以上のように、映像音声処理装置 1 0 は、非類似性測定基準と抽出した特徴量とを用いて、類似したセグメントの各種チェーンを検出することができる。

つぎに、図 5 中ステップ S 5 におけるチェーン解析について説明する。映像音声処理装置 1 0 は、検出したチェーンを用いて、ビデオデータの局所的ビデオ構造及び／又は大局的ビデオ構造を決定して出力する。ここでは、ビデオデータに発生する基本的な構造パターンを検出するのに、チェーン解析の結果をどのように用いるのかについて具体的な例を挙げて説明する。

まず、ビデオデータに発生する局所的な構造パターンであるシーンについて説明する。

シーンは、上述したように、セグメントのレベルより上位に位置づけられた最も基本的な局所的ビデオ構造の単位であり、意味的に関連する一連のセグメントから構成される。映像音声処理装置 1 0 は、チェーンを用いて、これらのシーンを検出することができる。映像音声処理装置 1 0 におけるシーン検出において、チェーンが満たすべき条件とは、そのチェーンが含む全てのセグメントに関して、互いに連続したセグメント間の時間間隔が、時間閾値と称される或る定められた値を超えないことである。ここでは、この条件を満た

すチェーンを局所チェーンと称する。

映像音声処理装置 10 は、チェーンを用いてシーンを検出するために、図 16 に示すような一連の処理を行う。

まず、映像音声処理装置 10 は、同図に示すように、ステップ S 9 1 乃至ステップ S 9 4 において、局所チェーンリストを求める。

すなわち、映像音声処理装置 10 は、ステップ S 9 1 において、上述した基本類似チェーン検出アルゴリズムを用いて、1 組の初期チェーンリストを求める。

次に、映像音声処理装置 10 は、ステップ S 9 2 において、求めた初期チェーンリストにおける各チェーン C について、チェーン C が局所チェーンでない場合には、チェーン C を、局所チェーンの条件範囲において最長であるところの局所サブチェーン $C = C_1, \dots, C_n$ の並びに分解する。

その後、映像音声処理装置 10 は、ステップ S 9 3 において、チェーンリストからチェーン C を除去する。

さらに、映像音声処理装置 10 は、ステップ S 9 4 において、各サブチェーン C_i をチェーンリストに追加する。この工程が終了すると、全てのチェーンが局所的となる。

次に、映像音声処理装置 10 は、ステップ S 9 5 において、チェーンリストの中から、時間的に交差する 1 対の重複しているチェーン C_1, C_2 、すなわち、 $\exists C_1, C_2 \mid [C_1^{\text{start}}, C_1^{\text{end}}] \cap [C_2^{\text{start}}, C_2^{\text{end}}]$ であるところのチェーン C_1, C_2 を求める。

続いて、映像音声処理装置 10 は、ステップ S 9 6 において、このような重複しているチェーン C_1, C_2 が存在するか否かを判別する。

ここで、重複しているチェーン C_1 、 C_2 が存在しない場合には、映像音声処理装置10は、チェーンリストに含まれた各チェーン毎に1つのシーンが存在するものとして、一連の処理を終了する。

一方、重複しているチェーン C_1 、 C_2 が存在する場合には、映像音声処理装置10は、ステップS97において、重複しているチェーン C_1 、 C_2 を合わせて、新たなチェーン C_M を形成する。

さらに、映像音声処理装置10は、ステップS98において、チェーンリストから重複しているチェーン C_1 、 C_2 を除去して、チェーン C_M を追加し、その後再びステップS95の処理へと移行して、同様の処理を繰り返す。

このようにした結果、重複しているチェーンがチェーンリスト内に存在しなくなったとき、最終的に得られたチェーンリストに含まれた各チェーン毎に、1シーンが存在することになる。なお、チェーン C_i に対応するシーン S_i の境界は、 C^{start} 及び C^{end} で与えられる。

ところで、セグメントの中には、いかなるチェーンにも割り当てられずに残るものがあるが、映像音声処理装置10は、既定値としては、2つの検出されたシーン間に残ったこのようなセグメントをまとめて1つのシーンとする。

このような一連の処理によって、映像音声処理装置10は、チェーンを用いることによって、ビデオデータにおける局所的な構造パターンであるシーンを検出することができる。

このような処理を先に図2に示した会話場面に適用する場合を考える。この場合、映像音声処理装置10は、ステップS91乃至ステップS94において、話し手のセグメントのそれぞれについて、

局所チェーンを求める。そして、映像音声処理装置 10 は、ステップ S 97 において、これらのチェーンをまとめ、シーン全体を表す単一の大きいチェーンを形成することになる。

このように、映像音声処理装置 10 は、会話場面におけるシーンを検出することができる。

なお、映像音声処理装置 10 においては、シーンを検出した際に、シーン内の全てのセグメントがチェーンに含まれる訳ではないことには注意を要する。

また、映像音声処理装置 10 は、上述したアルゴリズムを逐次的に行うことによって、シーンを逐次的に検出することもできる。

つぎに、大局的な構造パターンとして、ニュース項目を検出する場合について説明する。

上述したように、ニュース番組は、そのニュース項目が、例えば、まずアンカーによる導入文で始まり、現場からの 1 以上のリポートが続くといった周期的構造を有している。すなわち、このようなビデオ構造は、アンカーショットから次のアンカーショットの直前までを 1 周期とした単純な周期的構造であるとみなすことができる。

映像音声処理装置 10 は、チェーンを用いてニュース項目を自動的に検出するために、図 17 に概略を示すような一連の処理を行う。

まず、映像音声処理装置 10 は、同図に示すように、ステップ S 101 において、上述した周期的チェーン検出アルゴリズムを用いて、周期的チェーンの検出を行う。この工程を行うことによって、映像音声処理装置 10 は、周期的チェーンのリストを得ることができる。ここで、各周期は、ニュース項目を表してもよく、表さなくてもよい。

次に、映像音声処理装置 10 は、ステップ S 102 において、その周期が、ビデオデータの全長の所定割合よりも短いところの周期的チェーンを全て除去する。すなわち、映像音声処理装置 10 は、この工程を行うことによって、ニュース項目を表す見込みのない短い周期の周期的チェーンを排除することができる。このような周期は、例えば司会者がゲストにインタビューをする場合或いは他の短時間周期がニュース放送において現れる場合に発生しうるものである。

そして、映像音声処理装置 10 は、ステップ S 103 において、ステップ S 102 において残った全ての周期的チェーンについて、時間的に最も短い周期的チェーンを求め、この周期的チェーンが他の周期的チェーンに重なる場合には、その周期的チェーンを周期的チェーンのリストから除去する。映像音声処理装置 10 は、いかなる周期的チェーンも他の周期的チェーンと重なることがなくなるまで、この処理を繰り返す。このステップ S 103 が終了した後に残った周期的チェーンのリストは、検出したニュース項目リストを含むこととなる。すなわち、ステップ 103 にて得られた周期的チェーンのリストの各周期は、それぞれ、1つのニュース項目を表す。

このようにして、映像音声処理装置 10 は、チェーンを用いてニュース項目を自動的に検出することができる。

なお、特筆すべきは、映像音声処理装置 10 は、例えば、ニュース放送のメイン、スポーツ、ビジネスの各セグメントの間といったニュース放送の途中にニュースキャスターが変わった場合にも、問題なく作用することができることである。

つぎに、スポーツ放送におけるプレイを検出する場合について説

明する。

多くのスポーツは、同じ一連の工程が何度も繰り返されることによりプレイが構成されるといった固定パターンを有するという特徴がある。例えば、野球の場合には、ピッチャーがボールを投げ、バッターがボールを打とうとすることによりプレイが構成される。ビデオデータにおいて、このようなプレイ構造を有する他のチームスポーツとしては、例えばフットボールやラグビーが挙げられる。

このプレイ構造が放送されると、ビデオデータは、プレイの各部分についてのセグメント群の繰り返しを表すこととなる。すなわち、ビデオデータは、ピッチャーを表すセグメントの後に、バッターを表すセグメントが続き、ボールが打たれた場合には、外野選手等を表すセグメントが入ることになる。そのため、野球放送に対して映像音声処理装置 10 によるチェーン検出を適用した場合には、ビデオデータにおいて、ピッチャーを表すセグメントが 1 チェーンとして検出され、バッターを表すセグメントが別の 1 チェーンを占め、その他のチェーンが外野や種々の光景にあたることになる。

すなわち、これらのスポーツ放送においては、プレイ構造が、上述した周期的チェーン検出方法を用いて検出することができる周期的映像となる。このような他の例として、テニスが挙げられる。テニスにおいて、ビデオデータは、サーブ、ボレー、サーブ、ボレーといったような周期を構成する。この場合、各サーブを表すセグメントは、映像的に互いに類似しているため、映像音声処理装置 10 は、プレイを検出するために、このようなセグメントを用いることができる。その結果、映像音声処理装置 10 による構造解析においては、近似的にゲームのプレイ構造を検出することができる。

さらに、他のスポーツ、特に個人競技においては、プレイ構造としては、1人の競技者が或る活動を完結するまで行うことになるが、各競技者は、全て近似的に同じ活動を行っているともみなすことができる。例えば、スキージャンプ競技では、各競技者が1回ジャンプを行い、次の競技者が続いて同様のジャンプを行う。すなわち、ジャンプ競技の放送におけるビデオデータは、競技者がジャンプの準備に入り、助走路を滑り降りて、着地するというセグメントの並びからなるのが一般的である。これより、ビデオデータは、このような一連のセグメントを、各競技者毎に繰り返すことで構成される。このような放送におけるビデオデータに対してチェーン検出を適用した場合には、ジャンプの各段階毎に類似した一連のチェーンを検出することになる。したがって、各競技者毎の周期は、周期的チェーン検出方法を用いて抽出することができる。

映像音声処理装置10において、チェーン解析によりスポーツ放送におけるプレイを自動的に検出する際には、適当でないチェーンを排除するために、さらなる制約を設ける必要がある場合がある。どのような制約が適切であるかは、スポーツの種類によって異なるが、例えば、映像音声処理装置10は、検出された周期的チェーンのうち、その周期が十分長いものだけをプレイとして検出するという経験的なルールを用いることができる。

すなわち、映像音声処理装置10は、チェーンを用いてスポーツ放送におけるプレイを自動的に検出するために、図18に概略を示すような一連の処理を行う。

まず、映像音声処理装置10は、同図に示すように、ステップS111において、上述した周期的チェーン検出アルゴリズムを用い

て、周期的チェーンを検出する。

そして、映像音声処理装置 10 は、ステップ S 1 1 2 において、得られたチェーンのリストに対して品質条件を適用し、そのチェーンリストをフィルタリングして、本質的でないチェーンを除去する。品質条件としては、例えば、プログラムの大部分にわたるような周期的チェーンのみを残すといったことが挙げられる。勿論、映像音声処理装置 10 は、対象とするスポーツに特有の制約条件を追加してもよい。

このようにして、映像音声処理装置 10 は、チェーン解析によりスポーツ放送におけるプレイを自動的に検出することができる。

つぎに、周期検出とシーン検出とを組み合わせるとピックを検出する場合について説明する。

例えば、ドラマ、コメディ、バラエティといった多くのテレビ番組におけるビデオデータは、上述したシーンにより構成されている。しかし、ビデオデータは、その上位の構造として、幾つかの関連シーンの並びから構成されるピックなる構造を有する場合がある。このピックは、必ずしも、常にスタジオ司会者による紹介セグメントに始まるようなニュース放送におけるピックと類似したものであるとは限らない。例えば、視覚的な例として、紹介セグメントの代わりにロゴイメージのセグメント或いは総合司会者のセグメントが用いられたり、或いは聴覚的な例として、新たなピックが始まる度に、常に同じテーマ音楽が流されたりする場合がある。

或る番組におけるビデオデータが、このようなピック構造を有しているか否かは、周期検出とシーン検出とを組み合わせることによって、判断することができる。

そのため、映像音声処理装置 10 は、チェーンを用いた周期検出とシーン検出とを組み合わせたトピック検出を行うために、図 19 に概略を示すような一連の処理を行う。

まず、映像音声処理装置 10 は、同図に示すように、ステップ S 121 において、基本類似チェーン検出を行い、1 組の基本類似チェーンリストを識別する。

次に、映像音声処理装置 10 は、ステップ S 122 において、周期的チェーン検出を行い、1 組の周期的チェーンのリストを識別する。

続いて、映像音声処理装置 10 は、ステップ S 123 において、ステップ S 121 において求めた基本類似チェーンリストを用い、先に図 16 に示したアルゴリズムを適用して、シーン構造を抽出する。映像音声処理装置 10 は、この結果、シーンのリストを得ることができる。

そして、映像音声処理装置 10 は、ステップ S 124 において、ステップ S 122 において求めた周期的チェーンのリストを、ステップ S 123 において検出した各シーン要素と比較する。ここで、映像音声処理装置 10 は、検出したシーンのリストに含まれるシーンよりも短い周期の周期的チェーンを全て除去する。この結果得られた残りの周期的チェーンは、各周期が幾つかのシーンを有しているが、この各周期はそれぞれ、候補トピックとして識別されることとなる。

このようにして、映像音声処理装置 10 は、チェーンを用いた周期検出とシーン検出とを組み合わせることによって、トピック検出を行うことができる。

なお、映像音声処理装置 10 は、ステップ S 1 2 4 において、その他の制約や品質条件を設けることによって、トピック検出の精度を高めることもできる。

以上のように、映像音声処理装置 10 は、検出した各種チェーンを用いて、ビデオデータの各種局所的ビデオ構造及び／又は各種大局的ビデオ構造を決定して出力することができる。

以上説明してきたように、本発明の実施の形態として示す映像音声処理装置 10 は、互いに類似する複数の映像セグメント又は音声セグメントから構成される類似チェーンを検出することが可能である。そして、映像音声処理装置 10 は、これらの類似チェーンを解析することによって、高レベルのビデオ構造を抽出することができる。特に、映像音声処理装置 10 は、局所的ビデオ構造及び大局的ビデオ構造の解析を共通の枠組みで行うことができる。

この映像音声処理装置 10 は、完全に自動的に処理を行うことができ、ユーザが事前にビデオデータの内容の構造を知る必要はない。

また、映像音声処理装置 10 は、逐次的なチェーン検出を用いることにより、逐次的にビデオ構造を解析することも可能であり、さらに、プラットフォームの計算能力が十分強力であるならば、ビデオ構造解析を実時間で行うことが可能である。これにより、映像音声処理装置 10 は、事前に記録されたビデオデータその他、ライブのビデオ放送にも用いることができる。例えば、映像音声処理装置 10 は、スポーツ放送におけるプレイ検出において、ライブのスポーツ放送に適用可能である。

さらに、映像音声処理装置 10 は、ビデオ構造を検出した結果、ビデオブラウジングのための新たな高レベルアクセスの基礎を与え

ることができる。すなわち、映像音声処理装置 10 は、セグメントではなくトピックといった高レベルのビデオ構造を用いてビデオデータの内容を映像化することにより、内容に基づいたビデオデータへのアクセスを可能とする。例えば、映像音声処理装置 10 は、シーンを表示することにより、ユーザは、番組の要旨をすばやく知ることができる、興味のある部分を迅速に見つけることができる。

さらにまた、映像音声処理装置 10 は、ニュース放送におけるトピック検出の結果を用いることにより、ユーザに、ニュース項目単位での選択や視聴を可能とする等、ニュース放送に対して、強力で新しい方法のアクセスを可能とする。

また、映像音声処理装置 10 は、ビデオ構造検出の結果、ビデオデータの要約を自動的に作成するための基礎を与えることができる。一般に、筋の通った要約を作成するためには、ビデオデータに含まれる任意のセグメントを組み合わせるのではなく、ビデオデータを再構成可能な意味を持つ成分に分解し、それを元に適切なセグメントを組み合わせることが必要である。映像音声処理装置 10 により検出されたビデオ構造は、そのような要約を作成するための基礎的な情報を提供するものである。

さらに、映像音声処理装置 10 では、ビデオデータを、そのジャンル別に解析することが可能である。例えば、映像音声処理装置 10 は、テニスの試合のみを検出することを可能とする。

これより、映像音声処理装置 10 は、放送局におけるビデオ編集システムに組み込まれることにより、ビデオデータを、その内容に基づいて編集することを可能とする。

さらにまた、映像音声処理装置 10 は、一般家庭において、ホー

ムビデオを解析したり、ホームビデオからビデオ構造を自動的に抽出するのに用いることができる。さらに、映像音声処理装置 10 は、ビデオデータの内容の要約や、その内容に基づいた編集を行うのに用いることができる。

一方、映像音声処理装置 10 は、ビデオチェーンを、人手によるビデオデータの内容の解析を補足するツールとして使用することが可能である。特に、映像音声処理装置 10 は、チェーン検出の結果を映像化することにより、ビデオデータの内容のナビゲーションやビデオ構造解析が容易にすることができる。

また、映像音声処理装置 10 は、そのアルゴリズムが非常に単純であり計算上の効率もよいため、セットトップボックスやデジタルビデオレコーダ、ホームサーバ等の家庭用電子機器にも適用することができる。

なお、本発明は、上述した実施の形態に限定されるものではなく、例えば、セグメント間の類似性測定のために用いる特徴量や、適用可能なビデオデータの内容等は、上述したもの以外でもよいことは勿論であり、その他、本発明の趣旨を逸脱しない範囲で適宜変更が可能であることはいうまでもない。

産業上の利用可能性

以上詳細に説明したように、本発明にかかる信号処理方法は、供給された信号の内容の意味構造を反映するパターンを検出して解析する信号処理方法であって、信号を構成する連続したフレームのひと続きから形成されるセグメントから、その特徴を表す少なくとも

1 つ以上の特徴量を抽出する特徴量抽出工程と、特徴量を用いて、特徴量のそれぞれ毎に、セグメントの対の間の類似性を測定する測定基準を算出して、この測定基準によりセグメントの対の間の類似性を測定する類似性測定工程と、特徴量と測定基準とを用いて、セグメントのうち、互いに類似する複数のセグメントから構成される類似チェーンを検出する検出工程とを備える。

したがって、本発明にかかる信号処理方法は、信号において類似したセグメントが構成する基本的な構造パターンを検出することができ、これらの構造パターンがどのように組み合わせられているかを解析することによって、高レベルの構造を抽出することができる。

また、本発明にかかる映像音声処理装置は、供給されたビデオ信号の内容の意味構造を反映する映像及び／又は音声のパターンを検出して解析する映像音声処理装置であって、ビデオ信号を構成する連続した映像及び／又は音声フレームのひと続きから形成される映像及び／又は音声セグメントから、その特徴を表す少なくとも 1 つ以上の特徴量を抽出する特徴量抽出手段と、特徴量を用いて、特徴量のそれぞれ毎に、映像及び／又は音声セグメントの対の間の類似性を測定する測定基準を算出して、この測定基準により映像及び／又は音声セグメントの対の間の類似性を測定する類似性測定手段と、特徴量と測定基準とを用いて、映像及び／又は音声セグメントのうち、互いに類似する複数の映像及び／又は音声セグメントから構成される類似チェーンを検出する検出手段とを備える。

したがって、本発明にかかる映像音声処理装置は、ビデオ信号において類似した映像及び／又は音声セグメントの基本的な構造パターンを決定して出力することが可能であり、これらの構造パターン

がどのように組み合わせられているかを解析することによって、高レベルのビデオ構造を抽出することが可能となる。

請求の範囲

1. 供給された信号の内容の意味構造を反映するパターンを検出して解析する信号処理方法であって、

上記信号を構成する連続したフレームのひと続きから形成されるセグメントから、その特徴を表す少なくとも1つ以上の特徴量を抽出する特徴量抽出工程と、

上記特徴量を用いて、上記特徴量のそれぞれ毎に、上記セグメントの対の間の類似性を測定する測定基準を算出して、この測定基準により上記セグメントの対の間の類似性を測定する類似性測定工程と、

上記特徴量と上記測定基準とを用いて、上記セグメントのうち、互いに類似する複数のセグメントから構成される類似チェーンを検出する検出工程とを備えること

を特徴とする信号処理方法。

2. 上記類似チェーンを用いて解析し、上記信号の局所的構造及び／又は大局的構造を決定して出力する解析工程を備えること

を特徴とする請求の範囲第1項記載の信号処理方法。

3. 上記信号とは、ビデオデータにおける映像信号と音声信号との少なくとも1つであること

を特徴とする請求の範囲第1項記載の信号処理方法。

4. 上記類似チェーンは、当該類似チェーンが含む類似セグメントの間の関係に制約を有すること

を特徴とする請求の範囲第1項記載の信号処理方法。

5. 上記類似チェーンは、当該類似チェーンの構造に制約を有する

こと

を特徴とする請求の範囲第1項記載の信号処理方法。

6. 上記類似チェーンは、当該類似チェーンが含む全てのセグメントが互いに類似した関係にある基本類似チェーンであること

を特徴とする請求の範囲第4項記載の信号処理方法。

7. 上記類似チェーンは、当該類似チェーンが含む全てのセグメントにおいて、隣接するセグメントが互いに類似した関係にあるリンク類似チェーンであること

を特徴とする請求の範囲第4項記載の信号処理方法。

8. 上記類似チェーンは、当該類似チェーンが含む全てのセグメントにおいて、セグメントのそれぞれが、当該セグメントから所定の数だけ後方に配置されたセグメントと互いに類似した関係にある周期的チェーンであること

を特徴とする請求の範囲第4項記載の信号処理方法。

9. 上記類似チェーンは、当該類似チェーンが含む全てのセグメントにおいて、隣接するセグメントの各対における時間間隔が、所定の時間よりも短い局所チェーンであること

を特徴とする請求の範囲第5項記載の信号処理方法。

10. 上記類似チェーンは、当該類似チェーンが含む全てのセグメントにおいて、セグメントが近似的に等時間間隔で出現する均一チェーンであること

を特徴とする請求の範囲第5項記載の信号処理方法。

11. 上記検出工程は、上記特徴量と上記測定基準とを用いて、互いに類似しているセグメントを検出してまとめて候補チェーンを形成する候補チェーン検出工程と、

上記候補チェーンのそれぞれ毎に数的基準に対応する品質測定基準を算出して、上記信号の構造パターン解析における上記候補チェーンの重要性及び関連性を測定し、上記品質測定基準が所定の品質測定基準閾値を上回る候補チェーンのみを出力するフィルタリング工程とを有すること

を特徴とする請求の範囲第6項記載の信号処理方法。

12. 上記信号におけるセグメントのうち、セグメントが供給された時間順にしたがって当該セグメントを1つずつ逐次処理すること

を特徴とする請求の範囲第2項記載の信号処理方法。

13. 上記検出工程は、対象とするセグメントに関する上記特徴量と上記測定基準とを用いて、当該セグメントを含む候補チェーンを随時更新して求める候補チェーン検出工程と、

上記候補チェーンのそれぞれ毎に数的基準に対応する品質測定基準を算出して、上記信号の構造パターン解析における上記候補チェーンの重要性及び関連性を測定し、上記品質測定基準が所定の品質測定基準閾値を上回る候補チェーンのみを出力するフィルタリング工程とを有すること

を特徴とする請求の範囲第12項記載の信号処理方法。

14. 上記検出工程は、周期的チェーンの初期候補を求める初期周期的チェーン検出工程と、

上記周期的チェーンの初期候補の中から、時間的に交差する重複チェーンを求める重複チェーン検出工程と、

上記重複チェーンの整合を求める整合工程とを有すること

を特徴とする請求の範囲第8項記載の信号処理方法。

15. 上記解析工程により、上記類似チェーンを用いて、上記信号

の局所的構造として、セグメントの意味に基づく部分集合であるシーンを検出して出力すること

を特徴とする請求の範囲第2項記載の信号処理方法。

16. 上記解析工程により、上記類似チェーンを用いて、上記信号の大局的構造として、互いに類似するセグメントが反復的に発生する構造パターンを検出して出力すること

を特徴とする請求の範囲第2項記載の信号処理方法。

17. 上記構造パターンとして、ニュース放送におけるニュース項目を検出して出力すること

を特徴とする請求の範囲第16項記載の信号処理方法。

18. 上記構造パターンとして、プレイが反復的に発生するスポーツ放送におけるビデオ構造を検出して出力すること

を特徴とする請求の範囲第16項記載の信号処理方法。

19. 上記解析工程により、上記類似チェーンを用いて、セグメントの意味に基づく部分集合であるシーンのうち、関連するシーンをまとめたトピック構造を検出して出力すること

を特徴とする請求の範囲第2項記載の信号処理方法。

20. 供給されたビデオ信号の内容の意味構造を反映する映像及び／又は音声のパターンを検出して解析する映像音声処理装置であって、

上記ビデオ信号を構成する連続した映像及び／又は音声フレームのひと続きから形成される映像及び／又は音声セグメントから、その特徴を表す少なくとも1つ以上の特徴量を抽出する特徴量抽出手段と、

上記特徴量を用いて、上記特徴量のそれぞれ毎に、上記映像及び

／又は音声セグメントの対の間の類似性を測定する測定基準を算出して、この測定基準により上記映像及び／又は音声セグメントの対の間の類似性を測定する類似性測定手段と、

上記特徴量と上記測定基準とを用いて、上記映像及び／又は音声セグメントのうち、互いに類似する複数の映像及び／又は音声セグメントから構成される類似チェーンを検出する検出手段とを備えること

を特徴とする映像音声処理装置。

21. 上記類似チェーンを用いて解析し、上記ビデオ信号の局所的ビデオ構造及び／又は大局的ビデオ構造を決定して出力する解析手段を備えること

を特徴とする請求の範囲第20項記載の映像音声処理装置。

22. 上記類似チェーンは、当該類似チェーンが含む類似の映像及び／又は音声セグメントの間の関係に制約を有すること

を特徴とする請求の範囲第20項記載の映像音声処理装置。

23. 上記類似チェーンは、当該類似チェーンの構造に制約を有すること

を特徴とする請求の範囲第20項記載の映像音声処理装置。

24. 上記類似チェーンは、当該類似チェーンが含む全ての映像及び／又は音声セグメントが互いに類似した関係にある基本類似チェーンであること

を特徴とする請求の範囲第22項記載の映像音声処理装置。

25. 上記類似チェーンは、当該類似チェーンが含む全ての映像及び／又は音声セグメントにおいて、隣接する映像及び／又は音声セグメントが互いに類似した関係にあるリンク類似チェーンであるこ

と

を特徴とする請求の範囲第 2 2 項記載の映像音声処理装置。

26. 上記類似チェーンは、当該類似チェーンが含む全ての映像及び／又は音声セグメントにおいて、映像及び／又は音声セグメントのそれぞれが、当該セグメントから所定の数だけ後方に配置された映像及び／又は音声セグメントと互いに類似した関係にある周期的チェーンであること

を特徴とする請求の範囲第 2 2 項記載の映像音声処理装置。

27. 上記類似チェーンは、当該類似チェーンが含む全ての映像及び／又は音声セグメントにおいて、隣接する映像及び／又は音声セグメントの各対における時間間隔が、所定の時間よりも短い局所チェーンであること

を特徴とする請求の範囲第 2 3 項記載の映像音声処理装置。

28. 上記類似チェーンは、当該類似チェーンが含む全ての映像及び／又は音声セグメントにおいて、映像及び／又は音声セグメントが近似的に等時間間隔で出現する均一チェーンであること

を特徴とする請求の範囲第 2 3 項記載の映像音声処理装置。

29. 上記検出手段は、上記特徴量と上記測定基準とを用いて、互いに類似している映像及び／又は音声セグメントを検出してまとめて候補チェーンを形成し、上記候補チェーンのそれぞれ毎に数的基準に対応する品質測定基準を算出して、上記ビデオ信号の構造パターン解析に対する上記候補チェーンの重要性及び関連性を測定し、上記品質測定基準が所定の品質測定基準閾値を上回る候補チェーンのみを出力すること

を特徴とする請求の範囲第 2 4 項記載の映像音声処理装置。

30. 上記ビデオ信号における映像及び／又は音声セグメントのうち、映像及び／又は音声セグメントが供給された時間順にしたがって当該映像及び／又は音声セグメントを1つずつ逐次処理することを特徴とする請求の範囲第21項記載の映像音声処理装置。

31. 上記検出手段は、対象とする上記現映像及び／又は音声セグメントに関する上記特徴量と上記測定基準とを用いて、当該映像及び／又は音声セグメントを含む候補チェーンを随時更新して求め、上記候補チェーンのそれぞれ毎に数値的基準に対応する品質測定基準を算出して、上記ビデオ信号の構造パターン解析における上記候補チェーンの重要性及び関連性を測定し、上記品質測定基準が所定の品質測定基準閾値を上回る候補チェーンのみを出力すること

を特徴とする請求の範囲第30項記載の映像音声処理装置。

32. 上記検出手段は、周期的チェーンの初期候補を求め、上記周期的チェーンの初期候補の中から、時間的に交差する重複チェーンを求め、上記重複チェーンの整合を求めること

を特徴とする請求の範囲第26項記載の映像音声処理装置。

33. 上記解析手段は、上記類似チェーンを用いて、上記ビデオ信号の局所的ビデオ構造として、映像及び／又は音声セグメントの意味に基づく部分集合であるシーンを検出して出力すること

を特徴とする請求の範囲第21項記載の映像音声処理装置。

34. 上記解析手段は、上記類似チェーンを用いて、上記ビデオ信号の大局的ビデオ構造として、互いに類似する映像及び／又は音声セグメントが反復的に発生する構造パターンを検出して出力すること

を特徴とする請求の範囲第21項記載の映像音声処理装置。

35. 上記解析手段は、上記構造パターンとして、ニュース放送におけるニュース項目を検出して出力すること

を特徴とする請求の範囲第34項記載の映像音声処理装置。

36. 上記解析手段は、上記構造パターンとして、プレイが反復的に発生するスポーツ放送におけるビデオ構造を検出して出力すること

を特徴とする請求の範囲第34項記載の映像音声処理装置。

37. 上記解析手段は、上記類似チェーンを用いて、映像及び／又は音声セグメントの意味に基づく部分集合であるシーンのうち、関連するシーンをまとめたトピック構造を検出して出力すること

を特徴とする請求の範囲第21項記載の映像音声処理装置。

1/15

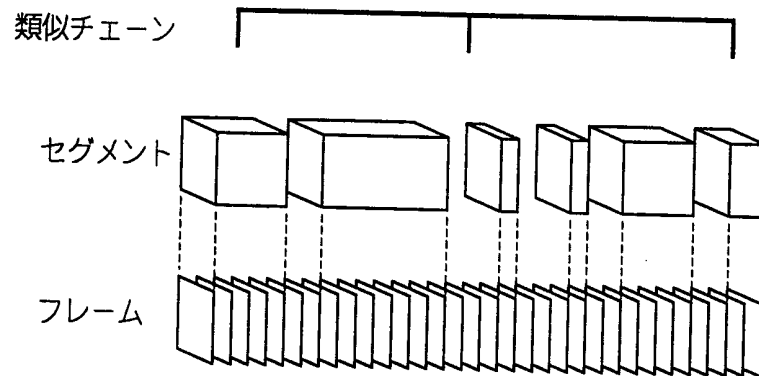


FIG. 1

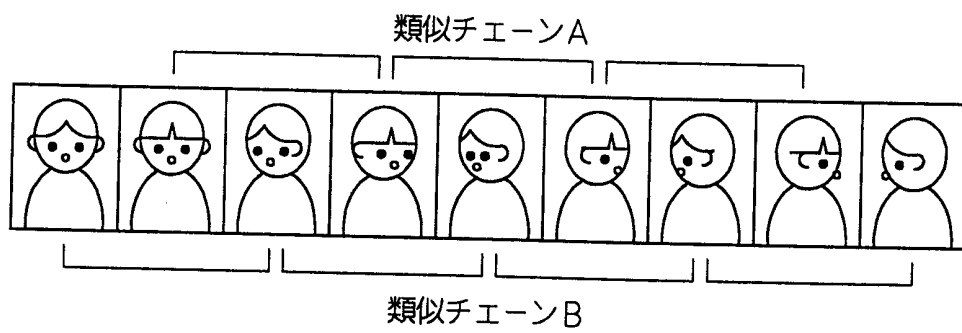


FIG. 2

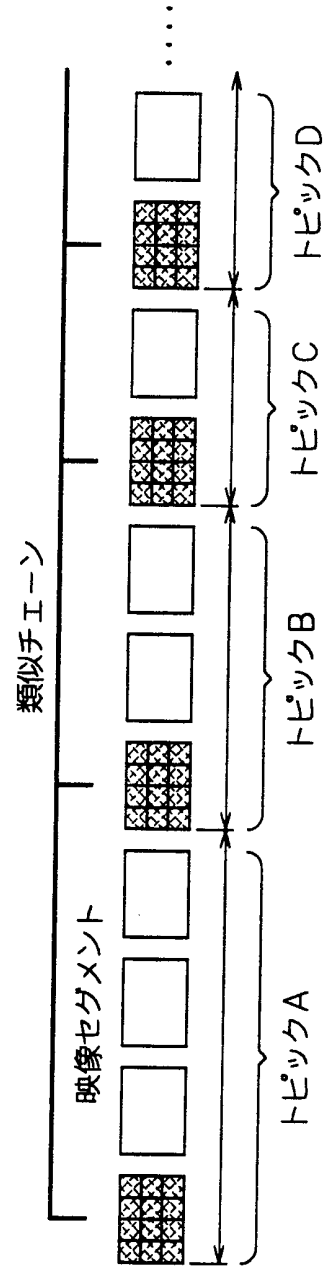


FIG. 3

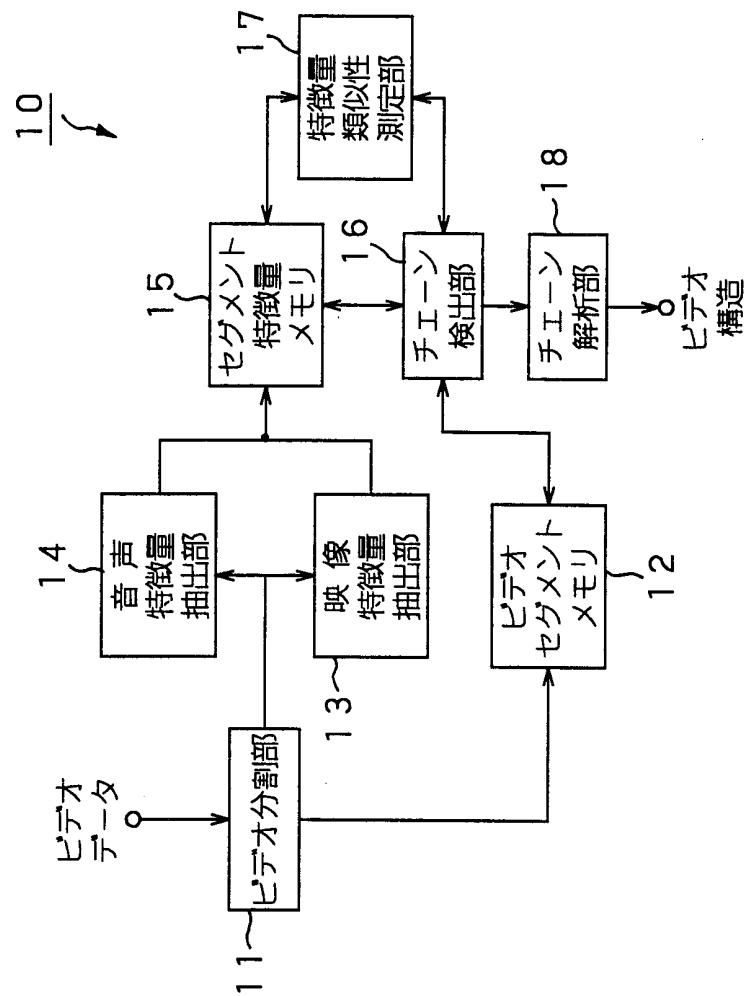


FIG. 4

4/15

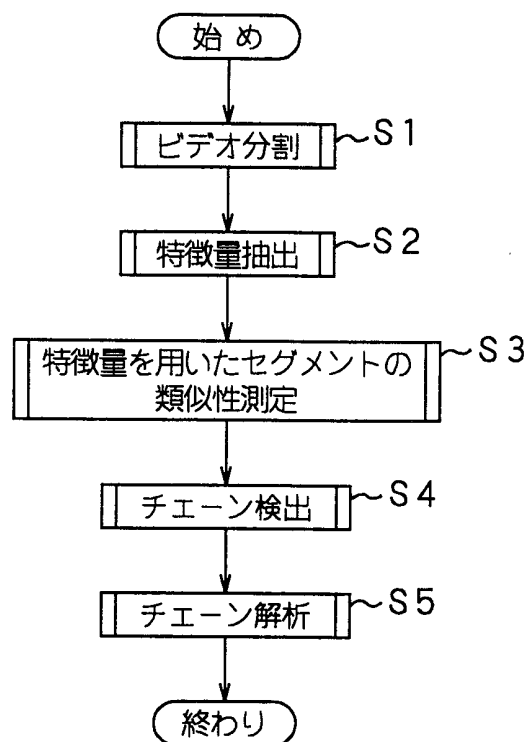


FIG. 5

5/15

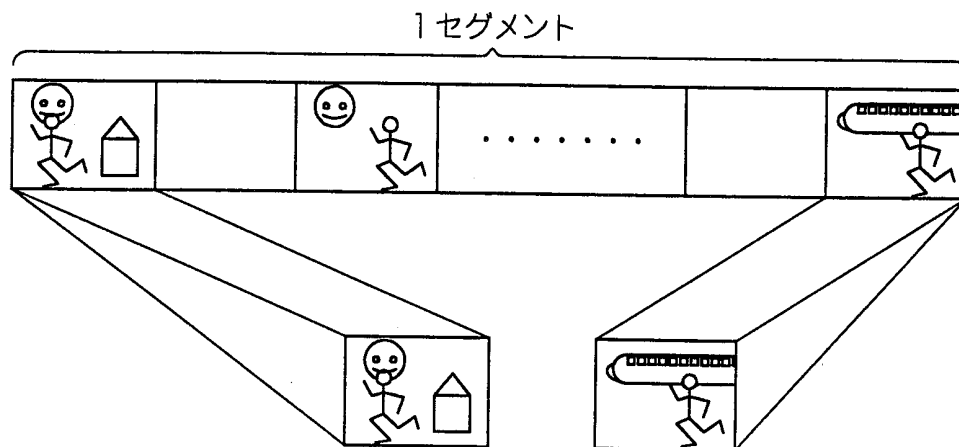


FIG. 6

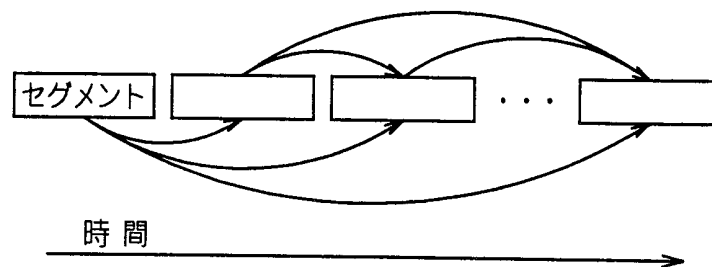


FIG. 7

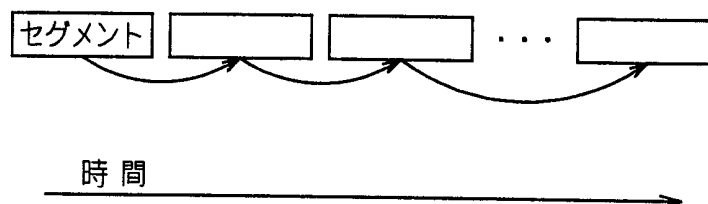


FIG. 8

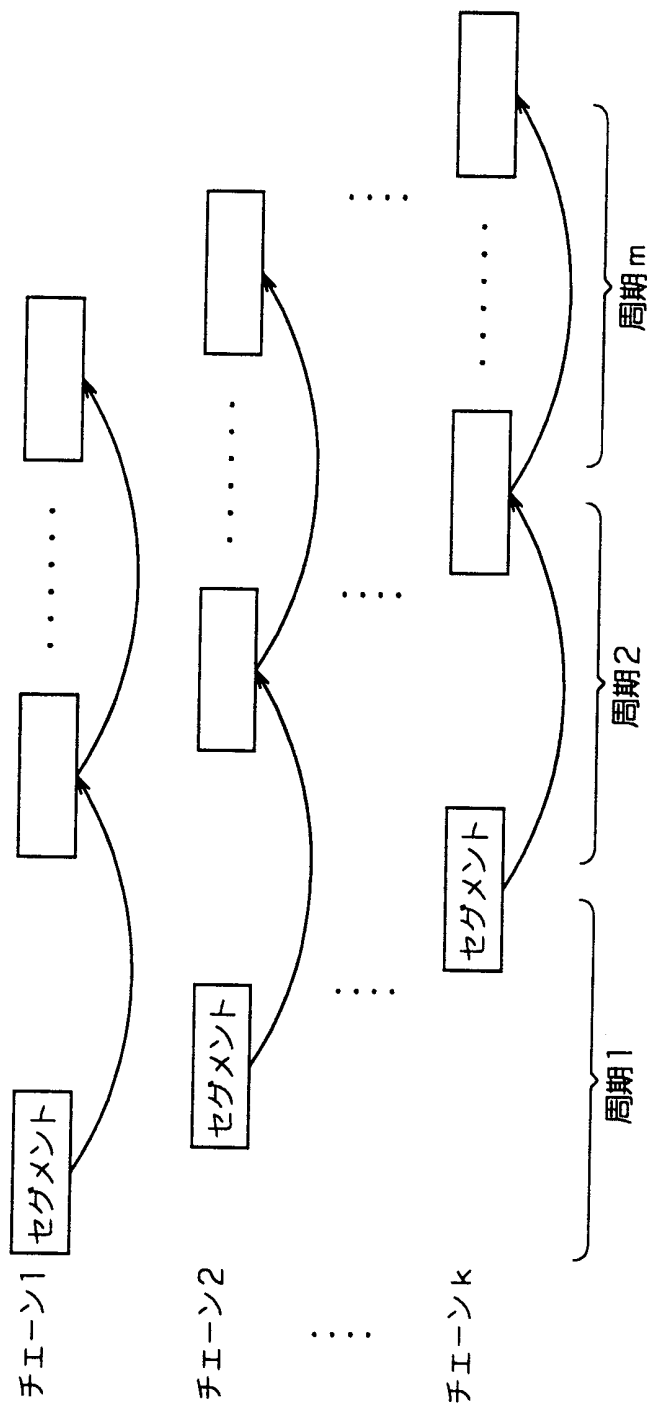


FIG. 9

7/15

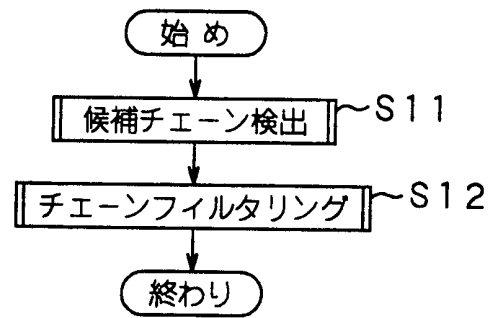


FIG. 10

8/15

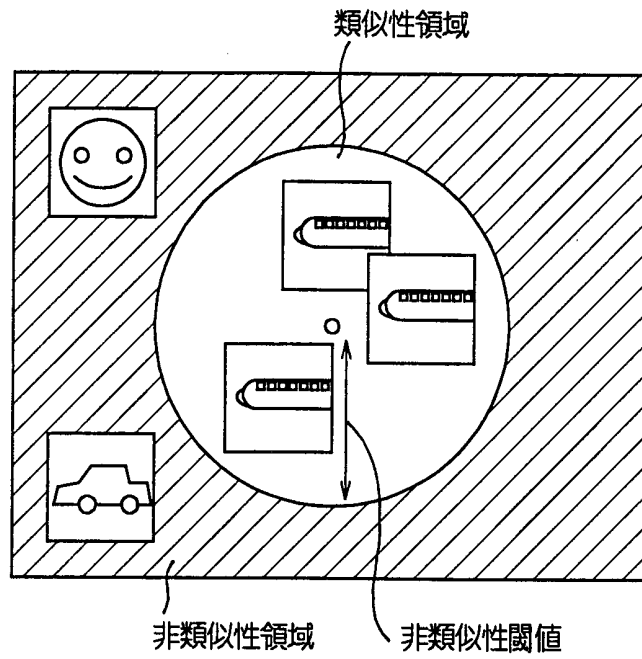


FIG. 11

9/15

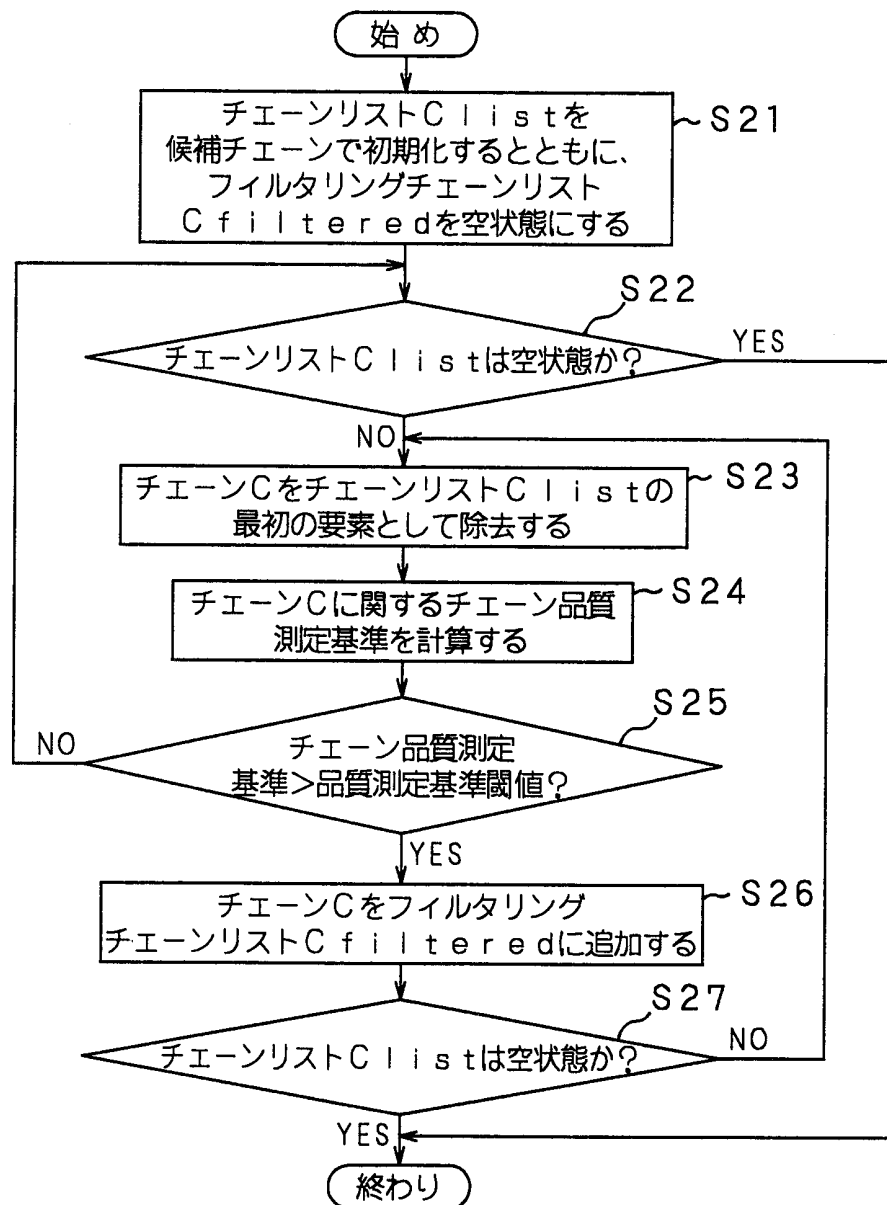


FIG. 12

10/15

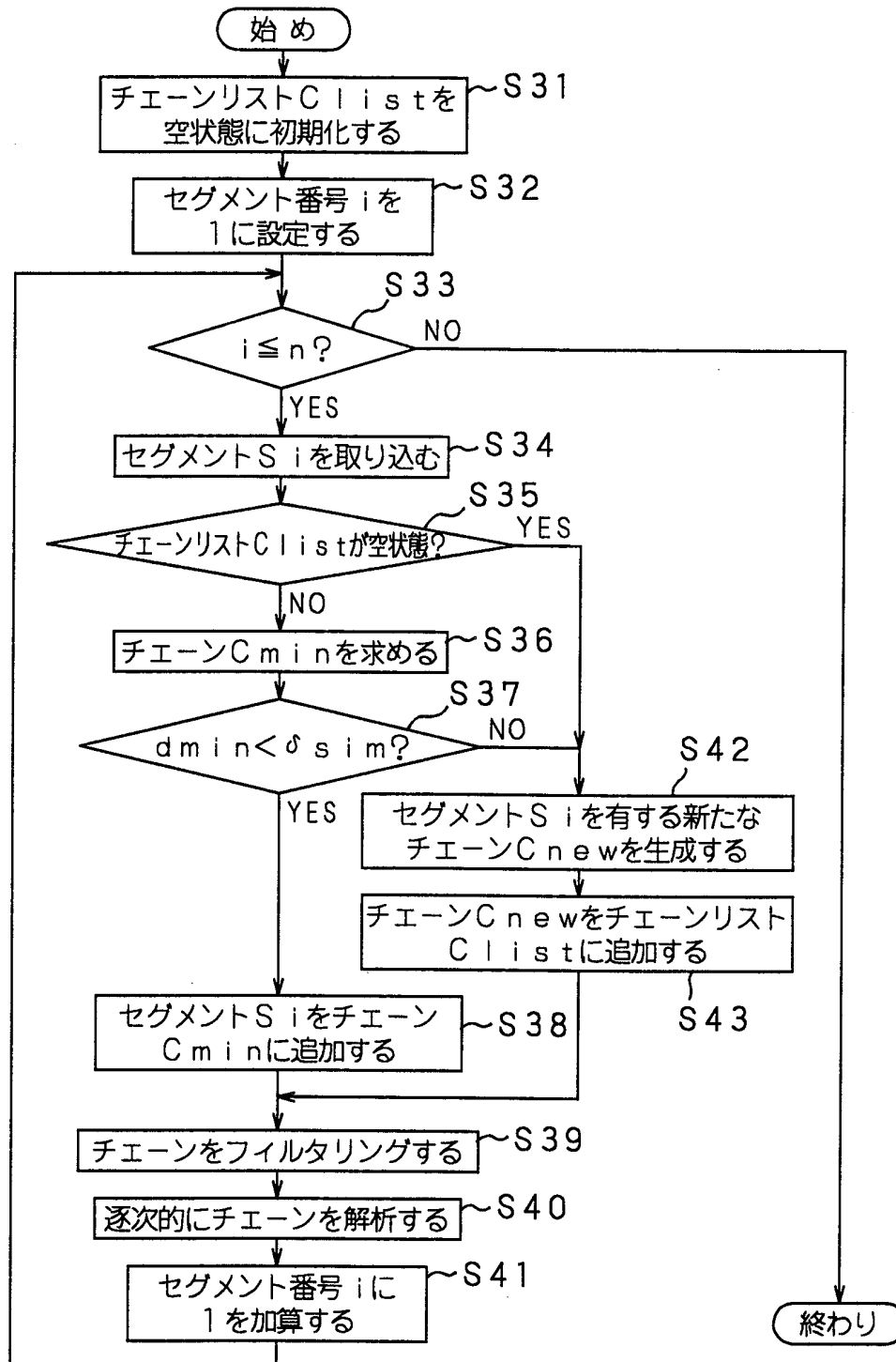


FIG. 13

11/15

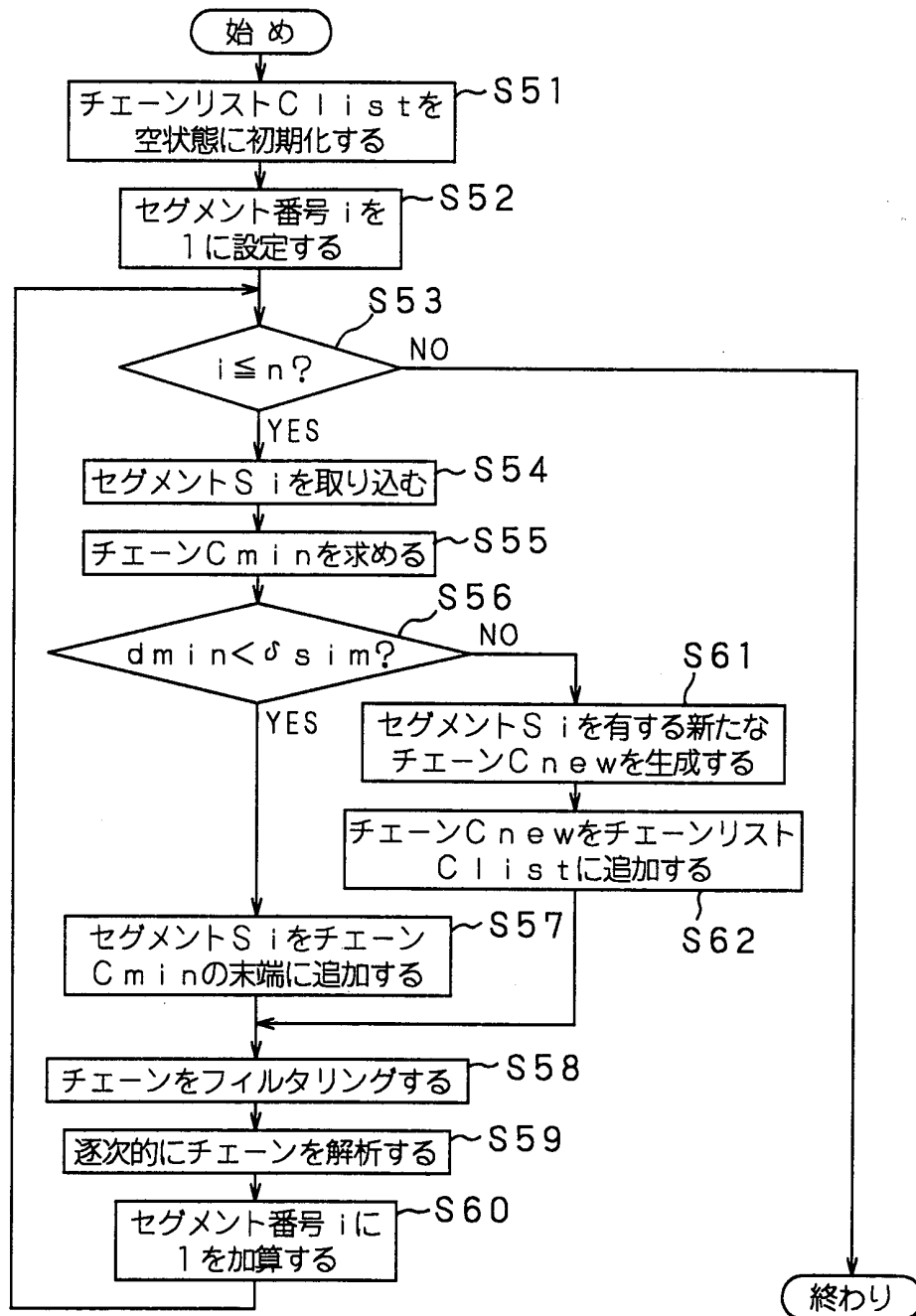


FIG. 14

12/15

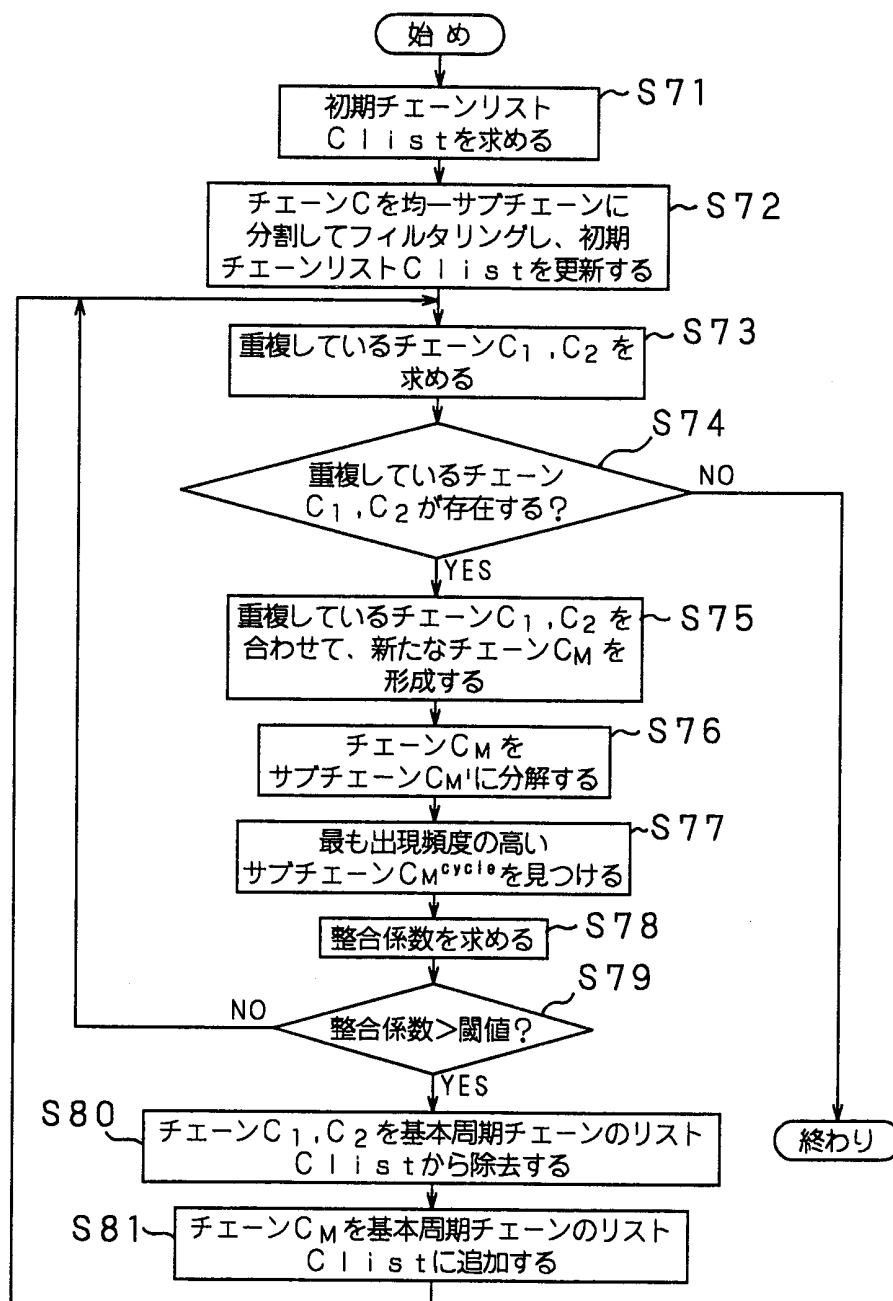


FIG. 15

13/15

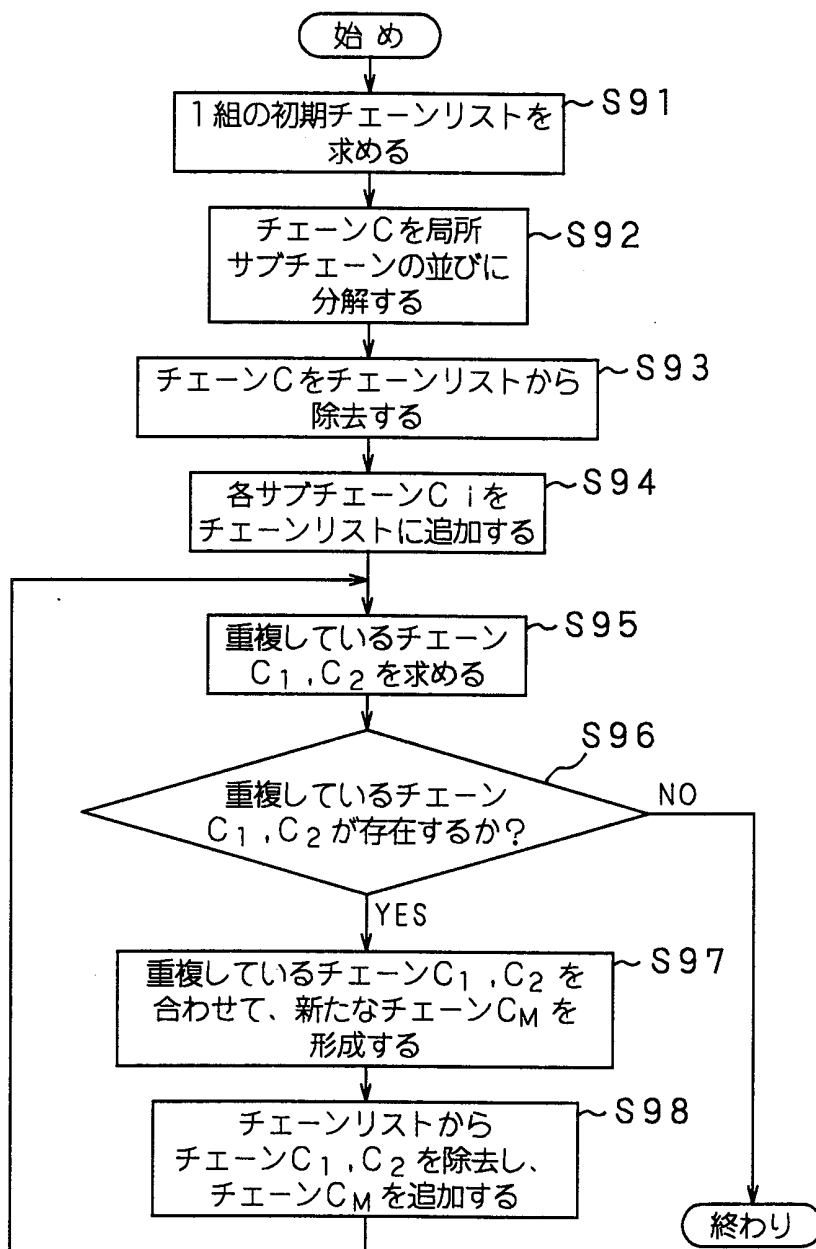


FIG. 16

14/15

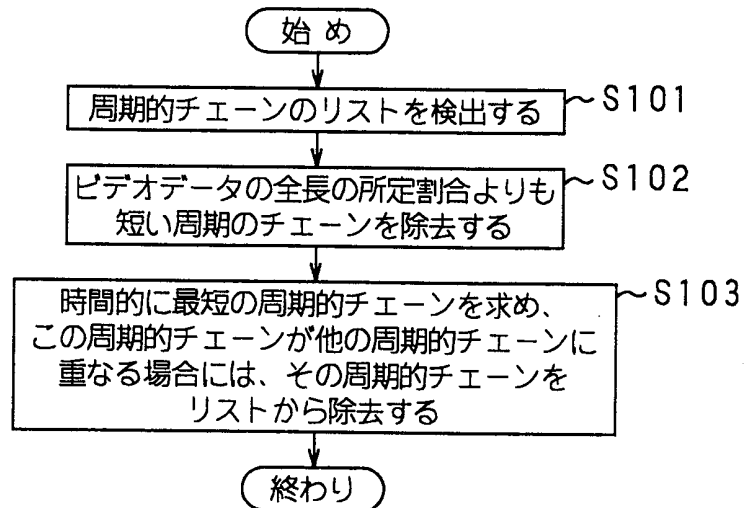


FIG. 17

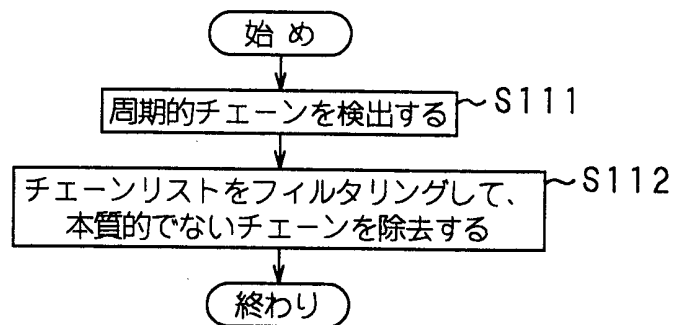


FIG. 18

15/15

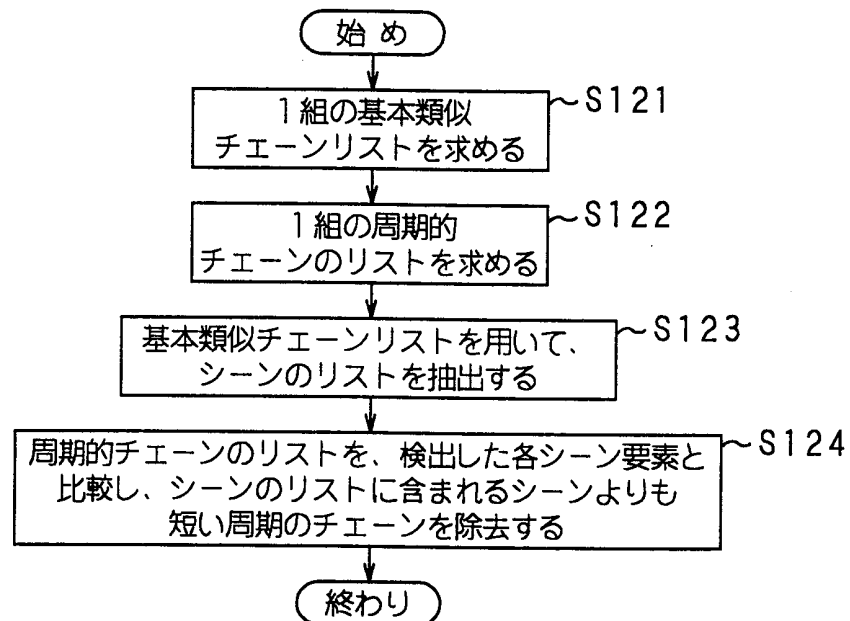


FIG. 19

INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP00/00422

A. CLASSIFICATION OF SUBJECT MATTER

Int.Cl⁷ H04N7/32
 Int.Cl⁷ H04N5/92
 Int.Cl⁷ G06T7/00

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)

Int.Cl⁷ H04N5/91~5/956 Int.Cl⁷ H04N7/24~7/68
 Int.Cl⁷ H04N5/222~5/257 Int.Cl⁷ G06T7/00
 Int.Cl⁷ H04N5/262~5/28

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Jitsuyo Shinan Koho 1922-1996 Toroku Jitsuyo Shinan Koho 1994-2000
 Kokai Jitsuyo Shinan Koho 1971-2000 Jitsuyo Shinan Toroku Koho 1996-2000

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	JP, 2-59976, A (Matsushita Electric Works, Ltd.), 28 February, 1990 (28.02.90), Full text; Figs. 1 to 5 (Family: none)	1, 3-5, 20, 22, 23
A	JP, 10-257436, A (Atsushi Matsushita, Kenichi Okada), 25 September, 1998 (25.09.98), Full text; Figs. 1 to 20 (Family: none)	1-37
A	JP, 7-193748, A (Nippon Telegr. & Teleph. Corp. <NTT>), 28 July, 1995 (28.07.95), Full text; Figs. 1 to 14 (Family: none)	1-37
A	EP, 711078, A (Matsushita Electric Ind. Co., Ltd.), 08 May, 1996 (08.05.96), Full text; all drawings & JP, 8-181995, A & US, 5751377, A	1-37

☐ Further documents are listed in the continuation of Box C.☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier 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

"I" 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
 18 April, 2000 (18.04.00)

Date of mailing of the international search report
 02 May, 2000 (02.05.00)

Name and mailing address of the ISA/
 Japanese Patent Office

Authorized officer

Facsimile No.

Telephone No.

A. 発明の属する分野の分類 (国際特許分類 (IPC)) Int. Cl ⁷ H04N7/32 Int. Cl ⁷ H04N5/92 Int. Cl ⁷ G06T7/00			
B. 調査を行った分野 調査を行った最小限資料 (国際特許分類 (IPC)) Int. Cl ⁷ H04N5/91~5/956 Int. Cl ⁷ H04N7/24~7/68 Int. Cl ⁷ H04N5/222~5/257 Int. Cl ⁷ G06T7/00 Int. Cl ⁷ H04N5/262~5/28			
最小限資料以外の資料で調査を行った分野に含まれるもの 日本国実用新案公報 1922-1996年 日本国公開実用新案公報 1971-2000年 日本国登録実用新案公報 1994-2000年 日本国実用新案登録公報 1996-2000年			
国際調査で使用した電子データベース (データベースの名称、調査に使用した用語)			
C. 関連すると認められる文献			
引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求の範囲の番号	
X	JP, 2-59976, A (松下電工株式会社) 28. 2月. 1990 (28. 02. 90) 全文, 第1-5図 (ファミリーなし)	1, 3-5, 20, 22, 23	
A	JP, 10-257436, A (松下温, 岡田謙一) 25. 9月. 1998 (25. 09. 98) 全文, 第1-20図 (ファミリーなし)	1-37	
☒ C欄の続きにも文献が列挙されている。 ☐ パテントファミリーに関する別紙を参照。			
* 引用文献のカテゴリー 「A」 特に関連のある文献ではなく、一般的技術水準を示すもの 「E」 国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの 「L」 優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献 (理由を付す) 「O」 口頭による開示、使用、展示等に言及する文献 「P」 国際出願日前で、かつ優先権の主張の基礎となる出願日の後に公表された文献 「T」 国際出願日又は優先日後に公表された文献であって出願と矛盾するものではなく、発明の原理又は理論の理解のために引用するもの 「X」 特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの 「Y」 特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの 「&」 同一パテントファミリー文献			
国際調査を完了した日 18. 04. 00		国際調査報告の発送日 02.05.00	
国際調査機関の名称及びあて先 日本国特許庁 (ISA/JP) 郵便番号100-8915 東京都千代田区霞が関三丁目4番3号		特許庁審査官 (権限のある職員) 安田 太 電話番号 03-3581-1101 内線 3531	

C (続き). 関連すると認められる文献		
引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求の範囲の番号
A	J P, 7-193748, A (日本電信電話株式会社) 28. 7月. 1995 (28. 07. 95) 全文, 第1-14図 (ファミリーなし)	1-37
A	EP, 711078, A (松下電器産業株式会社) 8. 5月. 1996 (08. 05. 96) 全文, 全図 & J P, 8-181995, A & US, 5751377, A	1-37