

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第4536940号  
(P4536940)

(45) 発行日 平成22年9月1日(2010.9.1)

(24) 登録日 平成22年6月25日(2010.6.25)

(51) Int. Cl.	F I		
HO4N 5/91 (2006.01)	HO4N	5/91	N
HO4N 5/262 (2006.01)	HO4N	5/262	
HO4N 5/76 (2006.01)	HO4N	5/76	B

請求項の数 16 (全 18 頁)

(21) 出願番号	特願2001-18785 (P2001-18785)	(73) 特許権者	000001007 キヤノン株式会社 東京都大田区下丸子3丁目30番2号
(22) 出願日	平成13年1月26日(2001.1.26)	(74) 代理人	100090273 弁理士 園分 孝悦
(65) 公開番号	特開2002-223412 (P2002-223412A)	(72) 発明者	池田 和世 東京都大田区下丸子3丁目30番2号 キヤノン株式会社内
(43) 公開日	平成14年8月9日(2002.8.9)	審査官	梅岡 信幸
審査請求日	平成20年1月24日(2008.1.24)	(56) 参考文献	特開平10-257436 (JP, A) 特開平11-120353 (JP, A) 特開平11-242542 (JP, A)

最終頁に続く

(54) 【発明の名称】 画像処理装置、画像処理方法、記憶媒体、及びコンピュータプログラム

(57) 【特許請求の範囲】

【請求項1】

動画を処理する画像処理装置であって、  
対象動画を時間的に部分動画に分割する分割手段と、  
上記分割手段で得られた部分動画から第1の代表画像を抽出する抽出手段と、  
上記分割手段で得られた部分動画を上記第1の代表画像の類似性に基づいて統合する統合手段と、

上記統合手段で得られた統合部分動画を構成する複数の部分動画に対応する第1の代表画像から所定数の第1の代表画像を選択する選択手段と、

上記選択手段で選択された所定数の第1の代表画像を合成することにより、上記統合手段で得られた統合部分動画に対する第2の代表画像を生成する生成手段とを備えることを特徴とする画像処理装置。

【請求項2】

上記選択手段は、上記複数の部分動画の再生時間に基づいて、上記第1の代表画像の選択を行うことを特徴とする請求項1記載の画像処理装置。

【請求項3】

上記選択手段は、上記再生時間の長い順に第1の代表画像を選択することを特徴とする請求項2記載の画像処理装置。

【請求項4】

上記選択手段は、類似する第1の代表画像の組に対応した部分画像の再生時間の合計に

10

20

基づいて、上記第 1 の代表画像の選択を行うことを特徴とする請求項 1 記載の画像処理装置。

【請求項 5】

上記選択手段は、上記第 1 の代表画像の中に所定画像の領域が存在するか否かの判定に基づいて、上記第 1 の代表画像の選択を行い、

上記生成手段は、上記選択手段で選択された第 1 の代表画像に基づいて、上記統合部分動画に対する第 2 の代表画像の生成を行うことを特徴とする請求項 1 記載の画像処理装置。

【請求項 6】

上記所定画像の領域は、顔領域を含むことを特徴とする請求項 5 記載の画像処理装置。 10

【請求項 7】

上記選択手段は、異なる上記第 1 の所定画像の領域が存在するか否かの判定に基づいて、上記第 1 の代表画像の選択を行うことを特徴とする請求項 5 記載の画像処理装置。

【請求項 8】

上記生成手段は、上記選択手段で選択された第 1 の代表画像に存在する所定画像の向きに基づいて、上記統合部分動画に対する第 2 の代表画像の生成を行うことを特徴とする請求項 5 記載の画像処理装置。

【請求項 9】

上記生成手段は、上記選択手段で選択された第 1 の代表画像に存在する所定画像の拡大画像に基づいて、上記統合部分動画に対する第 2 の代表画像の生成を行うことを特徴とする請求項 5 記載の画像処理装置。 20

【請求項 10】

動画を処理するための画像処理方法であって、

対象動画を時間的に部分動画に分割する分割ステップと、

上記分割ステップにより得られた部分動画を代表する第 1 の代表画像を抽出する抽出ステップと、

上記分割ステップにより得られた部分動画を上記第 1 の代表画像の類似性に基づいて統合する統合ステップと、

上記統合ステップで得られた統合部分動画を構成する複数の部分動画に対応する第 1 の代表画像から所定数の第 1 の代表画像を選択する選択ステップと、 30

上記選択ステップで選択された所定数の第 1 の代表画像を合成することにより、上記統合ステップにより得られた統合部分動画に対する第 2 の代表画像を生成する生成ステップとを含むことを特徴とする画像処理方法。

【請求項 11】

動画を処理するための画像処理方法であって、

動画を時間的に部分動画に分割する分割ステップと、

上記部分動画を代表する第 1 の代表画像を抽出する抽出ステップと、

上記部分動画を上記第 1 の代表画像の類似性に基づいて統合する統合ステップと、

上記第 1 の代表画像から特定のオブジェクト領域を検出する検出ステップと、

上記検出ステップでの検出結果により、上記特定のオブジェクト領域が存在すると判定された第 1 の代表画像を優先的に、上記統合ステップにより統合される前の各部分動画に対する第 1 の代表画像の中から所定数の第 1 の代表画像を選択する選択ステップと、 40

上記選択ステップにより選択された第 1 の代表画像を合成することにより、上記統合ステップにより統合された後の統合部分動画を代表する第 2 の代表画像を生成する生成ステップとを含むことを特徴とする画像処理方法。

【請求項 12】

上記特定のオブジェクト領域は、顔領域を含み、

上記検出ステップは、上記顔領域の人物をも特定するステップを含み、

上記選択ステップは、異なる人物が存在する第 1 の代表画像を優先的に選択するステップを含むことを特徴とする請求項 11 記載の画像処理方法。 50

## 【請求項 1 3】

上記検出ステップは、上記特定のオブジェクト領域のオブジェクトの向きをも特定するステップを含み、

上記生成ステップは、上記特定のオブジェクトが存在する各第 1 の代表画像の特定のオブジェクトの向きが中心を向いた状態の画像を、上記統合部分動画に対する第 2 の代表画像として生成するステップを含むことを特徴とする請求項 1 1 記載の画像処理方法。

## 【請求項 1 4】

動画を処理するための画像処理方法であって、

動画を時間的に部分動画に分割する分割ステップと、

上記部分動画を代表する第 1 の代表画像を抽出する抽出ステップと、

上記第 1 の代表画像の中の特定のオブジェクト領域を検出する検出ステップと、

上記部分動画を上記第 1 の代表画像の類似性に基づいて統合する統合ステップと、

上記統合ステップにより統合される前の各部分動画に対する第 1 の代表画像の中から、上記統合ステップによる統合部分動画に対する第 1 の代表画像を選択する選択ステップと

、  
上記選択ステップにより選択された第 1 の代表画像を合成することにより、上記統合部分動画を代表する画像を生成する生成ステップとを含み、

上記生成ステップは、上記検出ステップにより、上記第 1 の代表画像に特定のオブジェクト領域が存在する場合、上記特定のオブジェクト領域を拡大した部分画像から上記統合部分動画に対する第 2 の代表画像を生成するステップを含むことを特徴とする画像処理方法。

## 【請求項 1 5】

請求項 1 0 ~ 1 4 の何れか 1 項に記載の画像処理方法の処理ステップをコンピュータに実行させるためのコンピュータプログラムを記録したコンピュータ読取可能な記憶媒体。

## 【請求項 1 6】

請求項 1 0 ~ 1 4 の何れか 1 項に記載の画像処理方法の処理ステップをコンピュータに実行させるためのコンピュータプログラム。

## 【発明の詳細な説明】

## 【0 0 0 1】

## 【発明の属する技術分野】

本発明は、例えば、動画を処理するコンピュータやビデオ録画装置等に用いられる、画像処理装置、画像処理方法、コンピュータ読取可能な記憶媒体、及びコンピュータプログラムに関するものである。

## 【0 0 0 2】

## 【従来の技術】

例えば、動画はその構成上、階層構造を形成することが多い。ここでは、以下の説明において混乱を避けるために、動画の階層構造に関する主なる用語について、次のように定義する。

・撮影を中断せずに一台のカメラで撮影して得られた一連の動画像を「ショット」と言う。

・動画像の構成や内容の連続性等に基づいて複数の「ショット」を統合したもの、あるいは、1つの場면을複数のカメラで撮影して得られた複数の「ショット」を統合したものを「シーン」と言う。また、複数のシーンを統合したものも同じく「シーン」と言う。

・パンやズーム等のカメラの動きを考慮して「ショット」を分割、あるいは、写っている物体の出入りを考慮して「ショット」を分割したものを「サブショット」と言う。

## 【0 0 0 3】

そこで、上記の階層構造を利用して動画の検索や編集等を行う画像処理装置として、例えば、特開平 5 - 3 0 4 6 4 号や特開平 5 - 2 8 2 3 7 9 号等で提案された装置がある。このような画像処理装置では、動画を階層構造として画面表示し、当該表示画面上から動画の検索や編集等を行えるようになされている。

10

20

30

40

50

## 【 0 0 0 4 】

図 1 5 は、上記画像処理装置において、動画を階層構造として表示した画面を示したものである。

上記図 1 5 において、“ 7 0 5 ” は、シーンの代表画像を示し、“ 7 0 4 ” は、当該シーンに含まれるショットの代表画像を示し、“ 7 0 1 ” ~ “ 7 0 3 ” は、当該ショットに含まれるサブショットの代表画像を示す。

## 【 0 0 0 5 】

上記の代表画像としては、サブショットや、ショット及びシーンを代表するフレームを 1 つだけ選び、そのフレーム（代表フレーム）の縮小画像を用いている。

代表フレームの選択方法としては、例えば、ショットやシーンの先頭のフレームを自動的に選択する方法や、ユーザから指定されたフレームを代表フレームとする方法等が提案されている。

10

## 【 0 0 0 6 】

一方、複数のショットを統合してシーンを形成する方法としては、例えば、「繰り返しショットの統合による階層化アイコンを用いたビデオ・インターフェース」（情報処理学会論文誌：V o 1 . 3 9 , N o . 5 , 1 9 9 8 ）等に記載された方法がある。この方法では、ドラマ等で複数の人間が対話するシーンでは各発話者をアップにしたショットが発話毎に切り替わり同一シーン中に類似するショットが複数回出現する、という規則性を利用して、自動的にショットをシーンに統合するようになされている。

20

## 【 0 0 0 7 】

## 【発明が解決しようとする課題】

しかしながら、上述したような従来の画像処理装置及び方法では、動画を階層構造として画面表示するために、シーンやショットの代表フレームを 1 枚だけを選択しても、その代表フレームが、当該シーンやショットの内容を十分に表現しているとは言えなかった。このため、あるシーン或いはショットの内容を把握するためには、それ下位のノードの代表フレームを見なければならなかった。

## 【 0 0 0 8 】

また、複数のショットを自動的に統合して形成されたシーンが、例えば、対話シーンである場合、統合して得られた当該シーンの代表フレームとして、発話者の 1 人のみが存在するフレームが選択される場合があった。このため、この代表フレームからは、対話シーンであるということが把握できない場合があった。

30

## 【 0 0 0 9 】

また、動画中に人間が存在するショットは、人間が存在しないショットに比べて重要なショットであることが一般的であるが、任意のショットの代表フレームを選択すると、人間が存在しないショットの代表フレームが選択される場合があった。これは、シーンの内容を十分に代表していない。

## 【 0 0 1 0 】

また、選択された代表フレームに人間が存在する場合、これを縮小して画面表示することになるため、当該表示画面上で、誰が出現しているのか識別しづらかった。

## 【 0 0 1 1 】

そこで、本発明は、上記の欠点を除去するために成されたもので、動画が階層構造として表示された画面から、各階層の動画（部分動画）の内容を容易に且つ正確に認識することが可能な、画像処理装置、画像処理方法、コンピュータ読取可能な記憶媒体、及びコンピュータプログラムを提供することを目的とする。

40

## 【 0 0 1 2 】

## 【課題を解決するための手段】

本発明の画像処理装置は、動画を処理する画像処理装置であって、対象動画を時間的に部分動画に分割する分割手段と、上記分割手段で得られた部分動画から第 1 の代表画像を抽出する抽出手段と、上記分割手段で得られた部分動画を上記第 1 の代表画像の類似性に基づいて統合する統合手段と、上記統合手段で得られた統合部分動画を構成する複数の部

50

分動画に対応する第1の代表画像から所定数の第1の代表画像を選択する選択手段と、上記選択手段で選択された所定数の第1の代表画像を合成することにより、上記統合手段で得られた統合部分動画に対する第2の代表画像を生成する生成手段とを備えることを特徴とする。

【0023】

本発明の画像処理方法は、動画を処理するための画像処理方法であって、対象動画を時間的に部分動画に分割する分割ステップと、上記分割ステップにより得られた部分動画を代表する第1の代表画像を抽出する抽出ステップと、上記分割ステップにより得られた部分動画を上記第1の代表画像の類似性に基づいて統合する統合ステップと、上記統合ステップで得られた統合部分動画を構成する複数の部分動画に対応する第1の代表画像から所定数の第1の代表画像を選択する選択ステップと、上記選択ステップで選択された所定数の第1の代表画像を合成することにより、上記統合ステップにより得られた統合部分動画に対する第2の代表画像を生成する生成ステップとを含むことを特徴とする。

10

【0030】

また、本発明の画像処理方法の他の特徴とするところは、動画を処理するための画像処理方法であって、動画を時間的に部分動画に分割する分割ステップと、上記部分動画を代表する第1の代表画像を抽出する抽出ステップと、上記部分動画を上記第1の代表画像の類似性に基づいて統合する統合ステップと、上記第1の代表画像から特定のオブジェクト領域を検出する検出ステップと、上記検出ステップでの検出結果により、上記特定のオブジェクト領域が存在すると判定された第1の代表画像を優先的に、上記統合ステップにより統合される前の各部分動画に対する第1の代表画像の中から所定数の第1の代表画像を選択する選択ステップと、上記選択ステップにより選択された第1の代表画像を合成することにより、上記統合ステップにより統合された後の統合部分動画を代表する第2の代表画像を生成する生成ステップとを含むことを特徴とする。

20

【0033】

また、本発明の画像処理方法のその他の特徴とするところは、動画を処理するための画像処理方法であって、動画を時間的に部分動画に分割する分割ステップと、上記部分動画を代表する第1の代表画像を抽出する抽出ステップと、上記第1の代表画像の中の特定のオブジェクト領域を検出する検出ステップと、上記部分動画を上記第1の代表画像の類似性に基づいて統合する統合ステップと、上記統合ステップにより統合される前の各部分動画に対する第1の代表画像の中から、上記統合ステップによる統合部分動画に対する第1の代表画像を選択する選択ステップと、上記選択ステップにより選択された第1の代表画像を合成することにより、上記統合部分動画を代表する画像を生成する生成ステップとを含み、上記生成ステップは、上記検出ステップにより、上記第1の代表画像に特定のオブジェクト領域が存在する場合、上記特定のオブジェクト領域を拡大した部分画像から上記統合部分動画に対する第2の代表画像を生成するステップを含むことを特徴とする。

30

【0035】

本発明の記憶媒体は、上記の何れかに記載の画像処理方法の処理ステップをコンピュータに実行させるためのコンピュータプログラムを記録したことを特徴とする。

【0037】

本発明のコンピュータプログラムは、上記の何れかに記載の画像処理方法の処理ステップをコンピュータに実行させることを特徴とする。

40

【0038】

【発明の実施の形態】

以下、本発明の実施の形態について図面を用いて説明する。

【0039】

本発明は、例えば、図1に示すような画像処理装置100に適用される。

本実施の形態の画像処理装置100は、上記図1に示すように、CPU101、ROM102、RAM103、CD-ROMドライブ104、HDドライブ106、キーボード107、ディスプレイ108、マウス109、及びプリンタ110が、システムバス111

50

を介して互いに通信可能なように接続された構成としている。

【0040】

CPU101は、画像処理装置100全体の動作制御を司るものであり、例えば、ROM102等に予め記憶された処理プログラムを読み出して実行することで、図2に示すような機能を実現する。

すなわち、画像処理装置100は、上記図2に示すように、処理対象動画を時間的に分割する動画分割部151と、動画分割部151で得られた部分動画を代表する画像(代表画像)を抽出する代表画像抽出部153と、動画分割部151で得られた一連の部分動画を意味的にまとまりのある部分動画として統合する部分動画統合152と、代表画像抽出部153で得られた代表画像(部分動画統合152での統合前の部分画像の代表画像)の中  
10  
から人物(顔)が存在する画像を優先的に規定枚数選択する代表画像選択部155と、代表画像選択部155で選択された代表画像に基づいて部分動画統合152での統合後の部分動画(統合部分動画)の代表画像を作成する代表画像作成部154とを備えている。

【0041】

ROM102には、例えば、図3に示すように、CPU101での動作制御に必要な処理プログラム(制御手順プログラム)102a等が格納されている。

【0042】

RAM103は、例えば、図4に示すように、基本I/Oプログラムの格納領域103a及びオペレーションシステムプログラムの格納領域103bと共に、動画検索プログラムの格納領域103c、動画データベース103d、ショット情報の格納領域103e、  
20  
シーン統合情報の格納領域103f、及び代表画像分類情報の格納領域103g等を含んでいる。

【0043】

RAM103の格納領域103cに格納される動画検索プログラムは、例えば、図5に示すように、CD-ROM105に記憶されている。

したがって、図6に示すように、動画検索プログラム105aが記憶されたCD-ROM105が、画像処理装置100のCD-ROMドライブ104にセットされることで、動画検索プログラム105aは、上記図4に示したように、RAM103の格納領域103cへ格納(ロード)されることになる。

【0044】

動画検索プログラム105aがRAM103へ格納されCPU101から実行可能状態となると、これと同時に、RAM103の格納領域103eへのショット情報の格納等が、  
30  
HDドライブ106から行なわれる。また、動画検索プログラム105aの実行で使用されるメモリとしてのRAM103の動画データベース103dの確保や、RAM103のシーン統合情報用の格納領域103f及び代表画像分類情報用の格納領域103g等の確保が行なわれる。

【0045】

図7は、RAM103の格納領域103eに格納されたショット情報を示したものである。  
40

ショット情報は、動画データベース103dに格納されている動画に対するショットの情報を含んでいる。具体的には例えば、ショット情報は、ショットを一意に識別するための  
ショットID、動画中における開始位置を示す開始時間、ショットの時間の長さを表す  
ショット時間、ショットを代表する画像のファイル名、ショット中に出現する人物のID、  
及びショット中の人物の顔の向きを表す顔方向の情報を含んでいる。

【0046】

図8は、RAM103の格納領域103fに格納されたシーン統合情報を示したものである。

シーン統合情報は、類似するショットを統合するための情報を含んでいる。具体的には例えば、シーン統合情報は、対象ショットを識別するためのショットID、対象ショットに  
類似するショットのショットID、当該類似するショットの時間の合計時間、当該類似す  
50

るショットの中で最も長いショットを示す最長ショットID、及び当該類似するショットに出現する人物の情報を含んでいる。

【0047】

図9は、RAM103の格納領域103gに格納された代表画像分類情報を示したものである。

代表画像分類情報は、統合されたショットの代表画像を選択する際に使用され、シーン統合情報を人物別にまとめた情報を含んでいる。具体的には例えば、代表画像分類情報は、対象人物を識別するための人物の情報、対象人物が出現する類似のショットのID、対象人物が出現する時間の合計時間、及び最も長く人物が出現するショットを示す最長ショットIDを含んでいる。

10

【0048】

図10～図13は、画像処理装置100の動作を示したものである。

例えば、CPU101は、図10～図13のフローチャートに従った処理プログラム(制御手順プログラム)をROM102から読み出して実行する。これにより、画像処理装置100は、次のように動作する。

【0049】

<メイン処理>

ステップS101：図10参照

CPU101は、例えば、CR-ROMドライブ104を介して、CD-ROM105に格納された動画検索プログラム等をRAM103へロードすると共に、HDドライブ106から対象動画及びショット情報等をRAM103へロードする。

20

また、CPU101は、RAM103に対して、シーン統合情報及び代表画像分類情報等の格納領域を確保し、必要な初期化処理を実行する。

【0050】

ステップS102～ステップS104：

CPU101は、キーボード107或いはマウス109によるユーザ指示に従って、処理を分岐させる(ステップS102)。

すなわち、CPU101は、ユーザから動画検索の指示がなされた場合にはステップS103の処理を実行し、ユーザから動画登録の指示がなされた場合にはステップS104の処理を実行する。

30

【0051】

<動画検索処理：ステップS103>

CPU101は、RAM103の動画データベース103dに格納された対象動画の中から、ユーザが所望するシーン(ユーザから指定されたシーン)を検索する処理を実行する。

【0052】

具体的には例えば、CPU101は、上記図15に示したような動画像の階層構造の画面を、ディスプレイ108へ表示させる。

これにより、ユーザは、ディスプレイ108の表示画面から、キーボード107或いはマウス109を用いて、所望するシーンの代表画像を検索指示する。

40

【0053】

尚、ここでの動画検索処理は、例えば、特開平5-30464号等に記載された方法や、任意の方法を適用可能である。ただし、特開平5-30464号等に記載された処理が、シーンを代表する画像として、シーン中のフレームを直接使用している構成であるのに対して、本実施の形態では、後述するステップS104で作成されるシーンの代表画像を使用する。

【0054】

<動画登録処理：ステップS104>

動画登録処理は、CPU101により実現される上記図2に示した構成により、指定された動画像をRAM103の動画データベース103dへ登録する処理である。

50

図 11 は、当該動画登録処理を示したものである。

【 0055】

ステップ S201 :

動画分割部 151 は、対象動画（指定動画）を先頭から解析して、シーンチェンジ（ショットの切り替わり）を検出し、その検出結果の情報をショット情報として、RAM103 の格納領域 103e（上記図 7 参照）に格納する。

シーンチェンジの検出方法としては、例えば、特開平 5 - 30464 号等に記載されているような、フレーム間の変化量の大きさから、ショットとショットの境界を検出する方法を適用可能である。

【 0056】

このとき RAM103 の格納領域 103e に格納されるショット情報は、ショット ID、開始時間、及びショット時間のみの情報である。ショット ID としては、使用されているショット ID の最大値に “ 1 ” を加えた値を用いる。開始時間及びショット時間については、シーンチェンジが検出されるフレームの、動画先頭からのフレーム番号から自動的に求めることが可能である。

【 0057】

ステップ S201 において、シーンチェンジ（ショットの切り替わり）が検出された場合、或いは処理が対象動画の末尾に到達した場合に、次のステップ S202 へ進む。

【 0058】

ステップ S202 :

代表画像抽出部 153 は、ステップ S201 で動画分割部 151 により検出されたショットに対するキーフレーム（部分動画を代表するフレーム）を抽出する。

キーフレームの抽出方法としては、例えば、ショットの先頭や中心、或いは末尾等、ショットの位置を指定することによって、ショットのキーフレームを決定する方法を適用可能である。

【 0059】

代表画像抽出部 153 は、キーフレームを抽出後、そのキーフレームの画像情報をファイルとして保持するために、そのファイル名を、ショット情報の代表画像ファイル名として、RAM103 の格納領域 103e に格納する。

ここでのファイル名としては、上記図 7 に示されるように、例えば、ショット ID = “ 100 ” のショットに関しては、そのショット ID を利用して、ファイル名 = “ 100 . bmp ” とすることで、ファイル名の重複を避けることができる。

【 0060】

ステップ S203 :

対象画像の全てに対して、ステップ S201 及びステップ S202 の処理を実行し終えるまで、ステップ S201 及びステップ S202 の処理を繰り返し実行する。そして、対象画像の全てに対して、ステップ S201 及びステップ S202 の処理を実行し終えた場合に、すなわち動画の末尾まで処理が到達した場合に、次のステップ S204 へ進む。

【 0061】

ステップ S204 :

部分動画統合部 152 は、ステップ S202 において代表画像抽出部 153 により抽出されたキーフレームの類似性に基づいて、複数のショットをまとめて 1 つのシーンとして統合し、その統合結果をシーン統合情報として、RAM103 の格納領域 103f（上記図 8 参照）に格納する。

ここでのシーン統合処理方法については、例えば、「繰り返しショットの統合による階層化アイコンを用いたビデオ・インターフェース」（情報処理学会論文誌：V o 1 . 3 9 , N o . 5 , 1 9 9 8 ）等に記載された方法を適用可能である。

【 0062】

シーン統合情報は、本ステップ S204 で実行した統合に関する一時的な情報であり、本ステップ S204 の実行前に、最初に必ず初期化されるようになされている。

10

20

30

40

50

例えば、本ステップS 2 0 4の処理実行後、上記図7に示したショット情報に対して、上記図8に示したようなシーン統合情報が得られる。ただし、当該シーン統合情報として格納される情報としては、類似ID、ショットID、合計時間、及び最長ショットIDのみの情報であり、人物の情報は含まれない。

**【0063】**

ステップS 2 0 5 :

代表画像選択部155は、詳細は後述するが、ステップS 2 0 4で部分動画統合部152により得られたシーンを代表するフレーム画像を、ステップS 2 0 2で代表画像抽出部153により得られたショットのキーフレーム(統合前のショットのキーフレーム)の中から2枚選択する。

10

**【0064】**

ステップS 2 0 6 :

代表画像作成部154は、詳細は後述するが、ステップS 2 0 5で代表画像選択部155により得られた2枚のキーフレームに基づいて、ステップS 2 0 4で部分動画統合152により得られたシーンの代表フレーム(代表画像)を作成する。

**【0065】**

ステップS 2 0 7 :

対象画像の全てに対して、ステップS 2 0 4～ステップS 2 0 6の処理を実行し終えるまで、ステップS 2 0 4～ステップS 2 0 6の処理を繰り返し実行する。そして、対象画像の全てに対して、ステップS 2 0 4～ステップS 2 0 6の処理を実行し終えた場合に、すなわち動画の末尾まで処理が到達した場合に、本処理終了となる。

20

**【0066】**

<代表画像選択処理：ステップS 2 0 5 >

図12は、代表画像選択部155による代表画像選択処理を示したものである。

**【0067】**

ステップS 3 0 1 :

代表画像選択部155は、RAM 1 0 3の格納領域103fに格納されたシーン統合情報に含まれるショットIDにより示されるショットのキーフレームに対して、当該フレームに存在する人物の顔領域を推定し、その人物と顔の方向を特定し、その結果をショット情報としてRAM 1 0 3の格納領域103e(上記図7参照)に格納すると共に、RAM 1 0 3の格納領域103f(上記図8参照)のシーン統合情報を更新する。

30

人物の顔領域を推定して当該人物を特定する方法としては、例えば、特開平9-251534号等に記載されている方法を適用可能である。また、顔の方向を特定する方法としては、例えば、特開平9-251534号等に記載されているような、人物の顔を上下左右等の様々な向きから撮った画像を辞書画像として用意しておく方法が適用可能である。

**【0068】**

ここで、上記図7に示したRAM 1 0 3の格納領域103eのショット情報において、顔領域の情報が示されていないが、当該ショット情報のショットIDに対しては、キーフレームに存在する人物の顔領域を矩形で表現したときの当該矩形を示す2点の座標の情報が対応付けられている。

40

また、上記図7に示したショット情報の格納領域103e、及び上記図8に示したシーン統合情報の格納領域103fが、本ステップS 3 0 1の処理後の状態である。上記図8において、例えば、類似ID="5"に対する人物の情報の欄が空欄であるのは、そのショットに人物が存在しないことを意味する。

**【0069】**

ステップS 3 0 2 :

代表画像選択部155は、代表フレーム(代表画像)を選択するために、RAM 1 0 3の格納領域103e及び格納領域103fに格納されたショット情報及びシーン統合情報に基づいて、RAM 1 0 3の格納領域103gに格納する代表画像分類情報を生成する。

**【0070】**

50

具体的には、代表画像選択部 155 は、RAM 103 の格納領域 103 f に格納されたシーン統合情報の先頭から順番に、対象類似 ID の情報を取得し、その人物に対する代表画像分類情報を生成する。

すなわち、対象人物が代表画像分類情報に登録されていない場合は、RAM 103 の格納領域 103 g (上記図 9 参照) において、新たに、代表画像分類情報に対してエントリを生成し、これに対応する類似 ID を格納し、当該類似 ID に対応させて合計時間及び最長ショット ID を格納する。

一方、対象人物が代表画像分類情報に登録されていれば、RAM 103 の格納領域 103 g において、対象人物に対して、類似 ID を追加し、合計時間を加算し、最長ショット ID の大小を比較して、必要に応じて最長ショット ID の更新を行う。

また、対象類似 ID の情報に対して人物の情報が含まれていない場合、RAM 103 の格納領域 103 g において、人物の情報欄は空欄にし、類似 ID 及び合計時間最長ショット ID をそのまま格納する。

#### 【0071】

したがって、上記図 7 に示した RAM 103 の格納領域 103 e に格納されたショット情報、及び上記図 8 に示した RAM 103 の格納領域 103 f に格納されたシーン統合情報からは、上記図 9 に示したような代表画像分類情報が生成される。

#### 【0072】

上記図 9 では、その一例として、それぞれのシーンに 1 人しか出現していないものとしているが、2 人同時に出現している場合であっても、代表画像分類情報の人物の欄には必ず 1 人の情報しか格納しないように構成しているため、1 つの類似シーンに対して出現する人数分、同様の処理を実行すればよい。ただし、この場合、最長ショット ID として、1 人のみのショットが選ばれるようにする。

#### 【0073】

ステップ S 303 :

代表画像選択部 155 は、上記図 9 の代表画像分類情報に基づいて、人物が出現する代表画像を選択する。

具体的には、代表画像選択部 155 は、当該代表画像分類情報において、人物の情報が存在する行で、且つ、合計時間が長い行から順番に、各行のソーティングを行う。これにより、当該ソーティング後の代表画像分類情報の先頭の行から選択することで、人物を優先的に、且つ、同一人物が重複して選ばれないように、代表画像を選ぶことができる。また、長時間出現する人物を優先的に代表画像として選ぶこともできる。

#### 【0074】

ステップ S 304 :

代表画像選択部 155 は、人物に基づいた代表画像の選択が終了したか否かを判別する。具体的には例えば、代表画像選択部 155 は、上記図 9 の代表画像分類情報において、人物のエントリが規定枚数(ここでは“2”)以上あれば、代表画像の選択が終わったものと判別して本処理を終了し、そうでなければ次のステップ S 305 へ進む。

#### 【0075】

ステップ S 305 :

代表画像選択部 155 は、上記図 9 の代表画像分類情報に基づいて、人物が出現しないショットから代表画像を選択する。

具体的には例えば、代表画像選択部 155 は、上記図 9 の代表画像分類情報において、人物の情報が存在しない行を合計時間が長いものから順に並ぶようにソーティングを行う。このとき、人物の情報欄に人物の ID が格納されている行の位置は変更しないようにする。これにより、代表画像分類情報の先頭から規定枚数(ここでは“2”)分の行の最長ショット ID に対応した代表画像を選ぶことにより、人物が出現しないショットに関しては、ショットの時間が長いショットから代表画像が選ばれることになる。

本ステップ S 305 の処理の終了後、本処理終了となる。

#### 【0076】

< 代表画像作成処理：ステップ S 2 0 6 >

図 1 3 は、代表画像作成部 1 5 4 による代表画像作成処理を示したものである。

【 0 0 7 7 】

ステップ S 4 0 1 :

代表画像作成部 1 5 4 は、上記図 9 の代表画像分類情報の先頭から順番に一行ずつの情報を取り出し、これに対応する代表画像を取り出す。ここでは、最長ショット ID に対応した上記図 7 に示したようなショット情報の代表画像ファイル名に示される代表画像が取り出されることになる。これにより、人物が存在する代表画像が人物が存在しない代表画像よりも優先され、それぞれの画像中で、ショットの合計時間が長い画像の方が優先され、人物が重複することなく代表画像が取り出せることになる。

10

【 0 0 7 8 】

ステップ S 4 0 2 :

代表画像作成部 1 5 4 は、ステップ S 4 0 1 で取り出した代表画像に人物が存在しているか否かを判別する。

具体的には例えば、代表画像作成部 1 5 4 は、上記図 7 のショット情報の対象となる情報に人物の情報が含まれている場合（当該情報欄が空欄でない場合）、人物が存在するものとして、次のステップ S 4 0 3 へ進み、そうでない場合には、後述するステップ S 4 0 5 へ進む。

【 0 0 7 9 】

ステップ S 4 0 3 :

代表画像作成部 1 5 4 は、ステップ S 4 0 1 で取り出した代表画像を、統合されたシーンの代表画像の中にはめ込む位置を決定する。

20

【 0 0 8 0 】

具体的には例えば、本実施の形態では、統合前のショットの代表画像 2 枚を用いて、統合後のシーンの代表画像を作成する構成としているので、はめ込み位置としては、左側と右側の 2 箇所としており、スナップ S 4 0 1 で取り出した代表画像の人物の顔の向きによって、はめ込み位置を決定する。すなわち、最長ショット ID に対応したショット情報中の顔方向が右であれば、シーン統合後の代表画像の左側をはめ込み位置として決定し、顔方向が左であれば、シーン統合後の代表画像の右側をはめ込み位置として決定する。また、既にはめ込まれている場合には、はめ込まれていないほうをはめ込み位置として決定する。

30

この結果、例えば、2 枚の代表画像が、図 1 4 ( a ) に示されるような人物 A が存在する画像、及び同図 ( b ) に示されるような人物 B が存在する画像である場合、同図 ( c ) に示すように、人物 A が左向きであることにより人物 A の代表画像は右側に、人物 B が右向きであることにより人物 B の代表画像を左側に、それぞれのめ込み位置を決定する。

【 0 0 8 1 】

ステップ S 4 0 4 :

代表画像作成部 1 5 4 は、ステップ S 4 0 3 で決定したはめ込み位置に基づいて、ステップ S 4 0 1 で取り出した代表画像を、シーン統合後の代表画像にはめ込む。

このとき、上記図 1 4 ( c ) に示すように、顔領域の部分を拡大してはめ込みを行う。例えば、ショットの代表画像中の顔領域は、既に矩形として求まっているので、はめ込み先のはめ込み領域の形と顔領域の形（矩形）を勘案し、顔領域ができるだけ大きくなり、顔領域がはめ込み領域に収まる程度に拡大、或いは縮小してはめ込みを行う。

40

その後、後述するステップ S 4 0 6 へ進む。

【 0 0 8 2 】

ステップ S 4 0 5 :

ステップ S 4 0 2 の判別の結果、人物が存在しない場合、すなわち代表画像が人物以外のショットの画像である場合、代表画像作成部 1 5 4 は、当該代表画像をそのままシーン統合後の代表画像にはめ込む。このときのはめ込み位置は、空いているはめ込み位置から左から順番に選ぶ等、任意の規則で選択するようにしてもよい。また、はめ込み領域の形に

50

合わせて、必要に応じて縮小を行って、はめ込みを行う。  
その後、次のステップS 4 0 6へ進む。

【0083】

ステップS 4 0 6：

代表画像作成部154は、上記図9の代表画像分類情報の全ての情報に対して、規定枚数以上のショットの代表画像を取り出し終えている場合には本処理終了とし、そうでない場合には再びステップS 4 0 1へと戻る。

【0084】

したがって、特に、ステップS 4 0 3及びステップS 4 0 4の処理により、シーン統合後の代表画像として、上記図14(c)に示したような画像が得られることになる。すなわち、対話シーンでは、シーン統合後の代表画像が、人物がお互いに向き合った状態で、また、顔領域の部分がはめ込み領域のサイズに合わせてはめ込まれた状態の画像となるので、対話シーンであるということが容易に且つ明確に認識できる。

10

【0085】

尚、本発明は、本実施の形態に限られることはなく、以下のような形態をも含まれる。

【0086】

(1)本実施の形態では、CPU101が、外部記憶装置としてのCD-ROM105から、上述したような画像処理装置100の機能を実施するための処理プログラム(動画検索プログラム)を直接RAM103にロードして実行するように構成したが、これに限られることはなく、CD-ROM105から当該処理プログラムを一旦HDドライブ106

20

に格納(インストール)しておき、当該処理プログラムを動作させる時点で、HDドライブ106からRAM103にロードするようにしてもよい。  
また、当該処理プログラムを記録する媒体としては、CD-ROM105に限られることはなく、例えば、FD(フロッピーディスク)やICメモリカード等であってもよい。  
また、当該処理プログラムを、ROM102に記録しておき、これをメモリマップの一部となるように構成し、直接CPU101で実行するように構成してもよい。

【0087】

(2)上記図11に示した動画登録の処理において、ステップS 2 0 4では、ショットをまとめて1つのシーンとして統合するために、ショットの代表フレームの類似性を利用して自動的に行うように構成したが、これに限られることはなく、例えば、類似するショットを人手で指定することで統合を行うようにしてもよい。

30

【0088】

(3)上記図11に示した動画登録の処理では、ステップS 2 0 1でシーンチェンジを検出した後に、対応するショットのキーフレームを抽出して代表画像とするように構成したが、これに限られることはなく、例えば、シーンチェンジを検出した後に、対応するショットを、ズームやパン等のカメラの動き等に基づいて、サブショットに分割し、その分割サブショットに対してキーフレームを検出し、ステップS 2 0 4～ステップS 2 0 6と同様の処理を実行することで、サブショットのキーフレームを合成してショットの代表画像を作成するようにしてもよい。

【0089】

(4)上記図11に示した動画登録の処理において、ステップS 2 0 7では、全てのショットに対してシーンの統合が終了した場合に本処理終了とするように構成したが、これに限られることはなく、例えば、さらに、階層を重ねてシーンとシーンの統合を行い、当該統合後のシーンに対して、ステップS 2 0 5及びステップS 2 0 6と同様の処理を実行して代表画像を作成するようにしてもよい。

40

【0090】

(5)本実施の形態では、シーンの代表画像を、2つのショットの代表フレームとしたが、これに限られることはなく、例えば、3つ以上のショットの代表フレームとしてもよい。この場合、予め、使用する代表フレームの個数に応じて、代表フレームをはめ込む領域を決定する。例えば、4つの代表フレームを使用する場合は、代表画像の上下左右の4分

50

割した領域を用意しておく。

【0091】

(6) 本実施の形態では、上記図11に示したステップS205の処理において、代表画像選択部155が顔領域の存在を判定するように構成したが、これに限られることはなく、例えば、同図に示したステップS202の処理において、代表画像抽出部153が顔領域の存在を判定するようにしてもよい。これにより、より適切なキーフレームを抽出することができる。

【0092】

(7) 本発明の目的は、本実施の形態のホスト及び端末の機能を実現するソフトウェアのプログラムコードを記憶した記憶媒体を、システム或いは装置に供給し、そのシステム或いは装置のコンピュータ(又はCPUやMPU)が記憶媒体に格納されたプログラムコードを読みだして実行することによっても、達成されることは言うまでもない。

この場合、記憶媒体から読み出されたプログラムコード自体が本実施の形態の機能を実現することとなり、そのプログラムコードを記憶した記憶媒体は本発明を構成することとなる。

プログラムコードを供給するための記憶媒体としては、ROM、フロッピーディスク、ハードディスク、光ディスク、光磁気ディスク、CD-ROM、CD-R、磁気テープ、不揮発性のメモリカード等を用いることができる。

また、コンピュータが読みだしたプログラムコードを実行することにより、本実施の形態の機能が実現されるだけでなく、そのプログラムコードの指示に基づき、コンピュータ上で稼動しているOS等が実際の処理の一部又は全部を行い、その処理によって本実施の形態の機能が実現される場合も含まれることは言うまでもない。

さらに、記憶媒体から読み出されたプログラムコードが、コンピュータに挿入された拡張機能ボードやコンピュータに接続された機能拡張ユニットに備わるメモリに書き込まれた後、そのプログラムコードの指示に基づき、その機能拡張ボードや機能拡張ユニットに備わるCPUなどが実際の処理の一部又は全部を行い、その処理によって本実施の形態の機能が実現される場合も含まれることは言うまでもない。

【0093】

(8) 本実施の形態では、人物(顔)により画像の選択・分類・登録処理を行なっているが、人物(顔)に限定されるものではない、様々なオブジェクト(例えば、車、動物等)に対応できることは言うまでもない。

【0094】

【発明の効果】

以上説明したように本発明では、統合部分動画を構成する複数の部分動画(統合前の部分動画)に対応する代表画像群から所定数の代表画像を選択し、上記選択した所定数の各代表画像を合成することにより、統合部分動画の代表画像を生成するように構成したので、当該代表画像により、長いシーンのほうが短いシーンよりも重要なシーンであると言われる当該シーンを、明確に把握することができ、統合部分動画の内容を十分に表現できることが可能となる。

【0095】

また、統合部分動画の代表画像の生成に使用する代表画像を選択する際に、所定画像の領域(顔領域等)が存在するものを優先的に代表画像として選択するように構成すれば、重要なショットを用いた代表画像を生成することができる。

このとき、さらに、異なる所定画像のものを代表画像として選択するように構成すれば、例えば、同一人物の顔領域が存在する代表画像が複数選択されてしまうことを防ぐことができるため、できるだけ多くの登場人物が含まれる統合部分動画の代表画像を作成することができ、統合部分動画の内容を十分に表すことができる。

また、所定画像の向き(登場人物の顔の向き等)を考慮して、統合部分動画の作成を行うように構成した場合、対話シーン等の内容をより明確に表現することができる。

また、所定画像を拡大して統合部分動画の代表画像を作成するように構成した場合、統合

10

20

30

40

50

前の部分動画の代表画像が縮小画像であっても、どのような登場人物が存在するか等、統合部分動画の内容を、より明確に把握することができる。

【0096】

よって、本発明は、動画を階層構造として画面表示するための装置或いはシステムに対して非常に有効であり、当該階層構造の各階層の動画（部分動画）の内容を容易に且つ正確に認識することができるようになる。

【図面の簡単な説明】

【図1】本発明を適用した画像処理装置の構成を示すブロック図である。

【図2】上記画像処理装置の機能的構成を示すブロック図である。

【図3】上記画像処理装置のROMを説明するための図である。

10

【図4】上記画像処理装置のRAMを説明するための図である。

【図5】上記画像処理装置の動作制御のための処理プログラム等が記憶されたCD-ROMを説明するための図である。

【図6】上記CD-ROMから処理プログラムが上記画像処理装置へ供給されることを説明するための図である。

【図7】上記RAM上のショット情報を説明するための図である。

【図8】上記RAM上のシーン統合情報を説明するための図である。

【図9】上記RAM上の代表画像分類情報を説明するための図である。

【図10】上記画像処理装置のメイン動作を説明するためのフローチャートである。

【図11】上記メイン動作の動画登録処理を説明するためのフローチャートである。

20

【図12】上記動画登録処理の代表画像選択処理を説明するためのフローチャートである。

【図13】上記動画登録処理の代表画像作成処理を説明するためのフローチャートである。

【図14】上記代表画像作成処理で得られる統合後のシーンの代表画像を説明するための図である。

【図15】画面表示された動画の階層構造を説明するための図である。

【符号の説明】

100 画像処理装置

101 CPU

30

102 ROM

103 RAM

104 CD-ROMドライブ

105 CD-ROM

106 HDドライブ

107 キーボード

108 ディスプレイ

109 マウス

110 プリンタ

111 システムバス

40

151 画像分割部

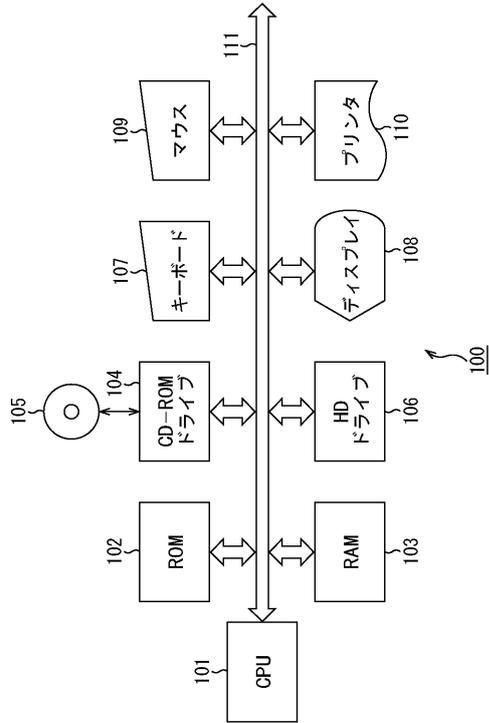
152 部分動画統合部

153 代表画像抽出部

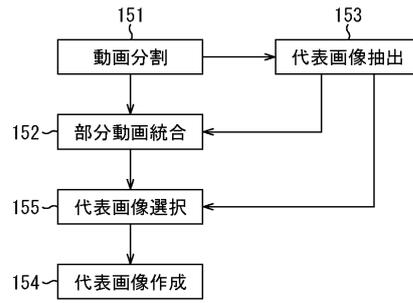
154 代表画像作成部

155 代表画像選択部

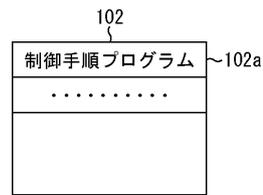
【図1】



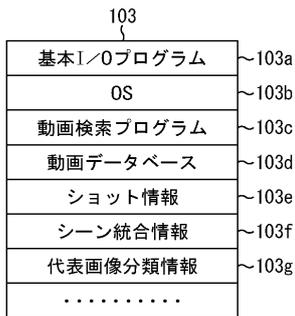
【図2】



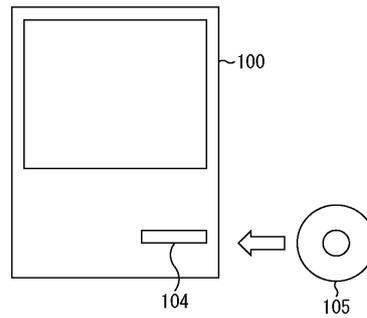
【図3】



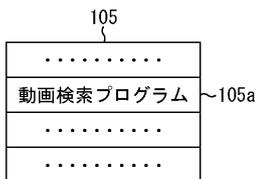
【図4】



【図6】



【図5】



【図 7】

103e

ショットID	開始時間	ショット時間	代表画像ファイル名	人物ID	顔方向
100	200.00	2.00	100.bmp	A	左
101	202.00	3.00	101.bmp	B	右
102	205.00	4.00	102.bmp	C	右
103	209.00	4.00	103.bmp	A	左
104	213.00	5.00	104.bmp	B	右
105	218.00	2.00	105.bmp	A	左
106	220.00	2.00	106.bmp		左
107	222.00	6.00	107.bmp	B	右
108	228.00	6.00	108.bmp	A	左
109	234.00	2.00	109.bmp	A	左

【図 8】

103f

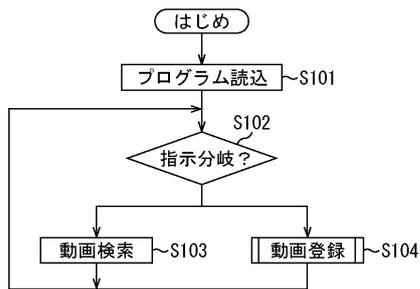
類似ID	ショットID	合計時間	最長ショットID	人物
1	100, 103, 108	12.00	108	A
2	101, 104, 107	14.00	107	B
3	102	4.00	102	C
4	105, 109	4.00	105	A
5	106	2.00	106	

【図 9】

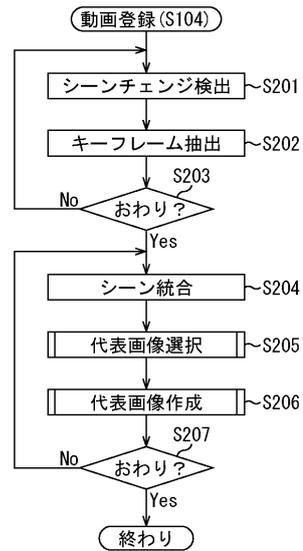
103g

人物	類似ID	合計時間	最長ショットID
A	1, 4	16.00	108
B	2	14.00	107
C	3	4.00	102
	5	2.00	106

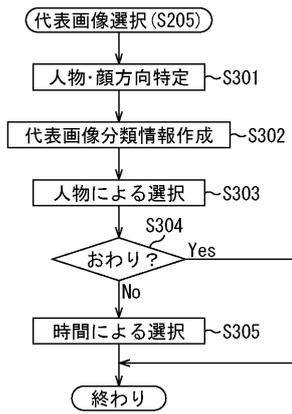
【図 10】



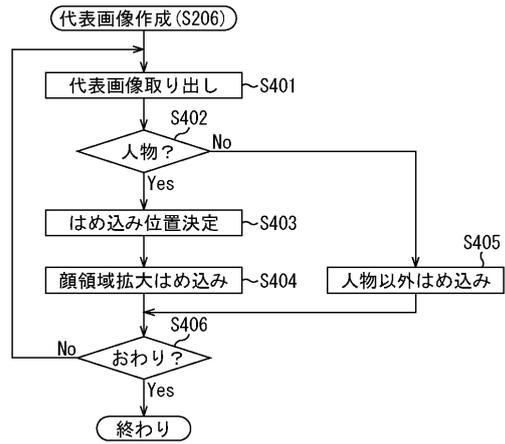
【図 11】



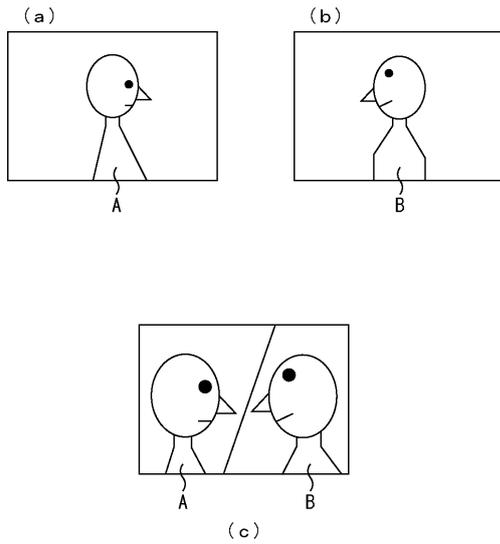
【図 1 2】



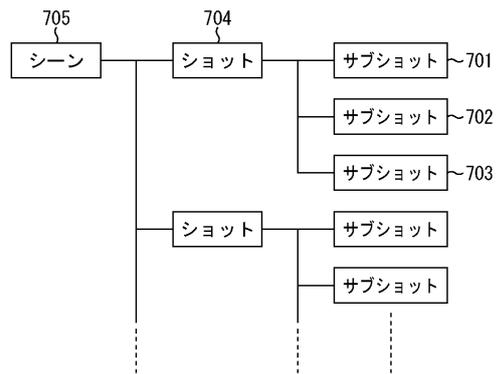
【図 1 3】



【図 1 4】



【図 1 5】



---

フロントページの続き

(58)調査した分野(Int.Cl. , D B名)

H04N 5/76 - 5/956

H04N 5/222- 5/257

H04N 5/262- 5/28

H04N 7/14 - 7/173

G11B 20/10 -20/16

G11B 27/00 -27/34