(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号
特許第5027400号
(P5027400)
(45)発行日　平成24年9月19日(2012.9.19)　(24)登録日　平成24年6月29日(2012.6.29)

(51)Int.Cl.　F I
H04N 5/76 (2006.01)　H04N 5/76 B
G06T 7/00 (2006.01)　G06T 7/00 300F

請求項の数 16　(全 16 頁)

(21)出願番号　特願2005-305799 (P2005-305799)
(22)出願日　平成17年10月20日 (2005.10.20)
(65)公開番号　特開2006-129480 (P2006-129480A)
(43)公開日　平成18年5月18日 (2006.5.18)
審査請求日　平成20年10月20日 (2008.10.20)
(31)優先権主張番号　10/978,172
(32)優先日　平成16年10月30日 (2004.10.30)
(33)優先権主張国　米国 (US)

(73)特許権者　500046438
マイクロソフト　コーポレーション
アメリカ合衆国　ワシントン州　９８０５２－６３９９　レッドモンド　ワン　マイクロソフト　ウェイ
(74)代理人　100140109
弁理士　小野　新次郎
(74)代理人　100089705
弁理士　社本　一夫
(74)代理人　100075270
弁理士　小林　泰
(74)代理人　100080137
弁理士　千葉　昭男
(74)代理人　100096013
弁理士　富田　博行

最終頁に続く

(54)【発明の名称】録画された会議のタイムラインに使用するための自動顔領域抽出

(57)【特許請求の範囲】
【請求項１】
ビデオ・サンプル中の二人以上の顔画像を検出するステップと、
前記ビデオ・サンプルに対応するオーディオ・サンプル中の二人以上の話し手の位置を検出するステップと、
前記二人以上の話し手のうちの主要な話し手を検出するステップと、
前記オーディオ・サンプルを、時間的持続性及び話し手の位置によりクラスタ化して話し手タイムラインを作成し、タイムライン・データベースに格納するステップと、
話し手位置毎に少なくとも１つの顔画像を顔データベースに格納するステップと、
前記顔データベースに格納された前記少なくとも１つの顔画像の中から、話し手位置毎に１つの顔画像を選択するステップと、
二人以上の話し手をその中に含む前記オーディオ／ビデオ（Ａ／Ｖ）サンプルを表示するステップと、
前記主要な話し手の顔画像、及び、前記話し手タイムラインに隣接して表示された前記選択された顔画像を別個に表示するステップと、
を備えることを特徴とする方法。
【請求項２】
前記二人以上の顔画像を検出するステップは、フェイス・トラッキングを使用して前記二人以上の顔画像を検出するステップをさらに備えることを特徴とする請求項１に記載の方法。

【請求項３】
前記二人以上の話し手の位置を検出するステップは、音源位置測定を使用して前記二人以上の話し手の位置を検出するステップをさらに備えることを特徴とする請求項１に記載の方法。
【請求項４】
前記顔データベースに格納された前記少なくとも１つの顔画像の中から、話し手位置毎に１つの顔画像を選択するステップは、最良の顔画像を選択するステップを含むことを特徴とする請求項１に記載の方法。
【請求項５】
前記最良の顔画像を選択するステップは、顔の最も正面からの撮影像を含む顔画像を前記最良の顔画像として選択するステップを含むことを特徴とする請求項４に記載の方法。
【請求項６】
前記最良の顔画像を選択するステップは、最少の動きを示す顔画像を前記最良の顔画像として選択するステップを含むことを特徴とする請求項４に記載の方法。
【請求項７】
前記最良の顔画像を選択するステップは、最大の対称性を示す顔画像を前記最良の顔画像として選択するステップを含むことを特徴とする請求項４に記載の方法。
【請求項８】
前記話し手の位置は、ビデオ・サンプルの座標で識別される話し手バウンディング・ボックスによって表されることを特徴とする請求項１に記載の方法。
【請求項９】
前記話し手の位置は、ビデオ・サンプルにおいて方位角および仰角によって識別される話し手の顔の角度によって示されることを特徴とする請求項１に記載の方法。
【請求項１０】
前記主要な話し手は、前記タイムラインに基づいて判断されることを特徴とする請求項１に記載の方法。
【請求項１１】
コンピュータに、請求項１～１０の何れか１項に記載されたステップを実行させるためのプログラム。
【請求項１２】
コンピュータに、請求項１～１０の何れか１項に記載されたステップを実行させるためのプログラムを記録したコンピュータ読み取り可能な記録媒体。
【請求項１３】
ビデオ・サンプル中の二人以上の顔画像を検出する手段と、
前記ビデオ・サンプルに対応するオーディオ・サンプル中の二人以上の話し手の位置を検出する手段と、
前記オーディオ・サンプルを、時間的持続及び話し手の位置によりクラスタ化して話し手タイムラインを作成し、タイムライン・データベースに格納する手段と、
話し手位置毎に少なくとも１つの顔画像を顔データベースに格納する手段と、
前記顔データベースに格納された前記少なくとも１つの顔画像の中から、話し手位置毎に１つの顔画像を選択する手段と、
前記タイムライン・データベースに格納されたタイムラインに基づいて、前記二人以上の話し手のうちの主要な話し手を検出する手段と、
二人以上の話し手をその中に含む前記オーディオ／ビデオ（Ａ／Ｖ）サンプルを表示する手段と、
前記主要な話し手の顔画像、及び、前記話し手タイムラインに隣接して表示された前記選択された顔画像を別個に表示する手段と、
を備えることを特徴とするシステム。
【請求項１４】
前記オーディオ・サンプル中の二人以上の話し手の位置を検出する手段は、音源ローカ

ライザを備えることを特徴とする請求項１３に記載のシステム。
【請求項１５】
前記ビデオ・サンプル中の二人以上の顔画像を検出する手段は、フェイス・トラッカを備えることを特徴とする請求項１３に記載のシステム。
【請求項１６】
前記主要な話し手は、前記タイムラインに基づいて判断されることを特徴とする請求項１３に記載のシステム。
【発明の詳細な説明】
【技術分野】
【０００１】
（関連出願の相互参照）
本出願は、２００２年６月２１日に出願し、本出願の譲受人であるＭｉｃｒｏｓｏｆｔ Ｃｏｒｐ．に譲渡された本発明者による米国特許出願第１０／１７７３１５号「A System and Method for Distributed Meetings」の一部継続出願である。本出願人は、参照によりその開示、教示する内容全てが本明細書に組み込まれる前記出願の出願日について優先権を主張するものである。
【０００２】
以下の説明は、一般にはビデオ画像処理に関する。より詳細には、以下の説明はビデオ再生に用いるインデックス付タイムラインを提供することに関する。
【背景技術】
【０００３】
２人以上の話し手（speaker）を含む状況を録画したビデオの再生、例えば、録画された会議の再生は、通常インデックス付のタイムライン（時間記録）と同時に表示される。タイムラインを使用することで、ユーザは１つまたは複数のタイムライン制御手段を操作して会議中の特定の時点に素早く移動することができる。ビデオが２人以上の話し手が含む場合、それぞれが特定の話し手に関連する複数のタイムラインを使うことができる。各タイムラインは、対応する話し手がいつ話しているかを示す。それによって、ユーザは、会議中特定の話し手が話している部分に進むことができる。
【０００４】
このような複数のタイムラインは、例えば「話し手１」、「話し手２」等のように、一般的なやり方でラベルを付けて各話し手を識別することができる。タイムラインに特定の話し手の名前を自動的に付与する現行の技術は不正確であり、ユーザおよびユーザに関連する声紋（voiceprints）および顔紋（faceprints）のデータベースが必要となる可能性もあり、これらはセキュリティおよびプライバシーの問題を伴う可能性がある
【特許文献１】米国特許出願公開第２００３／０２３４８６６号明細書「A System and Method for Distributed Meetings」
【発明の開示】
【発明が解決しようとする課題】
【０００５】
本発明の目的は、ビデオ画像処理におけるビデオ再生に用いるインデックス付タイムラインを提供することにある。
【課題を解決するための手段】
【０００６】
上記目的を達成するために、本発明によれば、（１）ビデオ・サンプル中の１つまたは複数の顔画像を検出すること、前記ビデオ・サンプルに対応するオーディオ・サンプル中の１人または複数の話し手を検出すること、話し手識別子によって話し手を識別する話し手タイムラインと、前記話し手タイムラインの各時点における話し手の位置とを格納すること、顔データベースに検出された各話し手毎に少なくとも１つの顔画像を格納すること、および検出された各話し手に話し手タイムラインと顔画像とを関連付けることを備えることを特徴とする方法、（２）１人または複数の話し手をその中に含むオーディオ／ビジ

ュアル（Ａ／Ｖ）サンプルを表示すること、各話し手に対応する話し手タイムラインであって、時間的な経過のどの時点で前記話し手タイムラインに対応する前記話し手が話しているかを示す、話し手タイムラインを表示すること、各話し手タイムラインに、前記話し手タイムラインに関連付けられた前記話し手に対応する前記話し手の顔画像を関連付けること、および前記顔画像を前記対応する話し手タイムラインと共に表示することを備えることを特徴とする方法、（３）実行可能な命令を含む１つまたは複数のコンピュータ可読媒体であって、前記命令は実行されたとき、Ａ／Ｖサンプル中の各話し手を話し手識別子で識別すること、前記Ａ／Ｖサンプル中の各話し手の位置を識別すること、前記Ａ／Ｖサンプル中で識別された各話し手毎に少なくとも１つの顔画像を抽出すること、前記Ａ／Ｖサンプル中で識別された各話し手毎に、それぞれが時間、話し手識別子および話し手の位置を示す話し手タイムラインを作成すること、および話し手の前記顔画像を、同じ話し手に対応する話し手タイムラインと関連付けることを備える方法を実施することを特徴とする１つまたは複数のコンピュータ可読媒体、（４）１つまたは複数のコンピュータ可読媒体であって、Ａ／Ｖサンプルの各話し手毎の話し手タイムラインであって、それぞれが時間的経過の複数の時点における話し手と話し手の位置とを識別する話し手タイムラインを含む話し手タイムライン・データベースと、話し手タイムラインで識別された各話し手毎に少なくとも１つの顔画像と、各顔画像を前記話し手タイムライン・データベース中の適切な話し手タイムラインに関連付ける話し手識別子とを含む顔データベースとを備えることを特徴とする１つまたは複数のコンピュータ可読媒体、ならびに（５）Ａ／Ｖサンプルと、前記Ａ／Ｖサンプルに登場する各話し手を識別する手段と、前記Ａ／Ｖサンプルで識別された各話し手毎に顔画像を識別する手段と、前記Ａ／Ｖサンプルで識別された各話し手毎に話し手タイムラインを作成する手段と、顔画像を適切な話し手タイムラインに関連付ける手段とを備えることを特徴とするシステム、が提供される。

【０００７】

前述の諸態様および本発明のその他の利点の多くは、以下の詳細な説明を添付の図面と併せて参照することにより、より理解されよう。

【発明を実施するための最良の形態】

【０００８】

以下の説明は、複数の話し手が存在する環境において各話し手の顔を自動的に検出し、その話し手の１つまたは複数の顔画像を、その話し手に対応するタイムラインの部分に関連付けるための様々な実装および実施形態に関する。この種の特殊ラベル付けは、タイムラインのどの部分が複数の話し手のうちの特定の１人に対応しているかを視聴者がより容易に判断できる点で、汎用的なラベル付けと比べて優れている。

【０００９】

以下の議論では、２人以上の参加者および／または話し手が登場する会議の録画にパノラマ式カメラを使用する例について説明する。複数のカメラを含むパノラマ式カメラについて説明されているが、以下の説明は単一のカメラにも、および２台以上のカメラを有するマルチカメラ装置にも関連する。

【００１０】

パノラマ画像が、会議内の顔を検出し追跡するフェイス・トラッカ（ＦＴ）に入力される。マイクロホン・アレイが、音に基づいて話し手の位置を検出する音源ローカライザ（ＳＳＬ）に入力される。フェイス・トラッカおよび音源ローカライザからの出力は、仮想シネマトグラファ（virtual cinematographer）に入力され、複数の話し手の位置を検出する。

【００１１】

話し手は、話し手クラスタリング・モジュールで後処理され、これは、２つ以上の個別タイムラインを含む集約タイムラインをより明確に区別するために、話し手を時間的かつ空間的にクラスタ化する。（集約）タイムラインは、タイムライン・データベースに格納される。各話し手毎に１つまたは複数の画像を格納する顔データベースが作成されるが、それぞれの顔のうち少なくとも１つは話し手に関連付けられたタイムラインに使用される

。

【００１２】

本明細書で提示され特許請求される概念を、１つまたは複数の適切な動作環境に関連して以下に詳しく説明する。以下に説明する要素のうち幾つかは、「A System and Method for Distributed Meetings」の表題を有する、２００２年６月２１日に出願された、本出願の親出願特許文献１）にも記載されている。

【００１３】

例示的動作環境

図１は、汎用コンピュータ／カメラ・デバイスを示すブロック図である。コンピューティング・システム環境１００は、適切なコンピューティング環境の一例に過ぎず、特許請求の主題の範囲または機能に関するいかなる限定を示唆するものでもない。また、コンピューティング・システム環境１００も、例示的動作環境１００に示される構成要素の１つまたはそれらの組合せに依存するか、あるいはそれらを必要としていると解釈されるべきではない。

【００１４】

説明される技術および対象物は、他の幾多の汎用または専用のコンピューティング・システム環境または構成で動作させることができる。使用に適した周知のコンピューティング・システム、環境および／または構成としては、パーソナル・コンピュータ、サーバ・コンピュータ、携帯型またはラップトップ型装置、マルチ・プロセッサ・システム、マイクロ・プロセッサに基づくシステム、セットトップボックス、プログラム可能な家庭用電化製品、ネットワークＰＣ、ミニ・コンピュータ、メインフレーム・コンピュータ、上記のシステムまたは装置のうちいずれかを含む分散コンピューティング環境などがあるが、ただしこれに限定されない。

【００１５】

以下の説明は、コンピュータで実行される、例えばプログラム・モジュールなどのコンピュータ実行命令という一般的なコンテクストで表現することができる。一般的に、プログラム・モジュールとしては、特定のタスクを実行しまたは特定の抽象データ型を実装するルーチン、プログラム、オブジェクト、構成要素、データ構造等がある。説明する実装形態は、通信ネットワークを介してリンクされたリモート処理デバイスによってタスクが実行される分散コンピューティング環境で実施することもできる。分散コンピューティング環境では、プログラム・モジュールは、メモリ記憶装置を含むローカルまたはリモートのコンピュータ記憶媒体に置くことができる。

【００１６】

図１を参照すると、本発明を実施するための例示的システムは、コンピュータ１１０の形をとる汎用コンピューティング・デバイスを含む。コンピュータ１１０の構成要素は、処理装置１２０、システム・メモリ１３０、およびシステム・メモリを含めたシステムの様々な構成要素を処理装置１２０に連結するシステム・バス１２１を含むが、それだけには限定されない。システム・バス１２１は、メモリ・バスまたはメモリ・コントローラ、周辺バス、および様々なバス・アーキテクチャのいずれかを使用したローカル・バスを含めて、様々なバス構造を有するものでよい。限定ではなく例として挙げると、こうしたアーキテクチャとしては、ＩＳＡ（Industry Standard Architecture）バス、ＭＣＡ（Micro Channel Architecture）バス、ＥＩＳＡ（Enhanced ISA）バス、ＶＥＳＡ（Video Electronics Standards Association）ローカル・バス、およびメザニン・バス（Mezzanine bus）とも呼ばれるＰＣＩ（Peripheral Component Interconnect）バスがある。

【００１７】

コンピュータ１１０は、一般には様々なコンピュータ可読媒体を含む。コンピュータ可読媒体は、コンピュータ１１０がアクセスできる利用可能などんな媒体でもよく、揮発性および不揮発性の媒体、着脱式または非着脱式の媒体のいずれも含まれる。限定ではなく例として挙げると、コンピュータ可読媒体は、コンピュータ記憶媒体および通信媒体を備えることができる。コンピュータ記憶媒体としては、コンピュータ可読命令、データ構造

、プログラム・モジュールまたは他のデータ等の情報の記憶に用いられる任意の方式または技術で実装される揮発性および不揮発性の媒体、着脱式または非着脱式の媒体が含まれる。コンピュータ記憶媒体としては、ＲＡＭ、ＲＯＭ、ＥＥＰＲＯＭ、フラッシュ・メモリまたは他のメモリ技術、ＣＤ－ＲＯＭ、ＤＶＤ（digital versatile disks）または他の光学式ディスク記憶、磁気カセット、磁気テープ、磁気ディスク記憶または他の磁気記憶デバイス、あるいは所望の情報を格納するために使用され、コンピュータ１１０がアクセスできるような他のいかなる媒体なども含まれる、それだけには限定されない。通信媒体は、搬送波もしくは他の搬送機構等の変調データ信号の形で、一般に、コンピュータ可読命令、データ構造、プログラム・モジュール、または他のデータを具現化し、任意の情報送達媒体がこれに含まれる。「変調データ信号」という言葉は、その１つまたは複数の特性が、信号に含まれる情報を符号化するように設定または変更された信号を意味する。限定ではなく例として挙げると、通信媒体としては、有線ネットワークまたは直接配線接続等の有線媒体、および音響、ＲＦ、赤外線およびその他の無線媒体がある。上記の任意の組合せも、コンピュータ可読媒体の範囲に含まれるべきである。

【００１８】

システム・メモリ１３０は、例えば読取り専用メモリ（ＲＯＭ）１３１やランダム・アクセス・メモリ（ＲＡＭ）１３２など揮発性および／または不揮発性のメモリの形をとるコンピュータ記憶媒体を含む。ＲＯＭ１３１は、一般に、例えば起動の際にコンピュータ１１０内の要素間での情報の転送を助ける基本ルーチンを含む基本入出力システム（ＢＩＯＳ）１３３を格納している。ＲＡＭ１３２は、一般に、ダイレクトに処理装置１２０からアクセス可能であり、かつ／または処理装置１２０が操作を加えているデータおよび／またはプログラム・モジュールを格納している。限定ではなく例として挙げると、図１はオペレーティング・システム１３４、アプリケーション・プログラム１３５、他のプログラム・モジュール１３６、およびプログラム・データ１３７を示す。

【００１９】

コンピュータ１１０は、他の着脱式／非着脱式、揮発性／不揮発性のコンピュータ記憶媒体を含むこともできる。ほんの一例として、図１は非着脱式で不揮発性の磁気媒体を読み書きするハードディスク・ドライブ１４１、着脱式で不揮発性の磁気ディスク１５２を読み書きする磁気ディスク・ドライブ１５１、およびＣＤ－ＲＯＭや他の光学式媒体など着脱式で不揮発性の光ディスク１５６を読み書きする光ディスク・ドライブ１５５を示す。この例示的動作環境で使用できる他の着脱式／非着脱式、揮発性／不揮発性のコンピュータ記憶媒体としては、磁気テープ・カセット、フラッシュ・メモリ・カード、ＤＶＤ、デジタル・ビデオ・テープ、ソリッド・ステートＲＡＭ（solid state RAM）、ソリッド・ステートＲＯＭ（solid state ROM）などがあるが、それだけには限定されない。ハードディスク・ドライブ１４１は、一般にインターフェース１４０などの非着脱式メモリ・インターフェースを介してシステム・バス１２１に接続されており、磁気ディスク・ドライブ１５１および光ディスク・ドライブ１５５は、一般にインターフェース１５０などの着脱式メモリ・インターフェースを介してシステム・バス１２１に接続されている。

【００２０】

上に説明し、図１に示すドライブおよびそれらに関連するコンピュータ記憶媒体は、コンピュータ１１０用のコンピュータ可読命令、データ構造、プログラム・モジュールまたは他のデータを格納する。図１では、例えばハードディスク・ドライブ１４１は、オペレーティング・システム１４４、アプリケーション・プログラム１４５、他のプログラム・モジュール１４６、およびプログラム・データ１４７を記憶しているものとして示される。これらの構成要素は、オペレーティング・システム１３４、アプリケーション・プログラム１３５、他のプログラム・モジュール１３６、およびプログラム・データ１３７と同じものでも、別のものでもよいことに留意されたい。ここでは、オペレーティング・システム１４４、アプリケーション・プログラム１４５、他のプログラム・モジュール１４６、およびプログラム・データ１４７が少なくとも異なるコピーであることを示すために別の番号を付けた。ユーザは、キーボード１６２や、一般にマウス、トラックボールまたは

タッチパッドと称されるポインティング・デバイス１６１などの入力装置を介して、コンピュータ１１０に命令および情報を入力する。他の入力装置（図示せず）としては、マイクロホン、ジョイスティック、ゲーム・パッド、パラボラ・アンテナ、スキャナなどがある。これらおよび他の入力装置は、システム・バス１２１に結合されているユーザ入力インターフェース１６０を介して処理装置１２０に接続されることが多いが、パラレル・ポート、ゲーム・ポート、ユニバーサル・シリアル・バス（ＵＳＢ）などその他のインターフェースおよびバス構造によって接続することもできる。モニタ１９１または他の表示装置も、ビデオ・インターフェース１９０などのインターフェースを介してシステム・バス１２１に接続されている。モニタに加え、コンピュータは、出力周辺インターフェース１９５を介して接続することができるスピーカ１９７やプリンタ１９６など他の周辺出力装置を含むこともできる。本発明にとって特に重要なことに、パーソナル・コンピュータ１１０の入力装置として、一連の画像１６４を取り込むカメラ１６３（例えばデジタル／電子のスチルまたはビデオ・カメラ、あるいはフィルム／写真式スキャナ）も含むことができる。さらに、カメラは１つのみ図示されているが、パーソナル・コンピュータ１１０の入力装置として複数のカメラを含むこともできる。１つまたは複数のカメラの画像１６４は、適切なカメラ・インターフェース１６５を介してコンピュータ１１０に入力される。このインターフェース１６５はシステム・バス１２１に接続され、それによって、画像をＲＡＭ１３２またはコンピュータ１１０に関連する他のデータ記憶装置の１つに送り格納することが可能になる。しかし、画像データは、カメラ１６３の使用を必要とせず、前述のどんなコンピュータ可読媒体からもコンピュータ１１０に入力することができることに留意されたい。

【００２１】

コンピュータ１１０は、リモート・コンピュータ１８０など１つまたは複数のリモート・コンピュータへの論理接続を使用したネットワーク環境で動作することもできる。リモート・コンピュータ１８０は、パーソナル・コンピュータ、サーバ、ルータ、ネットワークＰＣ、ピア・デバイスまたは他の共通ネットワーク・ノードでよく、図１ではメモリ記憶装置１８１しか図示されていないが、一般にはコンピュータ１１０に関連して説明した上述の多くまたは全ての要素を含む。図１に描かれた論理接続は、ローカル・エリア・ネットワーク（ＬＡＮ）１７１およびワイド・エリア・ネットワーク（ＷＡＮ）１７３を含むが、他のネットワークを含むこともできる。これらのネットワーク環境は、オフィス、事業体規模のコンピュータ・ネットワーク、イントラネットおよびインターネットにおいて一般的となっている。

【００２２】

ＬＡＮネットワーク環境で使用する場合、コンピュータ１１０はネットワーク・インターフェースまたはアダプタ１７０を介してＬＡＮ１７１に接続される。ＷＡＮネットワーク環境で使用する場合、コンピュータ１１０は一般に、インターネット等のＷＡＮ１７３上での通信を確立するための、モデム１７２または他の手段を含む。内部、外部のどちらでもよいモデム１７２は、ユーザ入力インターフェース１６０または他の適切な機構を介してシステム・バス１２１に接続することができる。ネットワーク環境では、コンピュータ１１０に関連して描かれたプログラム・モジュールあるいはその一部は、リモート・メモリ記憶装置に格納することができる。限定ではなく例として挙げると、図１ではリモート・アプリケーション・プログラム１８５がメモリ装置１８１上にあるものとして示されている。図示されたネットワーク接続は例示的なものであり、コンピュータ間の通信リンクを確立する他の手段を使用することもできることが理解できよう。

【００２３】

例示的パノラマ式カメラおよびクライアント装置

図２は、例示的なパノラマ式カメラ機器２００および例示的なクライアント装置２２２を示すブロック図である。特定の構成が示されているが、パノラマ式カメラ機器２００は、パノラマ式カメラまたはその機能的等価品を含むどんな装置でもよいことに留意されたい。本明細書に記載される１つまたは複数の技術を組み込んだ実用的適用例では、この図

でパノラマ式カメラ機器２００に含まれている構成要素よりも多くの、あるいは少ない構成要素を含むことができる。

【００２４】

パノラマ式カメラ機器２００は、プロセッサ２０２およびメモリ２０４を含む。パノラマ式カメラ機器２００は、複数のカメラ２０６（２０６＿１から２０６＿ｎで示される）が撮影した画像の幾つかをつなぎ合わせることでパノラマ画像を作成する。パノラマ画像は、完全な３６０度のパノラマ画像でも、その一部のみでもよい。パノラマ式カメラ機器２００が図示され説明されているが、ここで説明される技術は単一のカメラでも活用できることに留意されたい。

【００２５】

パノラマ式カメラ機器２００はさらに、マイクロホン・アレイ２０８を含む。以下に詳細に説明するが、マイクロホン・アレイは、音の方向が位置測定されるように構成される。言い換えれば、マイクロホン・アレイに入力された音の分析によって、検出された音が発生した方向が得られる。スピーカフォン（speakerphone）を使用可能にし、あるいはユーザに通知信号などを発するために、スピーカ２１０をパノラマ式カメラ機器２００に含めてもよい。

【００２６】

メモリ２０４は、較正データ、露出設定、ステッチング表（stitching table）等の幾つかのカメラ設定２１２を格納する。メモリ２０４には、１つまたは複数の他のカメラ・ソフトウェア・アプリケーション２１６と共に、カメラ機能を制御するオペレーティング・システム２１４も格納されている。

【００２７】

パノラマ式カメラ機器２００は、さらにパノラマ式カメラ機器２００との間でのデータの送受信に用いる入出力（Ｉ／Ｏ）モジュール２１８、およびカメラの機能が必要とする可能性のある他の種々のハードウェア２２０要素も含む。

【００２８】

パノラマ式カメラ機器２００は、少なくとも１つのクライアント装置２２２と通信する。クライアント装置２２２は、プロセッサ２２４、メモリ２２６、大容量記憶装置２４２（ハードディスク・ドライブ等）、および図で下に示すクライアント装置２２２に帰せられる機能を実行するために必要となるかもしれない他のハードウェア２３０を含む。

【００２９】

メモリ２２６は、フェイス・トラッカ（ＦＴ）モジュール２３０および音源位置測定（ＳＳＬ：sound source localization）モジュール２３２を格納している。フェイス・トラッカ・モジュール２３０および音源位置測定モジュール２３２は、仮想シネマトグラファ２３４と共に使って、カメラ・シーン内の人物を検出し、その人物が話しているかどうか、話しているとすればいつなのかを決定する。音源位置測定に関する幾つかの従来方式を使用することができる。親出願（特許文献１）に記載のものを含めて、様々なフェイス・トラッカ方式（または人物検出および追跡システム）が、本明細書の説明のように使用することができる。

【００３０】

メモリ２２６は、また、話し手クラスタリング・モジュール２３６を格納し、これは、２人以上の人物が話している際に主要な話し手を判断して、主要な話し手にタイムラインの特定部分を集中させるように、構成される。ほとんどの会議の状況では、２人以上の人物が同時に話すことがある。通常は、主要な話し手が話している最中に別の人物が短時間話し手に横やりを入れたり、あるいは話し手に重なって話す。話し手クラスタリング・モジュール２３６は、話し手を時間的かつ空間的にクラスタ化してタイムラインを整理するように構成される。

【００３１】

タイムライン２３８は、仮想シネマトグラファ２３４によって作成される。タイムライン２３８は、大容量記憶装置２４２上のタイムライン・データベース２４４に格納される

。タイムライン・データベース２４４は、複数のフィールドを含む。これらのフィールドとしては、時間、話し手の番号、およびカメラ画像（ｘ、ｙ、幅、高さ）内の話し手バウンディング・ボックス（bounding box）等があるが、必ずしもそれだけには限定されない。タイムライン・データベース２４４はさらに、話し手の１つまたは複数の顔の角度（方位角と仰角）を含むことができる。

【００３２】

メモリ２２６には、フェイス・エクストラクタ（顔領域抽出）モジュール２４０も格納され、カメラ画像の（フェイス・トラッカ２３０が識別した）顔バウンディング・ボックスから話し手の顔の画像を抽出するように構成される。フェイス・エクストラクタ・モジュール２４０は、抽出した顔画像を大容量記憶装置２４２上の顔データベース２４６に格納する。

【００３３】

少なくとも１つの実装では、１人または複数の話し手について複数の顔画像を格納することができる。どの顔画像がいつ使用されるかを決定するために、パラメータを指定することができる。あるいは、ユーザが、複数の顔画像から手動で特定の顔画像を選択することを可能とすることもできる。

【００３４】

少なくとも１つの代替実装では、それぞれの話し手に１つの顔画像のみが格納される。記憶される顔画像は、フェイス・エクストラクタ・モジュール２４０が抽出した単一の画像でもよいが、フェイス・エクストラクタ・モジュール２４０は、話し手の最良の画像を選択するように構成することもできる。

【００３５】

話し手の最良の画像の選択は、（正面の顔の画像が別の画像よりも優れた写真であると仮定して）正面の顔角度を識別することによって、最小の動きを示す顔画像を識別することによって、あるいは顔の対称性を最大化している顔画像を識別することによって、行うことができる。

【００３６】

録画された会議２４８も、後ほど呼び出して、再生できるように、大容量記憶装置２４２に記憶される。

【００３７】

図２に示され、図２と関連して説明した要素および機能は、後続の図面に関連して、以下に、より詳しく説明する。

【００３８】

例示的な再生画面

図３は、パノラマ画像３０２および顔画像タイムライン３０４を含む再生画面３００の線画である。第１の会議参加者３０３および第２の会議参加者３０５を含むパノラマ画像３０２が、と示されている。図の再生画面３００は、さらにタイトル・バー３０６および個人画像３０８も含んでいる。個人画像３０８は、オプション機能であり、そこでは、特定の個人、一般には主要な話し手、に焦点が当てられる。図３では、その個人画像３０８は、第１の会議参加者３０３の顔画像を表示している。

【００３９】

例示的な再生画面３００は、さらに再生ボタン、早送りボタン、巻き戻しボタンなど、メディア・プレイヤで通常見られる制御手段を備える制御セクション３１０を含む。再生画面３００は、再生画面３００の主題に関する情報を表示できる情報エリア３１２を含む。例えば、会議の題名、会議室の番号、会議出席者のリストなどを情報エリア３１２に表示してもよい。

【００４０】

顔画像タイムライン３０４は、第１の会議参加者３０３に対応する第１のサブタイムライン３１４、および第２の会議参加者３０５に対応する第２のサブタイムライン３１６を含む。それぞれのサブタイムライン３１４および３１６は、ある時間的連続体（temporal

continuum）の中で対応する会議参加者が話している箇所（セクション）を示す。ユーザは、サブタイムライン３１４、３１６の任意の時点に直接アクセスして、特定の会議参加者が話している会議の部分に直ちにアクセスすることができる。

【００４１】

第１の会議参加者３０３の第１の顔画像３１８が、第１のサブタイムライン３１４の隣に現れて、第１のサブタイムライン３１４が第１の会議参加者３０３に対応することを示す。第２の会議参加者３０５の顔画像３２０が第２のサブタイムライン３１６の隣に現れて、第２のサブタイムライン３１６が第２の会議参加者３０５に対応することを示す。

【００４２】

図４は、図３に示し説明した例示的再生画面３００に類似した要素を含む例示的再生画面４００を示す。図３に関して示し説明した要素および参照番号を、図４の例示的再生画面４００に関して使用する。

【００４３】

例示的再生画面４００は、パノラマ画像３０２および顔画像タイムライン３０４を含む。パノラマ画像３０２は、第１の会議参加者３０３および第２の会議参加者３０５を示す。タイトル・バー３０６が再生画面４００の上部の両端間に延びており、個人画像３０８は第２の会議参加者３０５を示している。

【００４４】

例示的再生画面４００は、さらに、ホワイト・ボードの前に位置する会議参加者（この場合は、第２の会議参加者３０５）を表示するホワイト・ボード話し手画像を含む。ホワイト・ボード話し手画像は、図３の再生画面３００に含まれていないが、ここではどのようにして、任意の特定再生画面３００、４００に他の画像をどのように含むことができるかを説明するために、使用している。

【００４５】

制御セクション３１０は、マルチメディア制御を含み、情報エリア３１２は再生画面４００で表示されている会議に関する情報を表示する。

【００４６】

顔画像タイムライン３０４は、第１のサブタイムライン３１４、第２のサブタイムライン３１６、および第３のサブタイムライン４０２を含む。図３には２つのサブタイムラインしか示してないが、タイムラインは管理できる限り任意の数のサブタイムラインを含むことができることに留意されたい。例えば、図４は、３つのサブタイムラインが存在する。

【００４７】

この例では会議参加者が２名しかいないにも関わらず、サブタイムラインが３つあることに留意されたい。これは、単一の話し手が２つ以上のサブタイムラインに関連付けられる可能性が存在するからである。この例では、第２のサブタイムライン３１６は、ホワイト・ボードの前にいるときの第２の会議参加者３０５に関連付けられており、第３のサブタイムライン４０４はホワイト・ボードの前以外の場所にいるときの第２の会議参加者３０５に関連付けられている。

【００４８】

この状況は、会議参加者が会議中に複数の場所を占める場合に起こる。仮想シネマトグラファ２３４はこの場合、３つの場所で話し手を検出した。仮想シネマトグラファ２３４は、これらの場所には２人の話し手しかいないことを必ずしも知らない。この機能は、ユーザが、ある特定の位置にいるときの話し手に、主に関心がある場合に、ユーザの役に立つ。例えば、ユーザは、録画した会議の中で、ある話し手がホワイト・ボードの前にいる箇所のみを再生したいと思うかもしれない。

【００４９】

例示的再生画面４００はさらに、第１の会議参加者３０３の第１の顔画像３１８、および第２の会議参加者３０５の第２の顔画像３２０を含む。さらに、第３の顔画像４０４も含まれ、これは第３のサブタイムライン４０２に関連する。第３の顔画像４０４は、第２

の会議参加者３０５の第２の場所に対応する。

【００５０】

例示的再生画面３００、４００を示すのに使用した技術については、後続の図面に関して以下により詳しく説明する。

【００５１】

例示的な方法論的実装：顔画像タイムラインの作成

図５は、顔画像付きのタイムラインを作成するための方法論的実装の例示的なフロー図５００である。例示的フロー図５００についての以下の説明では、前の図に示される要素および参照番号を引き続き参照する。

【００５２】

ブロック５０２で、パノラマ式カメラ機器２００が、１つまたは複数のビデオ画像を抽出してパノラマ画像を作成する。パノラマ画像は、画像内の顔を検出して追跡するフェイス・トラッカ２３０に入力される（ブロック５０４）。ほぼ同時に、ブロック５０６で、マイクロホン・アレイ２０８がパノラマ画像に対応する音を抽出し、その音をブロック５０８で、抽出した音に基づいて話し手の位置を検出する音源ローカライザ２３２に入力する。

【００５３】

仮想シネマトグラファ２３４は、フェイス・トラッカ２３０および音源ローカライザ２３２からのデータを処理して、ブロック５１０でタイムライン２３８を作成する。ブロック５１２で、先に述べたように、話し手クラスタリング・モジュール２３６が、話し手を時間的かつ空間的にクラスタ化して、タイムライン２３８の諸部分を統合して明確にする。

【００５４】

タイムラインは、次に挙げるフィールド、すなわち、時間、話し手の番号、画像内の話し手バウンディング・ボックス（ｘ、ｙ、幅、高さ）、話し手の顔の角度（方位角と仰角）など、を付けて、タイムライン・データベース２４４に記憶される。

【００５５】

フェイス・トラッカ２３０が導出したパノラマ画像および顔認識座標（すなわち、顔バウンディング・ボックス）を使用して、ブロック５１４でフェイス・エクストラクタ２４０が話し手の顔画像を抽出する。抽出された顔画像は、顔データベース２４６に記憶され、話し手番号に関連付けられる。

【００５６】

先に述べたように、フェイス・エクストラクタ２４０は、各話し手毎に複数の画像を抽出して、フェイス・エクストラクタ２４０がタイムライン２３８の中で最良の画像であると判断した画像を使用するように、構成することができる。

【００５７】

「最良」の画像を選択し、顔データベース２４６を作成するための例示的な方法論的実装を、図６に関して以下に示し説明する。

【００５８】

例示的な方法論的実装：顔データベースの作成

図６は、顔データベースを作成するための方法論的実装を示す例示的なフロー図６００である。図６についての以下の説明では、前の複数の図に示された要素および参照番号を引き続き参照する。

【００５９】

ブロック６０２で、フェイス・エクストラクタ２４０が先に述べたようにパノラマ画像から顔画像を抽出する。顔データベース２４６に話し手の顔画像がまだ記憶されていない場合（ブロック６０４の「Ｎｏ」ブランチ）、顔画像はブロック６１０で顔データベース２４６に記憶される。顔画像が記憶されているかどうかの判断は、顔画像に登場する人物の肖像の画像が既に記憶されている画像と類似するイメージを有するかどうかに必ずしも依存せず、むしろ識別された話し手がその話し手に対応する既に記憶されている画像を有

するかどうかに依存することに留意されたい。したがって、第１の場所にいる、ある話し手が、既に格納された顔画像を有し、その後、その話し手が第２の場所で検出された場合、その話し手が既に記憶されている顔画像を有するかどうかを判断するのに、第２の場所にいる話し手の顔画像を第１の場所にいる話し手の記憶済み顔画像と比較することはない。

【００６０】

話し手の顔画像が既に顔データベース２４６に記憶されている（以下、「記憶済み顔画像」と呼ぶ）場合（ブロック６０４の「Ｙｅｓ」ブランチ）、その顔画像はブロック６０６で記憶済み顔画像と比較される。フェイス・エクストラクタ２４０が、その顔画像が記憶済み顔画像よりも良いか、あるいはより好ましいと判断した場合（ブロック６０８の「Ｙｅｓ」ブランチ）、その顔画像は顔データベース２４６に記憶され、それによって以前の記憶済み顔画像が上書きされる。

【００６１】

顔画像が記憶済み顔画像よりも良くなかった場合（ブロック６０８の「Ｎｏ」ブランチ）、顔画像は破棄され、記憶済み顔画像が保持される。

【００６２】

どの顔画像がより優れた顔画像かを判断する基準は数多く、様々である。例えば、話し手が最も正面を向いたところをとらえた顔画像を「最良」の顔画像と判断するようにフェイス・エクストラクタ２４６を構成することができる。あるいは、第１の顔画像が動きを示し、第２の顔画像が動きを示さない場合、第２の顔画像が「最良」の顔画像であるとフェイス・エクストラクタ２４６が判断してもよい。あるいは、話し手の複数の画像のうち、どれが最大の対称性を示しているかを決定し、その顔画像をタイムラインに使用するようにフェイス・エクストラクタ２４６を構成することもできる。タイムラインに使用する最適の顔画像の判断には、ここに列挙した以外の基準も使うことができる。

【００６３】

他の話し手が存在する場合（ブロック６１２の「Ｙｅｓ」ブランチ）、処理はブロック６０２に戻り、一意の各話し手毎にこの処理が繰り返される。この場合も、このコンテクストで使用される「一意の話し手」は、異なる発言位置（speaking locations）に登場する人物が異なる話し手と解釈される可能性もあるので、必ずしも固有の人物を意味する必要はない。からである。識別すべき一意の話し手がそれ以上存在しない場合、処理は終了する（ブロック６１２の「Ｎｏ」ブランチ）。

【００６４】

結論

１つまたは複数の例示的実装を図示し説明したが、本明細書に添付した特許請求の範囲の精神および範囲から逸脱することなしに、様々な変更が行えることが理解されよう。

【図面の簡単な説明】

【００６５】

【図１】例示的な汎用コンピューティング／カメラ装置を示すブロック図である。

【図２】例示的なパノラマ式カメラおよびクライアント装置を表すブロック図である。

【図３】パノラマ画像および顔画像のタイムラインを含む例示的な再生画面を表現した図である。

【図４】パノラマ画像および顔画像のタイムラインを含む例示的な再生画面を示す図である。

【図５】顔画像付きのタイムライン作成の方法論的実装の例示的なフロー図である。

【図６】顔データベースの作成の方法論的実装を示す例示的なフロー図である。

【符号の説明】

【００６６】

１００　コンピューティング・システム環境

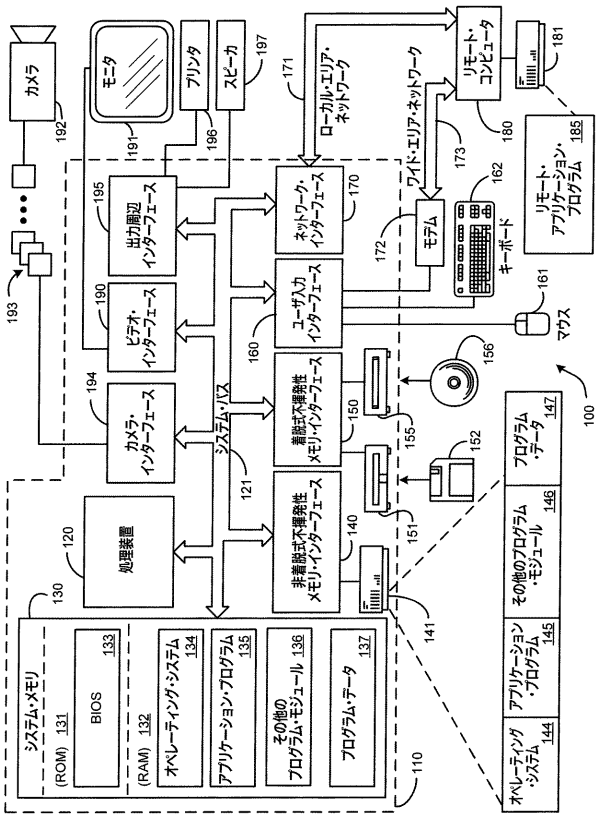
１１０　コンピュータ

１２０　処理装置

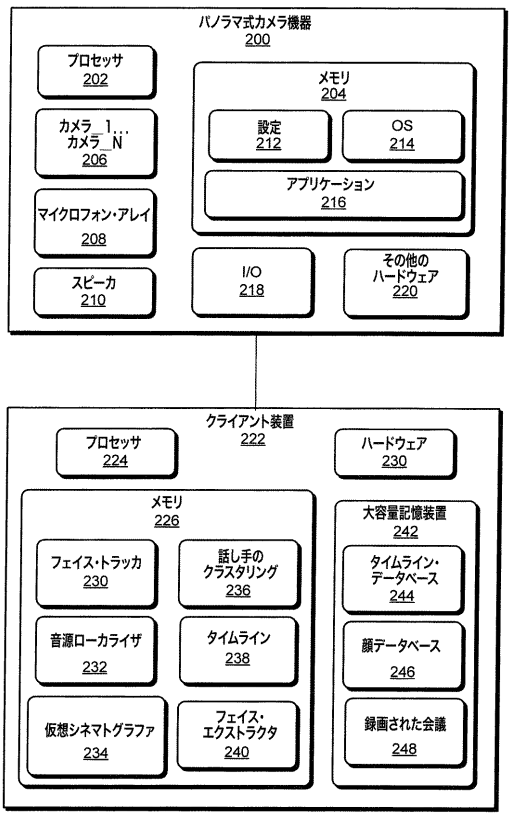
１２１　システム・バス
１３０　システム・メモリ
１３４　オペレーティング・システム
１３５　アプリケーション・プログラム
１３６　その他のプログラム・モジュール
１３７　プログラム・データ
１４０　非着脱式不揮発性メモリ・インターフェース
１４１　ハードディスク・ドライブ
１４４　オペレーティング・システム
１４５　アプリケーション・プログラム
１４６　その他のプログラム・モジュール
１４７　プログラム・データ
１５０　着脱式不揮発性メモリ・インターフェース
１５１　磁気ディスク・ドライブ
１５２　磁気ディスク
１５５　光ディスク・ドライブ
１５６　光ディスク
１６０　ユーザ入力インターフェース
１６１　マウス
１６２　キーボード
１７０　ネットワーク・インターフェース
１７１　ローカル・エリア・ネットワーク
１７２　モデム
１７３　ワイド・エリア・ネットワーク
１８０　リモート・コンピュータ
１８１　メモリ装置
１８５　リモート・アプリケーション・プログラム
１９０　ビデオ・インターフェース
１９１　モニタ
１９２　カメラ
１９４　カメラ・インターフェース
１９５　出力周辺インターフェース
１９６　プリンタ
１９７　スピーカ
２００　パノラマ式カメラ機器
２０２　プロセッサ
２０４　メモリ
２０６　カメラ＿１．．．カメラ＿Ｎ
２０８　マイクロホン・アレイ
２１０　スピーカ
２１２　設定
２１４　ＯＳ
２１６　アプリケーション
２１８　Ｉ／Ｏ
２２０　その他のハードウェア
２２２　クライアント装置
２２４　プロセッサ
２２６　メモリ
２３０　ハードウェア
２３０　フェイス・トラッカ

２３２　音源ローカライザ
２３４　仮想シネマトグラファ
２３６　話し手のクラスタリング
２３８　タイムライン
２４０　フェイス・エクストラクタ
２４２　大容量記憶装置
２４４　タイムライン・データベース
２４６　顔データベース
２４８　録画された会議
３００　再生画面
３０２　パノラマ画像
３０３　第１の会議参加者
３０４　顔画像タイムライン
３０５　第２の会議参加者
３０６　タイトル・バー
３０８　個人画像
３１０　制御セクション（コントロール）
３１２　情報エリア
３１４　サブタイムライン
３１６　サブタイムライン
３１８　第１の会議参加者３０３の顔画像
３２０　第２の会議参加者３０５の顔画像
４００　例示的再生画面

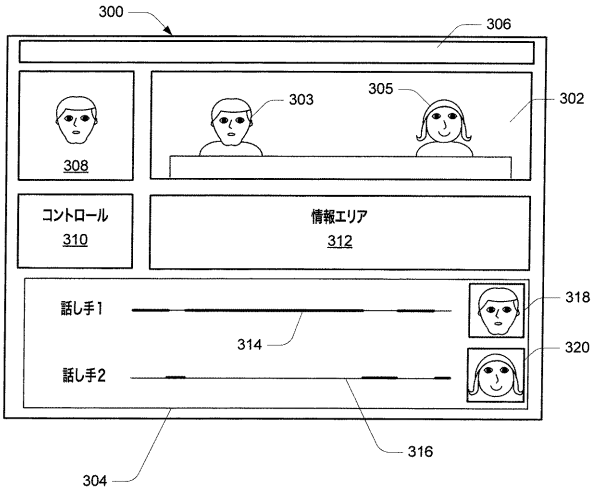
【図１】



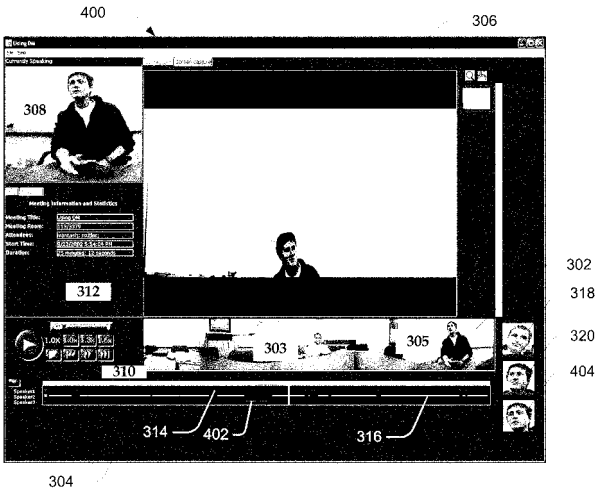
【図２】



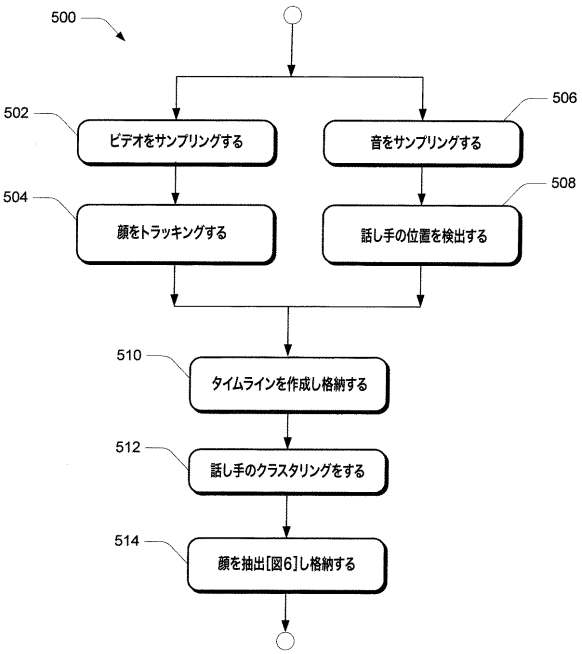
【図３】


300
306
303
305
302
308
コントロール
310
情報エリア
312
話し手1
318
314
話し手2
320
316
304

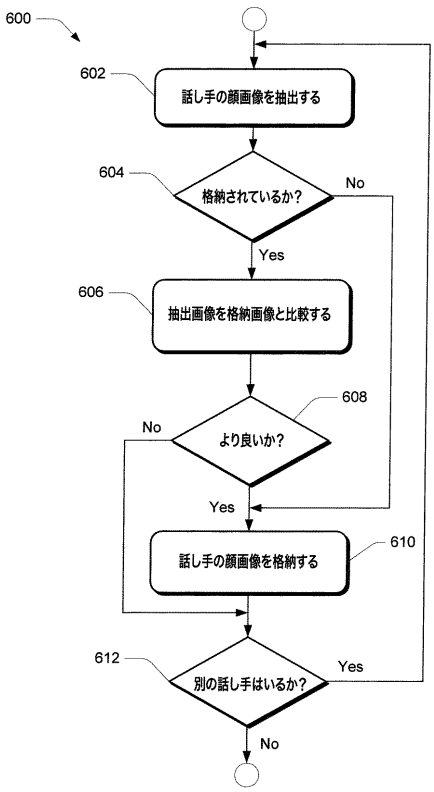
【図４】


400
306
308
312
302
318
303
305
320
404
310
314
402
316
304

【図５】


500
502
ビデオをサンプリングする
506
音をサンプリングする
504
顔をトラッキングする
508
話し手の位置を検出する
510
タイムラインを作成し格納する
512
話し手のクラスタリングをする
514
顔を抽出[図6]し格納する

【図６】


600
602
話し手の顔画像を抽出する
604
格納されているか?
No
Yes
606
抽出画像を格納画像と比較する
608
No
より良いか?
Yes
610
話し手の顔画像を格納する
612
別の話し手はいるか?
Yes
No

フロントページの続き

(74)代理人 100091063
弁理士 田中 英夫
(74)代理人 100153028
弁理士 上田 忠
(74)代理人 100120112
弁理士 中西 基晴
(74)代理人 100113974
弁理士 田中 拓人
(72)発明者 ロス ジー. カトラー
アメリカ合衆国 ９８０５２ ワシントン州 レッドモンド ワン マイクロソフト ウェイ マイクロソフト コーポレーション内

審査官 竹中 辰利

(56)参考文献 特開２０００－１２５２７４（ＪＰ，Ａ）
特開２００３－２３００４９（ＪＰ，Ａ）
特開２０００－０３６０５２（ＪＰ，Ａ）
特開２００２－２５１３９３（ＪＰ，Ａ）
再公表特許第９７／００９６８３（ＪＰ，Ａ１）

(58)調査した分野(Int.Cl.，ＤＢ名)
Ｈ０４Ｎ ５／７６
Ｇ０６Ｔ ７／００