

(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2013-183280
(P2013-183280A)

(43) 公開日 平成25年9月12日(2013.9.12)

(51) Int.Cl.			F I			テーマコード (参考)		
HO4N	5/91	(2006.01)	HO4N	5/91	N	5C053		
HO4N	5/93	(2006.01)	HO4N	5/91	C	5C122		
G11B	27/10	(2006.01)	HO4N	5/93	Z	5D077		
HO4N	5/225	(2006.01)	G11B	27/10	A			
			HO4N	5/225	F			

審査請求 未請求 請求項の数 14 O L (全 23 頁)

(21) 出願番号 特願2012-45673 (P2012-45673)
(22) 出願日 平成24年3月1日(2012.3.1)

(71) 出願人 00004112
株式会社ニコン
東京都千代田区有楽町1丁目12番1号
(74) 代理人 100064908
弁理士 志賀 正武
(74) 代理人 100108578
弁理士 高橋 詔男
(72) 発明者 大槻 正樹
東京都千代田区有楽町一丁目12番1号
株式会社ニコン内
Fターム(参考) 5C053 FA14 GB06 GB09 HA21 JA01
JA30 LA02 LA06
5C122 DA01 EA61 EA65 FH11 FH14
FJ04 GA31 GA34 HA29
5D077 CB04 DC21 DF01 GA03

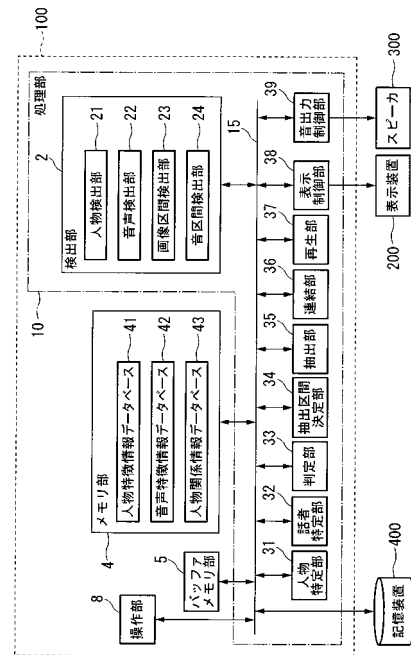
(54) 【発明の名称】 情報処理装置、撮像装置、及びプログラム

(57) 【要約】

【課題】 動画から所望の区間の動画を適切に抽出することができる情報処理装置を提供する。

【解決手段】 情報処理装置は、時系列に続く複数の画像のうち、第1の特徴が検出された画像が時系列に続く第1区間を検出する画像区間検出部と、複数の画像に時系列に対応付けられた音のうち、第2の特徴が検出された音が時系列に続く第2区間を検出する音区間検出部と、第1区間と第2区間とに基づいて、複数の画像および音から抽出する第3区間を決定する抽出区間決定部と、を備える。

【選択図】 図2



【特許請求の範囲】**【請求項 1】**

時系列に続く複数の画像のうち、第 1 の特徴が検出された画像が時系列に続く第 1 区間を検出する画像区間検出部と、

前記複数の画像に時系列に対応付けられた音のうち、第 2 の特徴が検出された音が時系列に続く第 2 区間を検出する音区間検出部と、

前記第 1 区間と前記第 2 区間とに基づいて、前記複数の画像および前記音から抽出する第 3 区間を決定する抽出区間決定部と、

を備えることを特徴とする情報処理装置。

【請求項 2】

前記抽出区間決定部は、

前記第 1 区間の画像および前記第 2 区間の音の何れも途切れないような区間を、前記第 3 区間として決定する。

ことを特徴とする請求項 1 に記載の情報処理装置。

【請求項 3】

前記抽出区間決定部は、

前記第 1 区間、および時間軸上において前記第 1 区間に少なくとも一部が重なる前記第 2 区間のそれぞれ全ての区間を合わせた区間を、前記第 3 区間として決定する

ことを特徴とする請求項 1 または 2 に記載の情報処理装置。

【請求項 4】

前記複数の画像から、前記第 1 の特徴として人物の画像を検出する人物検出部と、

前記音から、前記第 2 の特徴として人物の音声を検出する音声検出部と、

を備えることを特徴とする請求項 1 から 3 に記載の情報処理装置。

【請求項 5】

前記人物検出部により検出された人物の画像から人物の特徴を示す情報を抽出するとともに、抽出した前記人物の特徴を示す情報に基づいて、前記検出された人物の画像が示す人物が何れの人物であるかを特定する人物特定部、

を備え、

前記画像区間検出部は、

前記人物特定部が特定した結果に基づいて、同一の人物が検出された画像が時系列に続く区間を、前記第 1 区間として検出する

ことを特徴とする請求項 4 に記載の情報処理装置。

【請求項 6】

前記音声検出部により検出された音声から音声の特徴を示す情報を抽出するとともに、抽出した前記音声の特徴を示す情報に基づいて、前記検出された音声の話者である人物が何れの人物であるかを特定する話者特定部、

を備え、

前記音区間検出部は、

前記話者特定部が特定した結果に基づいて、同一の人物の音声が検出された音が時系列に続く区間を、前記第 2 区間として検出する

ことを特徴とする請求項 4 または 5 に記載の情報処理装置。

【請求項 7】

前記第 1 区間において検出された人物と、前記第 2 区間において検出された音声の話者である人物とが一致するか否かを判定する判定部、

を備え、

前記抽出区間決定部は、

前記判定部により、前記一致すると判定された場合、

前記第 1 区間と前記第 2 区間とに基づいて、前記複数の画像および前記音から抽出する前記第 3 区間を決定する

ことを特徴とする請求項 4 から 6 の何れか一項に記載の情報処理装置。

10

20

30

40

50

【請求項 8】

複数の人物それぞれが互いに関係があるか否かを示す情報である人物関係情報が記憶部に記憶されており、

前記判定部は、

前記人物関係情報に基づいて、前記第 1 区間において検出された人物と、前記第 2 区間において検出された音声の話者である人物とが前記互いに関係があるか否かを判定し、

前記抽出区間決定部は、

前記判定部により、前記互いに関係があると判定された場合、

前記第 1 区間と前記第 2 区間とに基づいて、前記複数の画像および前記音から抽出する前記第 3 区間を決定する

ことを特徴とする請求項 7 に記載の情報処理装置。

10

【請求項 9】

前記画像区間検出部は、

前記人物関係情報に基づいて、前記互いに関係がある人物が検出された画像が時系列に続く区間を、前記第 1 区間として検出する

ことを特徴とする請求項 8 に記載の情報処理装置。

【請求項 10】

前記音区間検出部は、

前記人物関係情報に基づいて、前記互いに関係がある人物の音声を検出された音が時系列に続く区間を、前記第 2 区間として検出する

ことを特徴とする請求項 8 または 9 に記載の情報処理装置。

20

【請求項 11】

前記第 3 区間の画像および音を抽出する抽出部と、

前記抽出部により抽出された前記第 3 区間の画像および音を再生する再生部と、
を備え、

前記再生部は、

複数の前記第 3 区間の画像および音を、それぞれ時系列の順に従って続けて再生することを特徴とする請求項 1 から 10 の何れか一項に記載の情報処理装置。

【請求項 12】

前記再生部は、

複数の前記第 3 区間のうちの第 1 の第 3 区間、および時系列の順に従った前記第 1 の第 3 区間の次の第 2 の第 3 区間を続けて再生する場合、前記第 1 の第 3 区間内における最後の所定の区間と、前記第 2 の第 3 区間内における最初の所定の区間とを重ねて再生することを特徴とする請求項 11 に記載の情報処理装置。

30

【請求項 13】

請求項 1 から 12 の何れか一項に記載の情報処理装置、
を備えることを特徴とする撮像装置。

【請求項 14】

コンピュータに、

時系列に続く複数の画像のうち、第 1 の特徴が検出された画像が時系列に続く第 1 区間を検出する画像区間検出ステップと、

前記複数の画像に時系列に対応付けられた音のうち、第 2 の特徴が検出された音が時系列に続く第 2 区間を検出する音区間検出ステップと、

前記第 1 区間と前記第 2 区間とに基づいて、前記複数の画像および前記音から抽出する第 3 区間を決定する抽出区間決定ステップと、

を実行させるためのプログラム。

40

【発明の詳細な説明】**【技術分野】****【0001】**

本発明は、情報処理装置、撮像装置、及びプログラムに関する。

50

【背景技術】

【0002】

動画の画像から人物を検出し、検出した人物又はその人物と関係のある人物が記録されている区間の動画を抽出して再生（動画ダイジェスト再生）する情報処理装置がある。（例えば、特許文献1参照）。上述のような情報処理装置によれば、長時間の動画あるいは複数の動画から、例えばユーザに指定された人物又はその人物と関係のある人物が画像に記録されている区間の動画のみを、再生するための動画として抽出することができる。

【先行技術文献】

【特許文献】

【0003】

【特許文献1】特開2011-82915号公報

【発明の概要】

【発明が解決しようとする課題】

【0004】

しかしながら、特許文献1に示す処理装置にあっては、動画に含まれている音声には考慮されていないため、所望の画像が記録されている区間の動画を抽出した際に音声が発話の途中で途切れてしまう場合があるという問題があった。

【0005】

本発明は、このような事情に鑑みてなされたもので、その目的は、動画から所望の区間の動画を適切に抽出することができる情報処理装置、撮像装置、及びプログラムを提供することにある。

【課題を解決するための手段】

【0006】

この発明は上述した課題を解決するためになされたもので、本発明は、時系列に続く複数の画像のうち、第1の特徴が検出された画像が時系列に続く第1区間を検出する画像区間検出部と、前記複数の画像に時系列に対応付けられた音のうち、第2の特徴が検出された音が時系列に続く第2区間を検出する音区間検出部と、前記第1区間と前記第2区間とに基づいて、前記複数の画像および前記音から抽出する第3区間を決定する抽出区間決定部と、を備えることを特徴とする情報処理装置である。

【0007】

また、本発明は、コンピュータに、時系列に続く複数の画像のうち、第1の特徴が検出された画像が時系列に続く第1区間を検出する画像区間検出ステップと、前記複数の画像に時系列に対応付けられた音のうち、第2の特徴が検出された音が時系列に続く第2区間を検出する音区間検出ステップと、前記第1区間と前記第2区間とに基づいて、前記複数の画像および前記音から抽出する第3区間を決定する抽出区間決定ステップと、を実行させるためのプログラムである。

【発明の効果】

【0008】

この発明によれば、動画から所望の区間の動画を適切に抽出することができる。

【図面の簡単な説明】

【0009】

【図1】本発明の第1実施形態による情報処理装置の概要を説明する図である。

【図2】第1実施形態による情報処理装置の構成の一例を示す概略ブロック図である。

【図3】情報処理装置における抽出処理の一例を示すフローチャートである。

【図4】第1区間、第2区間、及び第3区間の時間軸上の関係の第1例を表した図である。

【図5】第1区間、第2区間、及び第3区間の時間軸上の関係の第2例を表した図である。

【図6】第1区間、第2区間、及び第3区間の時間軸上の関係の第3例を表した図である。

10

20

30

40

50

【図 7】第 1 区間、第 2 区間、及び第 3 区間の時間軸上の関係の第 4 例を表した図である。

【図 8】情報処理装置における抽出処理の別の一例を示すフローチャートである。

【図 9】第 1 区間、第 2 区間、及び第 3 区間の時間軸上の関係の第 5 例を表した図である。

【図 10】第 1 区間、第 2 区間、及び第 3 区間の時間軸上の関係の第 6 例を表した図である。

【図 11】情報処理装置における連結処理の第 1 例を説明する図である。

【図 12】情報処理装置における連結処理の第 2 例を説明する図である。

【図 13】第 2 実施形態による撮像装置の構成の一例を示す概略ブロック図である。

10

【発明を実施するための形態】

【0010】

以下、図面を参照して、本発明の実施の形態について説明する。

< 第 1 実施形態 >

図 1 は、本発明の第 1 実施形態による情報処理装置 100 の概要を説明する図である。図 1 に示す情報処理装置 100 は、表示装置 200 と、スピーカ 300 と、記憶装置 400 と、それぞれ接続されている。これらの接続は、有線による接続であってもよいし、無線による接続であってもよい。なお、表示装置 200、スピーカ 300、又は記憶装置 400 は、情報処理装置 100 に含まれている構成であってもよい。

【0011】

20

情報処理装置 100 は、情報処理を実行する処理部 10 と、処理部 10 が処理を実行する上で用いるデータベース等が記憶されているメモリ部 4 と、処理を実行する上での作業領域となるバッファメモリ部 5 と、を備えている。例えば、情報処理装置 100 は、記憶装置 400 に記憶されている動画を読み出して、読み出した動画を画像信号と音声信号とにデコードする。また、情報処理装置 100 は、この画像信号を表示装置 200 に出力するとともに、音声信号をスピーカ 300 に出力する。すなわち、情報処理装置 100 は、記憶装置 400 に記憶されている動画を再生して、表示装置 200 及びスピーカ 300 に出力する。

【0012】

ここで、動画は、例えば、時系列に続く複数のフレーム画像（複数の画像）と、この複数のフレーム画像に時系列に対応付けられている音とを含んで構成されている。すなわち、このフレーム画像と音とは互いに同じフレーム周期で同期し、互いに時間軸において対応している。

30

【0013】

なお、動画を構成するフレーム画像及び音のそれぞれは、互いに時間軸において対応付けが可能な互いに異なるファイルに保存されていてもよい。また、複数のフレーム画像は、複数の静止画像を用いて構成されていてもよい。

【0014】

本実施形態の情報処理装置 100 は、例えば、動画から特定のシーン（例えば、ユーザが見たいシーン）のみを抽出して再生（所謂、ダイジェスト再生）する場合に、動画から抽出する区間を、複数のフレーム画像と音との何れの区間も適切になるように抽出する。

40

【0015】

このような処理を行う情報処理装置 100 の構成について、図 2 を参照して、詳しく説明する。

（情報処理装置の構成）

図 2 は、第 1 実施形態による情報処理装置 100 の構成の一例を示す概略ブロック図である。この図 2 を参照して、本実施形態の情報処理装置 100 の構成についてさらに詳しく説明する。なお、図 2 において、図 1 の各部に対応する構成には同一の符号を付けている。

【0016】

50

情報処理装置 100 は、処理部 10 と、メモリ部 4（記憶部）と、バッファメモリ部 5 と、操作部 8 と、を備えている。処理部 10 は、検出部 2 と、人物特定部 31 と、話者特定部 32 と、判定部 33 と、抽出区間決定部 34 と、抽出部 35 と、連結部 36 と、再生部 37 と、表示制御部 38 と、音出力制御部 39 と、を備えている。

【0017】

まず、処理部 10 の検出部 2 について説明する。

検出部 2 は、人物検出部 21 と、音声検出部 22 と、画像区間検出部 23 と、音区間検出部 24 とを備えており、動画うちの、人物が記録されている区間、又は音声記録されている区間を検出する。

【0018】

人物検出部 21 は、複数のフレーム画像それぞれから、人物の画像（第 1 の特徴）を検出する。例えば、人物検出部 21 は、フレーム画像から人物の顔の画像を検出することにより、複数のフレーム画像それぞれに人物の画像が含まれているか否か（人物が記録されているか否か）を検出する。

【0019】

音声検出部 22 は、音から人物の音声（第 2 の特徴）を検出する。すなわち、音声検出部 22 は、動画に含まれているフレーム画像に対応付けられた音から人物の音声を検出する。例えば、音声検出部 22 は、フレーム画像のフレーム単位、又は予め設定されたフレーム数単位において、音信号の振幅（大きさ）と周波数成分に基づいて、それぞれのフレームに対応した音に人物の音声が含まれているか否かを検出する。

【0020】

画像区間検出部 23 は、複数のフレーム画像のうち、所定の特徴（第 1 の特徴）が検出されたフレーム画像が時系列に続く区間（以下、第 1 区間）を検出する。例えば、画像区間検出部 23 は、人物検出部 21 により人物の画像が検出されたフレーム画像が時系列に続く第 1 区間を検出する。なお、画像区間検出部 23 は、わずかな区間（少ないフレーム数）において人物の画像が検出されなかったとしても、その前後の区間において人物の画像が継続して検出されていれば、その人物の画像が検出されなかった区間も人物が継続して検出されていた区間としてみなしてもよい。

【0021】

例えば、画像区間検出部 23 は、人物検出部 21 により人物の画像が検出されないフレーム画像が時系列に続く区間が、予め定められた第 1 のフレーム数以下であるか否かに基づいて、人物検出部 21 により人物の画像が検出されたフレーム画像が時系列に続く区間であるか否かを判定する。ここで、第 1 のフレーム数とは、人物の画像が継続して検出されている区間であるか否か（人物の画像が検出されている第 1 区間が継続しているか否か）を判定するために予め定められたフレーム数である。

【0022】

例えば、人物の画像が検出されないフレーム画像が時系列に続く区間が上述の第 1 のフレーム数以下である場合、画像区間検出部 23 は、人物の画像が検出されたフレーム画像が時系列に続く区間である（人物が検出されている第 1 区間が継続している）と判定する。一方、人物の画像が検出されないフレーム画像が時系列に続く区間が上述の第 1 のフレーム数を超えた場合、画像区間検出部 23 は、人物の画像が検出されたフレーム画像が時系列に続く区間（人物が検出されている第 1 区間）が終了したと判定する。すなわち、画像区間検出部 23 は、人物の画像が検出されない区間が第 1 のフレーム数以下の短い区間である場合、第 1 区間が継続していると判定し、人物の画像が検出されない区間が第 1 のフレーム数を超えて長い区間となった場合、第 1 区間が終了したと判定する。

【0023】

なお、画像区間検出部 23 は、単に人物の画像が検出されたフレーム画像が時系列に続く区間を第 1 区間として検出してもよいが、同一の人物の画像が検出された画像が時系列に続く区間を、第 1 区間として検出してもよい。例えば、画像区間検出部 23 は、検出された人物の画像が何れの人物の画像であるかを人物特定部 31 が特定した結果に基づいて

10

20

30

40

50

、同一の人物の画像が検出された画像が時系列に続く区間を、第1区間として検出してもよい。

【0024】

また、画像区間検出部23は、互いに関係がある人物の画像が検出されたフレーム画像が時系列に続く区間を、第1区間として検出してもよい。例えば、画像区間検出部23は、検出された人物の画像が何れの人物であるかを人物特定部31が特定した結果と、後述する、複数の人物それぞれが互いに関係があるか否かを示す情報である人物関係情報が登録されている人物関係情報データベース43と、に基づいて、互いに関係がある人物の画像が検出されたフレーム画像が時系列に続く区間を、第1区間として検出してもよい。

【0025】

音区間検出部24は、動画に含まれる音のうち所定の特徴(第2の特徴)が検出された音が時系列に続く区間(以下、第2区間)を検出する。例えば、音区間検出部24は、音声検出部22により検出された人物の音声の時系列に続く第2区間を検出する。なお、音区間検出部24は、わずかな区間(少ないフレーム数)において人物の音声を検出されなかったとしても、その前後の区間において人物が継続して検出されていれば、その人物の音声を検出されなかった区間も人物の音声が続いて検出されていた区間とみなして第2区間を検出してもよい。

【0026】

例えば、音区間検出部24は、音声検出部22により人物の音声を検出されない音が時系列に続く区間が、予め定められた第2のフレーム数以下であるか否かに基づいて、音声検出部22により人物の音声を検出された音が時系列に続く区間であるか否かを判定する。ここで、第2のフレーム数とは、人物の音声が続いて検出されているか否か(人物の音声が続いている第2区間が続いているか否か)を判定するために予め定められたフレーム数である。

【0027】

例えば、音声検出部22により人物の音声を検出されない音が時系列に続く区間が上述の第2のフレーム数以下である場合、音区間検出部24は、人物の音声が続いた音が時系列に続く区間である(人物の音声が続いている第2区間が続いている)と判定する。一方、人物の音声が続かない音が時系列に続く区間が上述の第2のフレーム数を超えた場合、音区間検出部24は、人物の音声が続いた音が時系列に続く区間(人物の音声が続いている第2区間)が終了したと判定する。すなわち、音区間検出部24は、人物の音声が続かない区間が第2のフレーム数以下の短い区間である場合、第2区間が続いていると判定し、人物の音声が続かない区間が第2のフレーム数を超えて長い区間となった場合、第2区間が終了したと判定する。

【0028】

なお、音区間検出部24は、単に人物の音声の時系列に続く区間を第2区間として検出してもよいが、同一の人物の音声が続いた画像が時系列に続く区間を、第2区間として検出してもよい。例えば、音区間検出部24は、検出された人物の音声の話者が何れの人物であるかを話者特定部32が特定した結果に基づいて、同一の人物の音声が続いた音が時系列に続く区間を、第2区間として検出してもよい。

【0029】

また、音区間検出部24は、互いに関係がある人物の音声が続いた画像が時系列に続く区間を、第2区間として検出してもよい。例えば、音区間検出部24は、検出された人物の音声の話者が何れの人物であるかを話者特定部32が特定した結果と、人物関係情報が登録されている人物関係情報データベース43と、に基づいて、互いに関係がある人物の音声が続いた画像が時系列に続く区間を、第2区間として検出してもよい。

【0030】

次に、処理部10の検出部2以外の各部について説明する。

【0031】

人物特定部31は、人物検出部21により検出された人物の画像から、人物の特徴を示

10

20

30

40

50

す情報を抽出するとともに、抽出した人物の特徴を示す情報に基づいて、検出された人物の画像が示す人物が何れの人物であることを特定する。例えば、メモリ部4には、複数の人物の特徴を示す情報が登録されている人物特徴情報データベース41が記憶されている。そして、人物特定部31は、抽出した人物の特徴を示す情報と人物特徴情報データベース41に登録されている複数の人物の特徴を示す情報とに基づいて、検出された人物の画像が示す人物が、人物特徴情報データベース41に登録されている複数の人物のうちの何れの人物であることを特定する。

【0032】

話者特定部32は、音声検出部22により検出された人物の音声から音声の特徴を示す情報を抽出するとともに、抽出した音声の特徴を示す情報に基づいて、検出された音声の話者である人物を特定する。例えば、メモリ部4には、複数の人物の音声の特徴を示す情報が登録されている音声特徴情報データベース42が記憶されている。そして、話者特定部32は、抽出した人物の音声の特徴を示す情報と音声特徴情報データベースに登録されている複数の人物の音声の特徴を示す情報とに基づいて、検出された音声の話者である人物が、音声特徴情報データベース42に登録されている複数の人物の音声うちの何れの人物であることを特定する。

10

【0033】

判定部33は、人物特定部31により第1区間において特定された人物と、話者特定部32により第2区間において特定された音声の話者である人物とが一致するか否かを判定する。例えば、判定部33は、第1区間において特定された人物の固有の識別情報と、第2区間において特定された音声の話者である人物の固有の識別情報とが一致するか否かを判定する。なお、この人物の固有の識別情報は、人物特徴情報データベース41、音声特徴情報データベース42、及び人物関係情報データベース43で人物を識別する情報として用いられている情報であり、後述する。

20

【0034】

また、判定部33は、人物特定部31により第1区間において特定された人物と、話者特定部32により第2区間において特定された音声の話者である人物とが互いに関係があるか否かを判定する。例えば、メモリ部4には、複数の人物それぞれが互いに関係があるか否かを示す情報である人物関係情報データベース43が記憶されている。そして、判定部33は、人物特定部31の特定結果、話者特定部32の特定結果、及び人物関係情報データベースに基づいて、第1区間において特定された人物と、第2区間において特定された音声の話者である人物とが互いに関係があるか否かを判定する。

30

【0035】

抽出区間決定部34は、上述した第1区間と第2区間とに基づいて、動画から抽出する区間(以下、第3区間)を決定する。例えば、抽出区間決定部34は、第1区間の画像及び第2区間の音の何れも途切れないような区間を、動画から抽出する第3区間として決定する。すなわち、抽出区間決定部34は、複数のフレーム画像及びこの複数のフレーム画像に時系列に対応付けられた音から抽出する第3区間を、上述した第1区間と第2区間とに基づいて、第1区間及び第2区間の何れの区間も途切れないような区間に決定する。

【0036】

例えば、ユーザに指定された人物のフレーム画像が続く第1区間の動画を抽出する場合、抽出区間決定部34は、第1区間、及び時間軸上において第1区間に少なくとも一部が重なる第2区間のそれぞれ全ての区間を合わせた区間を、抽出する第3区間として決定する。また、ユーザに指定された人物の音声が続く第2区間の動画を抽出する場合、抽出区間決定部34は、第2区間、及び時間軸上において第2区間に少なくとも一部が重なる第1区間のそれぞれ全ての区間を合わせた区間を、抽出する第3区間として決定する。

40

【0037】

すなわち、抽出区間決定部34は、人物の画像が含まれるフレーム画像が続く第1区間を基準に動画を抽出する際に、人物の音声が含まれる音が途切れないように第2区間も考慮して(合わせて)抽出する第3区間を決定する。一方、抽出部35は、人物の音声が含ま

50

まれる音が続く第2区間を基準に動画を抽出する際に、人物の画像が含まれるフレーム画像が途切れないように第1区間も考慮して（合わせて）抽出する第3区間を決定する。このような抽出区間の決定処理について、詳しくは後述する。

【0038】

抽出部35は、動画から第3区間のフレーム画像及び音を抽出する。例えば、抽出部35は、動画から、1つの第3区間または複数の第3区間のフレーム画像及び音を抽出する。なお、抽出部35は、複数の動画それぞれから、1つの第3区間または複数の第3区間のフレーム画像及び音を抽出してもよい。

【0039】

連結部36は、抽出部35により抽出された複数の第3区間のフレーム画像及び音を、それぞれ時系列の順に従って連結する。例えば、連結部36は、複数の第3区間のフレーム画像及び音を、それぞれ時系列の順に従って連結し、連結した動画を生成する。

10

【0040】

再生部37は、抽出部35により抽出された第3区間のフレーム画像及び音を、再生する。具体的には、再生部37は、抽出部35により抽出された第3区間のフレーム画像を時系列の順に従って画像信号に変換（デコード）して表示制御部38に供給するとともに、フレーム画像に時系列に対応した第3区間の音を音信号に変換（デコード）して、画像信号と時間軸上で同期させて音出力制御部39に供給する。

【0041】

また、再生部37は、抽出部35により抽出された複数の第3区間のフレーム画像及び音を、それぞれ時系列の順に従って続けて再生する。なお、再生部37は、複数の第3区間のフレーム画像及び音を再生する場合、連結部36により複数の第3区間のフレーム画像及び音がそれぞれ時系列の順に従って連結された動画を再生してもよい。

20

【0042】

表示制御部38は、再生部37から受けた画像信号を、表示装置200のフォーマットに対応する信号に変換して出力する。音出力制御部39は、再生部37から受けた音信号を、スピーカ300のフォーマットに対応する信号に変換して出力する。

【0043】

メモリ部4は、人物特徴情報データベース41、音声特徴情報データベース42、及び人物関係情報データベース43を有している。

30

人物特徴情報データベース41には、複数の人物の特徴を示す情報が登録されている。例えば、人物特徴情報データベース41には、複数の人物それぞれの固有の識別情報と、対応する人物の特徴を示す情報とが関連付けられて登録されている。この人物特徴情報データベース41を参照することにより、前述した人物特定部31は、検出された人物の特徴を示す情報に基づいて、固有の識別情報を選択して人物を特定することができる。

【0044】

音声特徴情報データベース42には、複数の人物の音声の特徴を示す情報が登録されている。例えば、音声特徴情報データベース42には、複数の人物それぞれの固有の識別情報と、対応する人物の音声の特徴を示す情報とが関連付けられて登録されている。この音声特徴情報データベース42を参照することにより、前述した話者特定部32は、検出された人物の音声の特徴を示す情報に基づいて、固有の識別情報を選択して話者である人物を特定することができる。

40

【0045】

人物関係情報データベース43には、複数の人物それぞれが互いに関係があるか否かを示す情報である人物関係情報が登録されている。例えば、人物関係情報データベース43には、複数の人物それぞれの互いの関係性や親密性を示す情報が登録されている。具体的には、人物関係情報データベース43には、例えば、互いに関係がある複数の人物の固有の識別情報には同一のフラグ情報が関連付けられて登録されている。なお、このフラグ情報は、人物の固有の識別情報に対して複数種類のフラグ情報が関連付けられていてもよい。すなわち、人物関係情報データベース43には、複数種類の関係が登録されていてもよ

50

い。なお、同一のフラグ情報に代えて、互いの親密度を示す情報としてもよい。

【0046】

バッファメモリ部5は、例えば、揮発性のメモリを有し、処理部10におけるフレーム画像や音に対する検出処理、抽出処理、または信号処理を実行する過程での作業領域として利用される。

【0047】

操作部8は、例えば、操作釦が設けられており操作入力を検出し、検出した操作を示す操作信号を処理部10に供給する。なお、操作部8は、タッチ操作を検出するタッチパネルを有する構成としてもよい。

【0048】

(抽出処理の説明)

このように構成された情報処理装置100において、動画から抽出する第3区間を決定する処理について具体的に説明する。情報処理装置100は、例えば、記憶装置400に記憶された動画から、人物が記録されている区間、ユーザにより指定された人物が記録されている区間、又はユーザにより指定された人物と関係がある人物(例えば関係性、親密性が高い人物)が記録されている区間を抽出して連結した動画を生成する(ダイジェスト動画を生成する)。また、情報処理装置100は、この連結した動画を再生する(ダイジェスト動画を再生する)。

【0049】

ここで、本実施形態の情報処理装置100は、例えば、上述したように人物が記録されている区間(第1区間)を抽出する際に、人物の音声(第2区間)、フレーム画像に含まれている人物の音声(第2区間)、又はフレーム画像に含まれている人物と関係がある人物の音声(第2区間)が途切れないように抽出する。

【0050】

具体的には、画像区間検出部23は、人物(又はユーザに指定された人物や、その指定された人物と関係がある人物)が検出されたフレーム画像が時系列に続く区間を第1区間として検出する。また、音区間検出部24は、音声検出部22により検出された人物の音声(又はユーザに指定された人物の音声や、その指定された人物と関係がある人物の音声)が時系列に続く区間を第2区間として検出する。そして、抽出区間決定部34は、第1区間の画像及び第2区間の音の何れも途切れないような区間を、動画から抽出する第3区間として決定する。例えば、動画から人物が記録されている第1区間を抽出する場合、抽出区間決定部34は、抽出する第3区間を、第1区間、及び時間軸上において第1区間に少なくとも一部が重なる第2区間のそれぞれ全ての区間を合わせた区間に決定する。

【0051】

図3は、情報処理装置100における抽出処理の一例を示すフローチャートである。この図3を参照して、情報処理装置100における抽出区間を決定して抽出する処理の動作の一例を説明する。

【0052】

この図では、情報処理装置100は、ユーザに指定された人物が記録されている第1区間を抽出する際に、第1区間に記録されている人物(第1区間において検出された人物と一致する人物)の音声(第2区間)が途切れないように抽出する。すなわち、情報処理装置100は、上述の第1区間と、時間軸上において第1区間に少なくとも一部が重なる区間であって、第1区間に記録されている人物(第1区間において検出された人物と一致する人物)の音声(第2区間)が記録されている区間である第2区間と、を合わせた第3区間を抽出する。

【0053】

まず、人物検出部21は、動画のフレーム画像から、ユーザに指定された人物の画像を検出する(ステップS11)。また、音声検出部22は、動画の音から人物の音声を検出する(ステップS12)。

【0054】

10

20

30

40

50

次に、画像区間検出部 2 3 は、人物検出部 2 1 により人物の画像が検出されたフレーム画像が時系列に続く第 1 区間を検出する。また、音区間検出部 2 4 は、時間軸上において第 1 区間に少なくとも一部が重なる区間、且つ音声検出部 2 2 により検出された人物の音声の時系列に続く区間である第 2 区間を検出する（ステップ S 1 3）。

【 0 0 5 5 】

続いて、人物特定部 3 1 は、人物検出部 2 1 により検出された人物の画像から、人物の特徴を示す情報を抽出するとともに、抽出した人物の特徴を示す情報に基づいて、検出された人物の画像が示す人物が何れの人物であるかを特定する。また、話者特定部 3 2 は、音声検出部 2 2 により検出された人物の音声から音声の特徴を示す情報を抽出するとともに、抽出した音声の特徴を示す情報に基づいて、検出された音声の話者である人物を特定する（ステップ S 1 4）。

10

【 0 0 5 6 】

判定部 3 3 は、人物特定部 3 1 により第 1 区間において特定された人物と、話者特定部 3 2 により第 2 区間において特定された音声の話者である人物とが一致するか否かを判定する（ステップ S 1 5）。

【 0 0 5 7 】

ステップ S 1 5 において、第 1 区間において特定された人物と、第 2 区間において特定された音声の話者である人物とが一致すると判定された場合、抽出区間決定部 3 4 は、抽出する区間を、第 1 区間及び第 2 区間のそれぞれ全ての区間を合わせた第 3 区間に決定する（抽出する区間を第 1 区間から第 3 区間に変更する、ステップ S 1 6）。そして処理部 1 0 は、ステップ S 1 8 に処理を進める。

20

【 0 0 5 8 】

一方、ステップ S 1 5 において、第 1 区間において特定された人物と、第 2 区間において特定された音声の話者である人物とが一致しないと判定された場合、抽出区間決定部 3 4 は、抽出する区間を、第 1 区間に決定する（ステップ S 1 7）。そして処理部 1 0 は、ステップ S 1 8 に処理を進める。

【 0 0 5 9 】

ステップ S 1 8 において、抽出部 3 5 は、抽出区間決定部 3 4 により決定された区間の動画を抽出する。すなわち、第 1 区間において特定された人物と、第 2 区間において特定された音声の話者である人物とが一致すると判定された場合、抽出部 3 5 は、第 3 区間（第 1 区間及び第 2 区間を合わせた区間）のフレーム画像と音とを抽出する。また、第 1 区間において特定された人物と、第 2 区間において特定された音声の話者である人物とが一致しないと判定された場合、抽出部 3 5 は、第 1 区間のフレーム画像と音とを抽出する。

30

【 0 0 6 0 】

このように、抽出区間決定部 3 4 は、人物の画像が含まれるフレーム画像が続く第 1 区間と、その人物の音声が含まれる音が連続する第 2 区間とに基づいて、動画から抽出する第 3 区間を、上述の第 1 区間及び第 2 区間のそれぞれ全ての区間を合わせた区間に決定する。そして、抽出部 3 5 は、抽出区間決定部 3 4 により決定された区間を抽出する。

【 0 0 6 1 】

これにより、情報処理装置 1 0 0 は、動画から、人物の画像が含まれるフレーム画像が続く第 1 区間を抽出する際に、その人物の音声が含まれる第 2 区間が途切れないように適切に抽出することができる。

40

【 0 0 6 2 】

（抽出区間の具体例）

次に、図 4 ~ 7 を参照して、時間軸上における第 1 区間、第 2 区間、及び第 3 区間の時間軸上の関係を具体的に説明する。図 4 ~ 7 は、第 1 区間、第 2 区間、及び第 3 区間の時間軸上の関係の例を表した図であって、それぞれ、第 1 区間と第 2 区間との時間軸上の関係が異なる例を示している。

【 0 0 6 3 】

図 4 は、第 1 区間、第 2 区間、及び第 3 区間の時間軸上の関係の第 1 例を表した図であ

50

る。この図は、抽出する人物としてユーザに指定された人物 A の画像が含まれるフレーム画像が連続する第 1 区間がフレーム f 3 ~ f 8 であって、この人物 A の音声記録されている第 2 区間がフレーム f 2 ~ f 10であることを示している。

【0064】

この図に示す例において、ユーザに指定された人物 A が記録されている第 1 区間を抽出する場合、情報処理装置 100 は、抽出する第 3 区間を、第 1 区間及び第 2 区間のそれぞれ全ての区間を合わせたフレーム f 2 ~ f 10 の区間に決定する（抽出する区間を第 1 区間から第 3 区間に変更する）。

【0065】

よって、第 1 区間の開始点（フレーム f 3）に対して第 2 区間の開始点（フレーム f 2）の方が前であって、第 1 区間の終了点（フレーム f 8）に対して第 2 区間の終了点（フレーム f 10）の方が後ろである場合、情報処理装置 100 は、第 1 区間を抽出する際に、第 1 区間及び第 2 区間のそれぞれ全ての区間を合わせた区間（すなわち、この例では第 2 区間）である第 3 区間（フレーム f 2 ~ f 10）を抽出する。そのため、情報処理装置 100 は、第 1 区間を抽出する際に、第 1 区間に記録されている人物 A の音声途切れないように（すなわち第 2 区間が途切れないように）最短の所望の区間を抽出することができる。

10

【0066】

図 5 は、第 1 区間、第 2 区間、及び第 3 区間の時間軸上の関係の第 2 例を表した図である。この図は、抽出する人物としてユーザに指定された人物 A の画像が含まれるフレーム画像が連続する第 1 区間がフレーム f 3 ~ f 8 であって、この人物 A の音声記録されている第 2 区間がフレーム f 4 ~ f 7 であることを示している。

20

【0067】

この図に示す例において、ユーザに指定された人物 A が記録されている第 1 区間を抽出する場合、情報処理装置 100 は、抽出する区間を、第 1 区間及び第 2 区間のそれぞれ全ての区間を合わせたフレーム f 3 ~ f 8 の第 3 区間に決定する（抽出する区間を第 1 区間と同じ第 3 区間とする）。

【0068】

よって、第 1 区間の開始点（フレーム f 3）に対して第 2 区間の開始点（フレーム f 4）の方が後ろであって、第 1 区間の終了点（フレーム f 8）に対して第 2 区間の終了点（フレーム f 7）の方が前である場合、情報処理装置 100 は、第 1 区間を抽出する際に、第 1 区間及び第 2 区間のそれぞれ全ての区間を合わせた区間（すなわち、この例では第 1 区間）である第 3 区間（フレーム f 3 ~ f 8）を抽出する。この場合、情報処理装置 100 は、第 1 区間を抽出する第 3 区間としても第 2 区間が途切れることがないため、第 1 区間に記録されている人物 A の音声途切れないように（すなわち第 2 区間が途切れないように）最短の所望の区間を抽出することができる。

30

【0069】

図 6 は、第 1 区間、第 2 区間、及び第 3 区間の時間軸上の関係の第 3 例を表した図である。この図は、抽出する人物としてユーザに指定された人物 A の画像が含まれるフレーム画像が連続する第 1 区間がフレーム f 3 ~ f 8 であって、この人物 A の音声記録されている第 2 区間がフレーム f 4 ~ f 10 であることを示している。

40

【0070】

この図に示す例において、ユーザに指定された人物 A が記録されている第 1 区間を抽出する場合、情報処理装置 100 は、抽出する第 3 区間を、第 1 区間及び第 2 区間のそれぞれ全ての区間を合わせたフレーム f 3 ~ f 10 の区間に決定する（抽出する区間を第 1 区間から第 3 区間に変更する）。

【0071】

よって、第 1 区間の開始点（フレーム f 3）に対して第 2 区間の開始点（フレーム f 4）の方が後ろであって、第 1 区間の終了点（フレーム f 8）に対して第 2 区間の終了点（フレーム f 10）の方が後ろである場合、情報処理装置 100 は、第 1 区間を抽出する際

50

に、第 1 区間及び第 2 区間のそれぞれ全ての区間を合わせた区間である第 3 区間（フレーム f 3 ~ f 10）を抽出する。そのため、情報処理装置 100 は、第 1 区間を抽出する際に、第 1 区間に記録されている人物 A の音声が入り切れないように（すなわち第 2 区間が入り切れないように）最短の所望の区間を抽出することができる。

【0072】

図 7 は、第 1 区間、第 2 区間、及び第 3 区間の時間軸上の関係の第 4 例を表した図である。この図は、抽出する人物としてユーザに指定された人物 A の画像が含まれるフレーム画像が連続する第 1 区間がフレーム f 3 ~ f 8 であって、この人物 A の音声が入り切れている第 2 区間がフレーム f 2 ~ f 7 であることを示している。

【0073】

この図に示す例において、ユーザに指定された人物 A が記録されている第 1 区間を抽出する場合、情報処理装置 100 は、抽出する第 3 区間を、第 1 区間及び第 2 区間のそれぞれ全ての区間を合わせたフレーム f 2 ~ f 8 の区間に決定する（抽出する区間を第 1 区間から第 3 区間に変更する）。

【0074】

よって、第 1 区間の開始点（フレーム f 3）に対して第 2 区間の開始点（フレーム f 2）の方が前であって、第 1 区間の終了点（フレーム f 8）に対して第 2 区間の終了点（フレーム f 7）の方が前である場合、情報処理装置 100 は、第 1 区間を抽出する際に、第 1 区間及び第 2 区間のそれぞれ全ての区間を合わせた区間である第 3 区間（フレーム f 2 ~ f 8）を抽出する。そのため、情報処理装置 100 は、第 1 区間を抽出する際に、第 1 区間に記録されている人物 A の音声が入り切れないように（すなわち第 2 区間が入り切れないように）最短の所望の区間を抽出することができる。

【0075】

なお、図 3 ~ 7 を用いて、情報処理装置 100 が、第 1 区間を抽出する際に、第 1 区間及び第 2 区間のそれぞれ全ての区間を合わせた第 3 区間を抽出する処理について説明したが、第 2 区間を抽出する際に、第 1 区間及び第 2 区間のそれぞれ全ての区間を合わせた第 3 区間を抽出する処理としてもよい。例えば、動画から人物の音声が入り切れている第 2 区間を抽出する場合、抽出区間決定部 34 は、抽出する第 3 区間を、第 2 区間、及び時間軸上において第 2 区間に少なくとも一部が重なる第 1 区間のそれぞれ全ての区間を合わせた区間に決定してもよい。

【0076】

（抽出処理の別の例）

図 8 は、情報処理装置 100 における抽出処理の別の一例を示すフローチャートである。前述の図 3 では、第 1 区間で検出された人物と第 2 区間で検出された音声の人物とが同一である場合に第 3 区間を抽出する処理の動作の一例を説明したが、この図 8 では、第 1 区間で検出された人物と第 2 区間で検出された音声の人物とが関係がある場合に第 3 区間を抽出する処理の動作の一例を説明する。なお、図 8 に示す処理は、図 3 に示すステップ S 15 の処理に代えて、ステップ S 15 a の処理を実行することが異なり、この図 8 において、図 3 の各処理と対応する処理には同一の符号を付け、その説明を省略する。

【0077】

判定部 33 は、人物特定部 31 により第 1 区間において特定された人物と、話者特定部 32 により第 2 区間において特定された音声の話者である人物とが互いに関係があるかを判定する（ステップ S 15 a）。

【0078】

ステップ S 15 a において、第 1 区間において特定された人物と、第 2 区間において特定された音声の話者である人物とが互いに関係があると判定された場合、抽出区間決定部 34 は、抽出する区間を、第 1 区間及び第 2 区間のそれぞれ全ての区間を合わせた第 3 区間に決定する（抽出する区間を第 1 区間から第 3 区間に変更する、ステップ S 16）。そして処理部 10 は、ステップ S 18 に処理を進める。

【0079】

10

20

30

40

50

一方、ステップ S 1 5 a において、第 1 区間において特定された人物と、第 2 区間において特定された音声の話者である人物とが互いに関係がないと判定された場合、抽出区間決定部 3 4 は、抽出する区間を、第 1 区間に決定する（ステップ S 1 7）。そして処理部 1 0 は、ステップ S 1 8 に処理を進める。

【 0 0 8 0 】

このように、抽出区間決定部 3 4 は、人物の画像が含まれるフレーム画像が続く第 1 区間と、その人物と関係がある人物の音声が含まれる音が連続する第 2 区間とに基づいて、動画から抽出する第 3 区間を、上述の第 1 区間及び第 2 区間のそれぞれ全ての区間を合わせた区間に決定する。そして、抽出部 3 5 は、抽出区間決定部 3 4 により決定された区間を抽出する。

10

【 0 0 8 1 】

例えば、情報処理装置 1 0 0 は、ユーザに指定された人物が検出された記録されている第 1 区間を抽出する際に、第 1 区間に記録されている人物と関係がある人物（第 1 区間において検出された人物と関係がある人物）の音声は途切れないように抽出する。すなわち、情報処理装置 1 0 0 は、第 1 区間と、時間軸上において第 1 区間に少なくとも一部が重なる区間であって、第 1 区間に記録されている人物と関係がある人物（第 1 区間において検出された人物と関係がある人物）の音声は記録されている区間である第 2 区間と、を合わせた第 3 区間を抽出する。

【 0 0 8 2 】

これにより、情報処理装置 1 0 0 は、動画から、人物の画像が含まれるフレーム画像が続く第 1 区間を抽出する際に、その人物と関係がある人物の音声が含まれる第 2 区間が途切れないように適切に抽出することができる。

20

【 0 0 8 3 】

図 9 は、第 1 区間、第 2 区間、及び第 3 区間の時間軸上の関係の第 5 例を表した図であり、図 4 に対して、第 1 区間で検出された人物 A と第 2 区間で検出された音声の人物 B とが互い関係がある人物である場合の例を示していることが異なる。また、この図 9 は、抽出する人物としてユーザに指定された人物 A の画像が含まれるフレーム画像が連続する第 1 区間がフレーム f 3 ~ f 8 であって、この人物 A と関係がある人物 B の音声は記録されている第 2 区間がフレーム f 2 ~ f 1 0 であることを示している。

【 0 0 8 4 】

この図に示す例において、ユーザに指定された人物 A が記録されている第 1 区間を抽出する場合、情報処理装置 1 0 0 は、抽出する第 3 区間を、第 1 区間及び第 2 区間のそれぞれ全ての区間を合わせたフレーム f 2 ~ f 1 0 の区間に決定する（抽出する区間を、第 1 区間から第 3 区間に変更する）。

30

【 0 0 8 5 】

よって、第 1 区間のフレーム画像に含まれる人物 A と第 2 区間の音に含まれる音声の話者である人物 B とが互いに関係がある場合、情報処理装置 1 0 0 は、第 1 区間を抽出する際に、第 1 区間及び第 2 区間のそれぞれ全ての区間を合わせた区間（すなわち、この例では第 2 区間）である第 3 区間（フレーム f 2 ~ f 1 0）を抽出する。そのため、情報処理装置 1 0 0 は、第 1 区間を抽出する際に、第 1 区間に記録されている人物 A と関係がある人物 B の音声は途切れないように（すなわち第 2 区間が途切れないように）最短の所望の区間を抽出することができる。

40

【 0 0 8 6 】

図 1 0 は、第 1 区間、第 2 区間、及び第 3 区間の時間軸上の関係の第 6 例を表した図である。また、図 1 0 は、第 1 区間が人物 A 又は人物 A と関係がある人物 B が記録されている区間であって、第 2 区間が人物 A の音声又は人物 B の音声は記録されている区間である場合の例を示している。

【 0 0 8 7 】

この図において、抽出する人物としてユーザに指定された人物 A の画像が含まれるフレーム画像が連続する区間がフレーム f 3 ~ f 5 であって、人物 A と関係がある人物 B の画

50

像が含まれるフレーム画像が連続する第 1 区間がフレーム f 6 ~ f 8 である。ここで、情報処理装置 100 は、人物 A 又は人物 A と関係がある人物 B が記録されている区間であるフレーム f 3 ~ f 8 を、第 1 区間として検出する。

【0088】

また、人物 A の音声が続く区間がフレーム f 2 ~ f 6 であって、人物 A と関係がある人物 B の画像が含まれるフレーム画像が連続する第 1 区間がフレーム f 7 ~ f 10 である。ここで、情報処理装置 100 は、人物 A の音声又は人物 A と関係がある人物 B の音声が記録されている区間であるフレーム f 2 ~ f 10 を、第 2 区間として検出する。

【0089】

ユーザに指定された人物 A 及び人物 A と関係がある人物 B が記録されている第 1 区間を抽出する場合、情報処理装置 100 は、抽出する第 3 区間を、第 1 区間及び第 2 区間のそれぞれ全ての区間を合わせたフレーム f 2 ~ f 10 の第 3 区間に決定する（抽出する区間を第 1 区間から第 3 区間に変更する）。

10

【0090】

よって、第 1 区間のフレーム画像に含まれる人物 A と人物 B とが互いに関係がある場合、情報処理装置 100 は、第 1 区間を抽出する際に、人物 A 又は人物 B の画像を含む第 1 区間、及び人物 A の音声又は人物 B の音声を含む第 2 区間、のそれぞれ全ての区間を合わせた区間（すなわち、この例では第 2 区間）である第 3 区間（フレーム f 2 ~ f 10）を抽出する。そのため、情報処理装置 100 は、人物 A 又は人物 B の画像を含む第 1 区間を抽出する際に、第 1 区間に記録されている人物 A の音声又は人物 B の音声が途切れないように（すなわち第 2 区間が途切れないように）最短の所望の区間を抽出することができる。

20

【0091】

（抽出した複数の区間を連結又は再生する処理）

次に、抽出した複数の第 3 区間を連結する処理について説明する。情報処理装置 100 の連結部 36 は、抽出部 35 により抽出された複数の第 3 区間の画像及び音を、それぞれ時系列の順に従って連結する。

【0092】

図 11 は、情報処理装置 100 における連結処理の第 1 例を説明する図である。この図は、第 1 の第 3 区間（図 11 (a) の第 3 区間 (1)）と第 2 の第 3 区間（図 11 (b) の第 3 区間 (2)）との 2 つの第 3 区間を連結する処理を示している。

30

【0093】

図 11 (a) に示す第 1 の第 3 区間は、人物 A が記録されている第 1 の第 1 区間（第 1 区間 (1)：フレーム f 3 ~ f 8）と、人物 A の音声が続く第 1 の第 2 区間（第 2 区間 (1)：フレーム f 2 ~ f 10）とのそれぞれ全ての区間を合わせた区間（第 3 区間 (1)：フレーム f 2 ~ f 10）である。

【0094】

また、図 11 (b) に示す第 2 の第 3 区間は、人物 B（人物 A と関係がある人物 B）が記録されている第 2 の第 1 区間（第 1 区間 (2)：フレーム f 23 ~ f 28）と、人物 B の音声が続く第 2 の第 2 区間（第 2 区間 (2)：フレーム f 22 ~ f 30）とのそれぞれ全ての区間を合わせた区間（第 3 区間 (2)：フレーム f 22 ~ f 30）である。

40

【0095】

情報処理装置 100 の連結部 36 は、時系列の順に従って、第 1 の第 3 区間の最後のフレーム f 10 の画像及び音と、第 2 の第 3 区間の最初のフレーム f 22 の画像及び音とを連結する。

【0096】

よって、連結部 36 は、抽出部 35 が抽出した複数の第 3 区間を時系列の順に従って連結することができる。

【0097】

50

次に、図 1 2 を参照して、抽出した複数の第 3 区間を連結する処理の別の例について説明する。図 1 2 は、情報処理装置 1 0 0 における連結処理の第 2 例を説明する図である。

【 0 0 9 8 】

この図に示す処理は、図 1 1 に示す処理と同様に第 1 の第 3 区間と第 2 の第 3 区間との 2 つの第 3 区間を連結する処理を示しているが、連結する際に第 1 の第 3 区間及び第 2 の第 3 区間のそれぞれ一部の区間を重ねて連結する点が図 1 1 に示す処理と異なる。

【 0 0 9 9 】

図 1 2 (a) に示す第 1 の第 3 区間は、人物 A が記録されている第 1 の第 1 区間 (第 1 区間 (1) : フレーム f 3 ~ f 8) と、人物 A の音声記録されている第 1 の第 2 区間 (第 2 区間 (1) : フレーム f 2 ~ f 1 0) とのそれぞれ全ての区間を合わせた区間 (第 3 区間 (1) : フレーム f 2 ~ f 1 0) である。ここで、第 1 の第 3 区間における最後のフレーム f 9 ~ f 1 0 の区間 (符号 1 0 1 参照) においては、第 1 の第 2 区間 (第 2 区間 (1)) が含まれ、且つ第 1 の第 1 区間 (第 1 区間 (1)) が含まれていない。

10

【 0 1 0 0 】

一方、図 1 2 (b) に示す第 2 の第 3 区間は、人物 B (人物 A と関係がある人物 B) が記録されている第 2 の第 1 区間 (第 1 区間 (2) : フレーム f 2 3 ~ f 2 8) と、人物 B の音声記録されている第 2 の第 2 区間 (第 2 区間 (2) : フレーム f 2 5 ~ f 3 0) とのそれぞれ全ての区間を合わせた区間 (第 3 区間 (2) : フレーム f 2 3 ~ f 3 0) である。ここで、第 2 の第 3 区間における最初のフレーム f 2 3 ~ f 2 4 の区間 (符号 1 0 2 参照) においては、第 2 の第 1 区間 (第 1 区間 (2)) が含まれ、且つ第 2 の第 2 区間 (第 2 区間 (2)) が含まれていない。

20

【 0 1 0 1 】

このような場合、情報処理装置 1 0 0 の連結部 3 6 は、第 1 の第 3 区間における最後のフレーム f 9 ~ f 1 0 の区間と、第 2 の第 3 区間における最初のフレーム f 2 3 ~ f 2 4 の区間と、を重ねて連結する。具体的には、図 1 2 (c) に示すように、第 1 の第 3 区間における最後のフレーム f 9 ~ f 1 0 の区間の音と、第 2 の第 3 区間における最初のフレーム f 2 3 ~ f 2 4 の区間のフレーム画像と、を重ねて連結する。すなわち、第 1 の第 3 区間における最後のフレーム f 9 ~ f 1 0 と、第 2 の第 3 区間における最初のフレーム f 2 3 ~ f 2 4 と、が重ねて連結された区間 (符号 1 0 3 参照) は、フレーム画像が第 2 の第 3 区間のフレーム f 2 3 ~ f 2 4 の画像であって、音が第 1 の第 3 区間のフレーム f 9 ~ f 1 0 の区間の音である。

30

【 0 1 0 2 】

このように、連結部 3 6 は、複数の第 3 区間のうちの第 1 の第 3 区間、及び時系列の順に従った第 1 の第 3 区間の次の第 2 の第 3 区間を連結する場合、第 1 の第 3 区間内における最後の所定の区間と、第 2 の第 3 区間内における最初の所定の区間とを重ねて連結する。

【 0 1 0 3 】

例えば、図 1 2 に示すように、第 1 の第 3 区間には、第 1 の第 1 区間及び第 1 の第 2 区間が含まれており、この第 1 の第 3 区間における最後の所定の区間において、第 1 の第 2 区間が含まれ、且つ第 1 の第 1 区間が含まれていないとする。また、第 2 の第 3 区間には、第 2 の第 1 区間及び第 2 の第 2 区間が含まれており、この第 2 の第 3 区間における最初の所定の区間において、第 2 の第 1 区間が含まれ、且つ第 2 の第 2 区間が含まれていないとする。このような場合、連結部 3 6 は、第 1 の第 3 区間における最後の所定の区間内の第 1 の第 2 区間と、第 2 の第 3 区間における最初の所定の区間内の第 2 の第 1 区間とのそれぞれ少なくとも一部を重ねて連結する。

40

【 0 1 0 4 】

なお、以下のようにしてもよい。

例えば、第 1 の第 3 区間には、第 1 の第 1 区間及び第 1 の第 2 区間が含まれており、この第 1 の第 3 区間における最後の所定の区間において、第 1 の第 1 区間が含まれ、且つ第 1 の第 2 区間が含まれていないとする。また、第 2 の第 3 区間には、第 2 の第 1 区間及び

50

第2の第2区間を含まれており、この第2の第3区間における最初の所定の区間において、第2の第2区間が含まれ、且つ第2の第1区間が含まれていないとする。このような場合、連結部36は、第1の第3区間における最後の所定の区間内の第1の第1区間と、第2の第3区間における最初の所定の区間内の第2の第2区間とのそれぞれ少なくとも一部を重ねて連結する。

【0105】

これにより、連結部36は、抽出部35が抽出した複数の第3区間を時系列の順に従って連結する際に、複数の第3区間における最初の所定の区間又は最後の所定の区間に不要な画像又は音が含まれている場合には、有用な画像又は有用な音声を用いて、第1区間と第2区間とを重ねて（オーバーラップさせて）連結することができる。よって、情報処理装置100は、有用な画像又は音声途切れることなく、且つ不要な画像又は音を極力抽出しないように動画を抽出することができる。

10

【0106】

なお、図12では、第1の第3区間における最後のフレームf9～f10の区間（符号101参照）と、第2の第3区間における最初のフレームf23～f24の区間（符号102参照）と、のそれぞれの区間の長さが同じ場合を例としているが、これらの区間の長さは互いに異なってもよい。すなわち、これらの区間の長さが互いに異なっていたとしても、連結部36は、第1の第3区間における最後の所定の区間内の第1の第1区間と、第2の第3区間における最初の所定の区間内の第2の第2区間とのそれぞれ少なくとも一部を重ねて連結してもよい。

20

【0107】

また、連結部36は、連結後の第3区間のフレーム画像と音とを、1つの動画ファイル（ダイジェスト動画ファイル）として生成する。また、再生部37は、複数の第3区間を時系列の順に従って連結した動画ファイルを生成し、その生成した動画ファイルを再生（ダイジェスト動画再生）する。

【0108】

これにより、情報処理装置100は、動画から抽出した複数の第3区間のフレーム画像と音とを、時系列の順に従って連結した1つの動画ファイル（ダイジェスト動画ファイル）として生成することができる。また、情報処理装置100は、動画から抽出した複数の第3区間のフレーム画像と音とを、時系列の順に従って再生（ダイジェスト動画再生）することができる。よって、情報処理装置100は、動画から所望の区間の動画を適切に抽出して再生することができる。

30

【0109】

なお、再生部37は、複数の第3区間を時系列の順に従って連結した動画ファイルを生成しないで、単に、複数の第3区間の画像及び音を時系列の順に従って続けて再生してもよい。

【0110】

また、再生部37は、複数の第3区間のうちの第1の第3区間、及び時系列の順に従った第1の第3区間の次の第2の第3区間を再生する場合、第1の第3区間内における最後の所定の区間と、第2の第3区間内における最初の所定の区間とを重ねて再生してもよい。

40

【0111】

例えば、図12に示すように、第1の第3区間には、第1の第1区間及び第1の第2区間が含まれており、この第1の第3区間における最後の所定の区間において、第1の第2区間が含まれ、且つ第1の第1区間が含まれていないとする。また、第2の第3区間には、第2の第1区間及び第2の第2区間が含まれており、この第2の第3区間における最初の所定の区間において、第2の第1区間が含まれ、且つ第2の第2区間が含まれていないとする。このような場合、再生部37は、第1の第3区間における最後の所定の区間内の第1の第2区間と、第2の第3区間における最初の所定の区間内の第2の第1区間とのそれぞれ少なくとも一部を重ねて再生してもよい。

50

【0112】

また、以下のようにしてもよい。

例えば、第1の第3区間には、第1の第1区間及び第1の第2区間が含まれており、この第1の第3区間における最後の所定の区間において、第1の第1区間が含まれ、且つ第1の第2区間が含まれていないとする。また、第2の第3区間には、第2の第1区間及び第2の第2区間が含まれており、この第2の第3区間における最初の所定の区間において、第2の第2区間が含まれ、且つ第2の第1区間が含まれていないとする。このような場合、再生部37は、第1の第3区間における最後の所定の区間内の第1の第1区間と、第2の第3区間における最初の所定の区間内の第2の第2区間とのそれぞれ少なくとも一部を重ねて再生してもよい。

10

【0113】

以上、説明してきたように、第1実施形態の情報処理装置100は、動画から所望の区間の動画を適切に抽出することができる。

【0114】

<第2実施形態>

次に、第2実施形態について説明する。第2実施形態は、本発明の情報処理装置を含む撮像装置の態様である。

(撮像装置の構成)

図13は、本発明の第2実施形態による撮像装置500の構成の一例を示す概略ブロック図である。この図13を参照して、本実施形態の撮像装置500の構成について説明する。なお、撮像装置500は、図2を用いて説明した情報処理装置100が備える各構成を含んでおり、その対応する構成には同一の符号をつけておりその説明を省略する。

20

【0115】

図13に示す撮像装置500は、検出部2、情報処理部3、メモリ部4、バッファメモリ部5、撮像部6、制御部7、操作部8、記憶媒体制御部9、表示制御部38、音出力制御部39、表示部11、音出力部12、及びバス15を備えている。

【0116】

ここで、図13に示す検出部2、情報処理部3、メモリ部4、バッファメモリ部5、操作部8、表示制御部38、及び音出力制御部39を含む構成が、図2の情報処理装置100と対応した構成である。なお、情報処理部3は、図2の人物特定部31、話者特定部32、判定部33、抽出区間決定部34、抽出部35、連結部36、及び再生部37を含んで構成されている。また、表示部11は図2の表示装置200と対応し、音出力部12は図2のスピーカ300と対応する。

30

【0117】

記憶媒体制御部9は、取り外しが可能な記憶媒体600(カードメモリ等)と接続され、この記憶媒体600への情報の書込み、読み出し、あるいは消去を行う。記憶媒体600は、撮像装置500に対して着脱可能に接続されるメモリであって、例えば、撮像部6によって撮影された画像、又は情報処理部3によって画像処理が実行された画像などを記憶する。この記憶媒体600は、図2の記憶装置400と対応する。なお、撮像装置500の構成において、記憶媒体600に代えて、通信網を介して接続可能な記憶装置(例えば記憶装置400)としてもよい。

40

【0118】

なお、撮像装置500が備える図2の情報処理装置100と対応した各部は、第1実施形態において説明した処理に加えて、撮像装置500において必要な処理を実行する。例えば、情報処理部3は、撮像装置500が撮像した画像に対して、ホワイトバランス調整、シャープネス調整、ガンマ補正、階調調整などの画像処理を実行するとともに、JPEG(Joint Photographic Experts Group)などの画像圧縮処理を実行する。メモリ部4は、例えば、撮影に関するシーン判定条件や撮影条件、各種の撮影モードの設定条件等、撮像装置500を制御するための制御条件を記憶する。バッファメモリ部5は、撮像部6によって撮像された画像、記憶媒体600から読み出された画像、情報処理部3による画像

50

処理の過程で生成された画像等を、一時的に記憶する。操作部 8 は、例えば、電源スイッチ、シャッターボタン、動作モードを選択するモードダイヤル、などを備えており、ユーザによって操作された操作入力を検出し、検出した操作内容を示す操作信号を制御部 7 又は情報処理部 3 に出力する。

【0119】

撮像部 6 は、レンズ部 6 1、撮像素子 6 2、及び A/D 変換部 6 3 を備えており、動画（動画像）または静止画（静止画像）等を撮像する。この撮像部 6 は、設定された撮像条件（例えば絞り値、露出等）に基づいて制御部 7 により制御され、レンズ部 6 1 を介して入力された光束を、撮像素子 6 2 の撮像面上に結像させる。撮像素子 6 2 は、撮像面に結像された光束を光電変換したアナログ信号を、A/D 変換部 6 3 に出力する。A/D 変換部 6 3 は、撮像素子 6 2 から入力されたアナログ信号をデジタル信号に変換し、この変換したデジタル信号を出力する。なお、上述したレンズ部 6 1 は、撮像装置 5 0 0 と一体化されている構成としてもよいし、撮像装置 5 0 0 に着脱可能に取り付けられている構成としてもよい。

10

【0120】

例えば、撮像部 6 は、操作部 8 のシャッターボタンが操作されることに応じて、静止画の撮像処理、又は動画の撮像処理を実行し、撮像した画像を出力する。そして、撮像部 6 によって撮像された静止画又は動画は、制御部 7 の制御により、バッファメモリ部 5 や情報処理部 3 を介して記憶媒体 6 0 0 に記録される。

【0121】

制御部 7 は、撮像装置 5 0 0 が備えている各部を制御する。例えば、制御部 7 は、操作部 8 から受け取る操作信号、又はメモリ部 4 に記憶されている制御条件等に基づいて、撮像装置 5 0 0 が備えている各部を制御する。

20

【0122】

バス 1 5 は、検出部 2、情報処理部 3、メモリ部 4、バッファメモリ部 5、撮像部 6、制御部 7、操作部 8、記憶媒体制御部 9、表示制御部 3 8、及び音出力制御部 3 9 と接続され、各部から出力された画像信号、音信号、制御信号、等を転送する。

【0123】

以上、図 1 3 を用いて説明した撮像装置 5 0 0 の構成においては、第 1 実施形態の情報処理装置 1 0 0 に対応する構成を有しており、第 1 実施形態において説明した動画の抽出処理と同様の処理が可能である。

30

【0124】

例えば、撮像装置 5 0 0 は撮像して記憶媒体 6 0 0 に記憶させた動画（フレーム画像及び音から）から特定の動画の区間を抽出する際の抽出区間を、前述した第 1 区間と第 2 区間とに基づいて、第 3 区間に決定することができる。また、撮像装置 5 0 0 は、複数の第 3 区間のフレーム画像及び音を、時系列の順に従って連結して 1 つの動画ファイル（ダイジェスト動画ファイル）として生成することができる。また、撮像装置 5 0 0 は、複数の第 3 区間のフレーム画像及び音を時系列の順に従って再生（ダイジェスト動画再生）することができる。

【0125】

これにより、第 2 実施形態の撮像装置 5 0 0 は、動画から所望の区間の動画を適切に抽出することができる。

40

【0126】

なお、上記第 1 実施形態（又は第 2 実施形態）では、情報処理装置 1 0 0（又は撮像装置 5 0 0）が、人物が検出されたフレーム画像が時系列に続く区間を第 1 区間として検出する例、及び、人物の音声を検出された音が時系列に続く区間を第 2 区間として検出する例を説明したが、これに限られるものではなく、情報処理装置 1 0 0（又は撮像装置 5 0 0）は、人物及び人物の音声以外の画像の特徴及び音の特徴に基づいて第 1 区間及び第 2 区間を検出してもよい。例えば、情報処理装置 1 0 0（又は撮像装置 5 0 0）は、画像の特徴（第 1 の特徴）及び音の特徴（第 2 の特徴）として、動物及び動物の鳴き声、楽団及び

50

楽団の演奏音等、に基づいて、第 1 区間及び第 2 区間を検出してよい。

【 0 1 2 7 】

また、上記実施形態の第 1 区間又は第 2 区間は、人物が続く区間又は音声が続く区間と完全に一致していなくてもよい。例えば、第 1 区間又は第 2 区間は、人物が続く区間又は音声が続く区間と数フレーム分の違いがあってもよく、その場合であっても同様の効果を奏する。また、上記実施形態の第 3 区間も同様に、第 1 区間と第 2 区間とを合わせた区間と完全に一致していなくてもよい。

【 0 1 2 8 】

また、上記実施形態では、抽出区間決定部 3 4 が、第 1 区間、及び時間軸上において第 1 区間に少なくとも一部が重なる第 2 区間のそれぞれ全ての区間を合わせた区間を第 3 区間として抽出する例を説明したが、これに限られものではない。例えば、抽出区間決定部 3 4 は、第 1 区間、及び時間軸上において第 1 区間に少なくとも一部が重ならない第 2 区間のそれぞれを第 3 区間として抽出してもよい。具体的には、抽出区間決定部 3 4 は、例えば、第 1 区間と第 2 区間とが互いに時間軸上において少なくとも一部が重なるか否かにかかわらず、人物が検出されたフレーム画像が時系列に続く第 1 区間と、人物の音声が検出された音が時系列に続く第 2 区間とのそれぞれを第 3 区間として抽出してもよい。

10

【 0 1 2 9 】

また、再生部 3 7 は、動画を再生する場合、画像又は音声にフェードイン又はフェードアウト処理を実行してもよい。例えば、再生部 3 7 は、第 3 区間の動画を再生する場合、画像又は音声をフェードインさせて再生を開始し、画像又は音声をフェードアウトさせて再生を終了してもよい。また、再生部 3 7 は、複数の第 3 区間の動画を時系列の順に従って続けて再生する場合、複数の第 3 区間それぞれの画像又は音声をフェードインさせて再生を開始し、フェードアウトさせて再生を終了してもよい。また、再生部 3 7 は、複数の第 3 区間のうちの第 1 の第 3 区間における最後の所定の区間と、時系列の順に従った第 1 の第 3 区間の次の第 2 の第 3 区間における最初の所定の区間とを重ねて再生する場合、その重ねて再生する画像又は音にフェードイン又はフェードアウト処理を実行してもよい。

20

【 0 1 3 0 】

また、上記実施形態では、情報処理装置 1 0 0 (第 1 実施形態)と、撮像装置 5 0 0 (第 2 実施形態)との例をそれぞれ説明したが、これに限られるものではなく、その他の電子機器に本発明の情報処理装置を適用することができる。この電子機器とは、例えば、表示装置、パーソナルコンピュータ、タブレット型端末、ゲーム機等、様々な電子機器であってもよい。

30

【 0 1 3 1 】

なお、上述の処理部 1 0 は専用のハードウェアにより実現されるものであってもよく、また、メモリ及び CPU (中央演算装置)により構成され、上述の処理部 1 0 の機能を実現するためのプログラムをメモリにロードして実行することによりその機能を実現させるものであってもよい。

【 0 1 3 2 】

また、上述の処理部 1 0 の機能を実現するためのプログラムをコンピュータ読み取り可能な記録媒体に記録して、この記録媒体に記録されたプログラムをコンピュータシステムに読み込ませ、実行することにより上述の処理部 1 0 の処理を行ってもよい。なお、ここでいう「コンピュータシステム」とは、OS や周辺機器等のハードウェアを含むものとする。

40

【 0 1 3 3 】

また、「コンピュータシステム」は、WWWシステムを利用している場合であれば、ホームページ提供環境(あるいは表示環境)も含むものとする。

また、「コンピュータ読み取り可能な記録媒体」とは、フレキシブルディスク、光磁気ディスク、ROM、CD-ROM等の可搬媒体、コンピュータシステムに内蔵されるハードディスク等の記憶装置のことをいう。さらに「コンピュータ読み取り可能な記録媒体」とは、インターネット等のネットワークや電話回線等の通信回線を介してプログラムを送

50

信する場合の通信線のように、短時間の間、動的にプログラムを保持するもの、その場合のサーバやクライアントとなるコンピュータシステム内部の揮発性メモリのように、一定時間プログラムを保持しているものも含むものとする。また上記プログラムは、前述した機能の一部を実現するためのものであってもよく、さらに前述した機能をコンピュータシステムにすでに記録されているプログラムとの組み合わせで実現できるものであってもよい。

【0134】

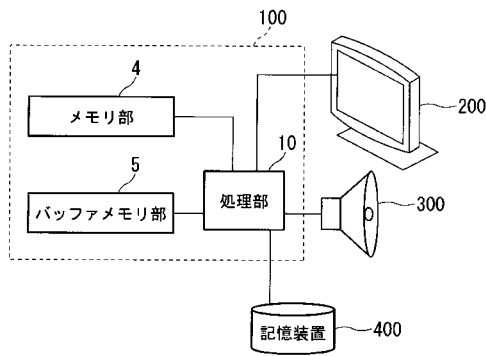
以上、この発明の実施形態を図面を参照して詳述してきたが、具体的な構成はこの実施形態に限られるものではなく、この発明の要旨を逸脱しない範囲の設計等も含まれる。

【符号の説明】

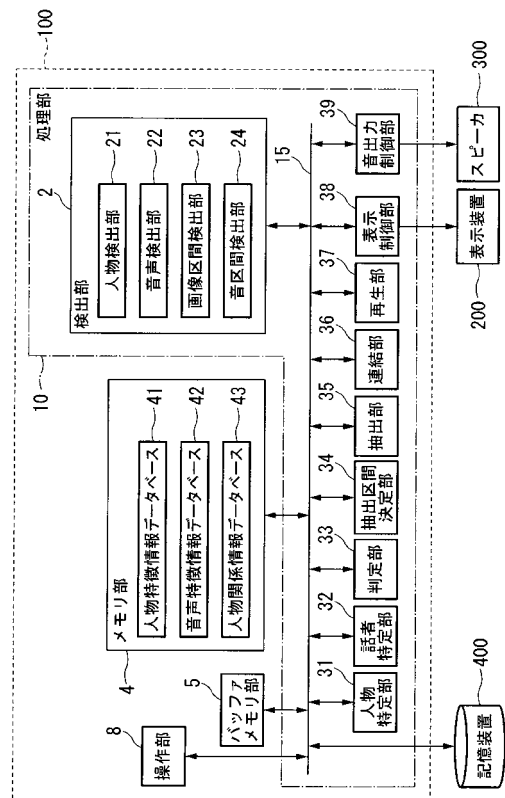
【0135】

2 検出部、4 メモリ部、21 人物検出部、22 音声検出部、23 画像区間検出部、24 音区間検出部、31 人物特定部、32 話者特定部、33 判定部、34 抽出区間決定部、35 抽出部、36 連結部、37 再生部、38 表示制御部、39 音出力制御部、100 情報処理装置、500 撮像処理装置

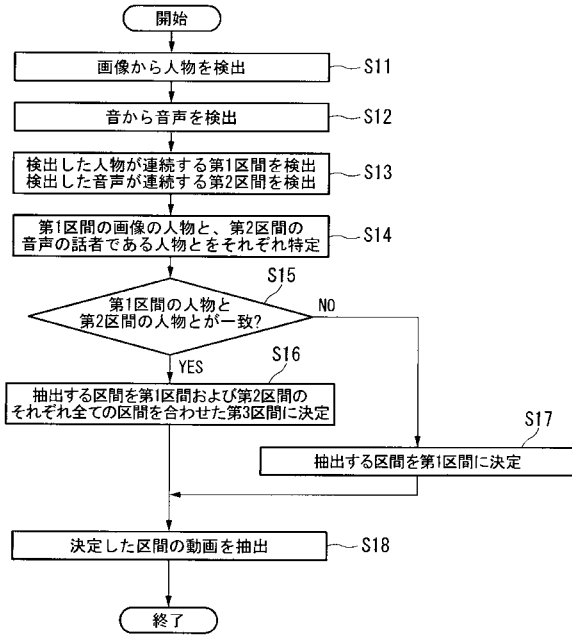
【図1】



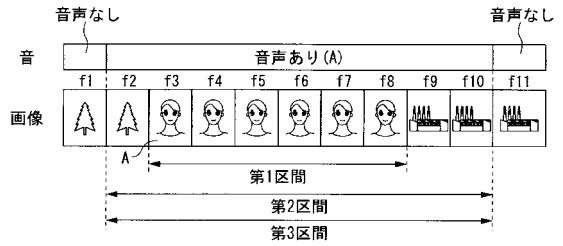
【図2】



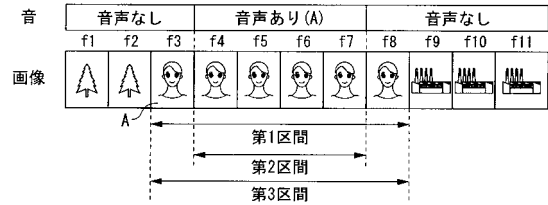
【 図 3 】



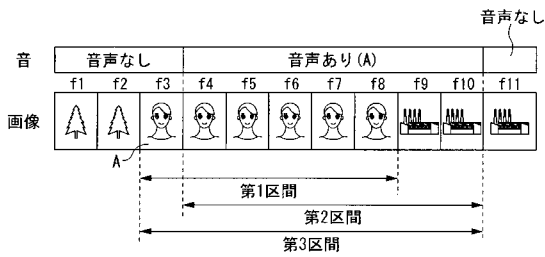
【 図 4 】



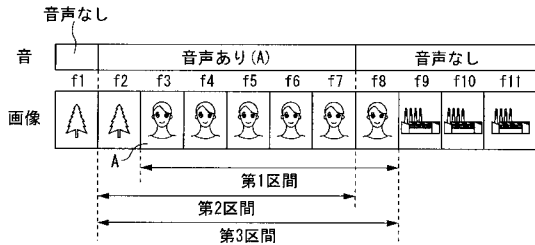
【 図 5 】



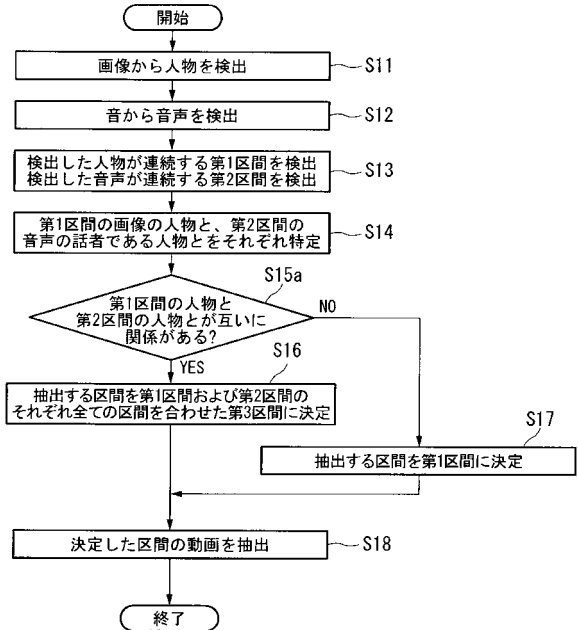
【 図 6 】



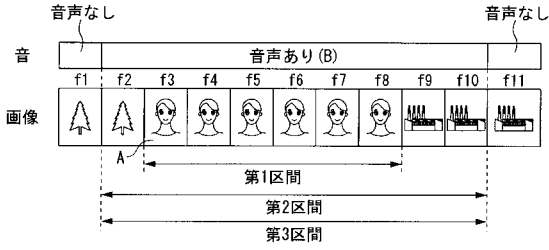
【 図 7 】



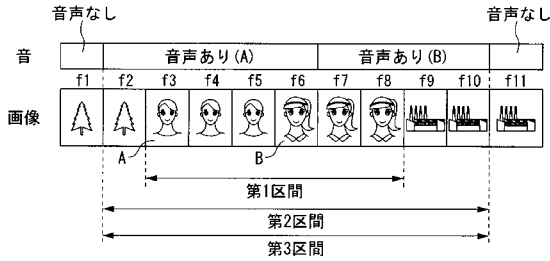
【 図 8 】



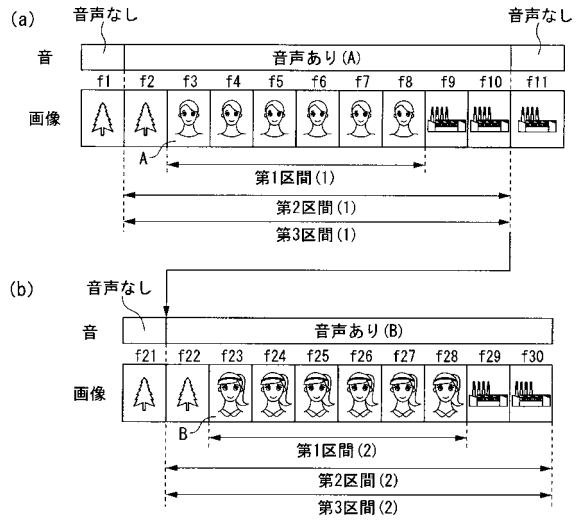
【図9】



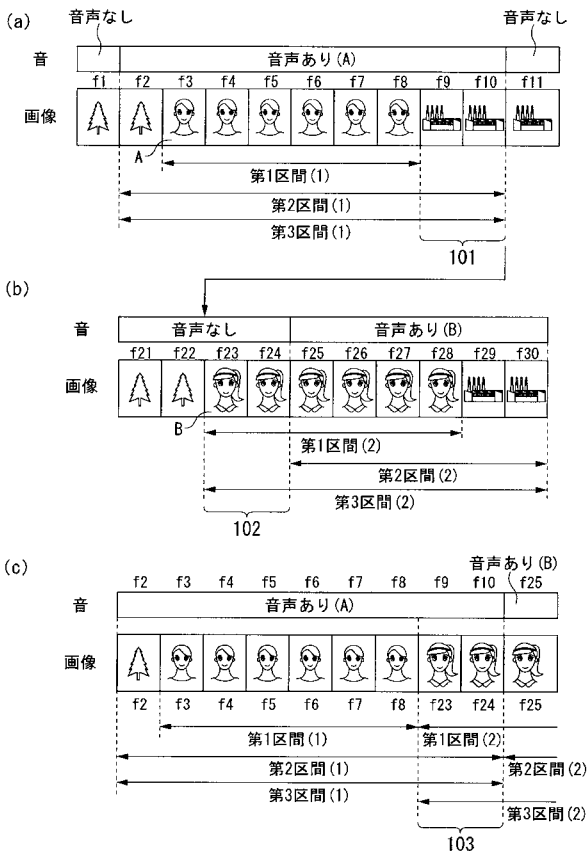
【図10】



【図11】



【図12】



【図13】

