

(19) 日本国特許庁 (JP)

(12) 特 許 公 報 (B2)

(11) 特許番号

特許第6486377号  
(P6486377)

(45) 発行日 平成31年3月20日 (2019.3.20)

(24) 登録日 平成31年3月1日 (2019.3.1)

(51) Int. Cl.

F I

H04N 7/15 (2006.01)

H04N 7/15

G06F 13/00 (2006.01)

G06F 13/00 550L

G06T 1/00 (2006.01)

G06T 1/00 340B

請求項の数 10 (全 23 頁)

(21) 出願番号 特願2016-554533 (P2016-554533)  
 (86) (22) 出願日 平成26年11月17日 (2014.11.17)  
 (65) 公表番号 特表2017-503442 (P2017-503442A)  
 (43) 公表日 平成29年1月26日 (2017.1.26)  
 (86) 国際出願番号 PCT/US2014/065855  
 (87) 国際公開番号 W02015/077159  
 (87) 国際公開日 平成27年5月28日 (2015.5.28)  
 審査請求日 平成29年10月20日 (2017.10.20)  
 (31) 優先権主張番号 14/084,090  
 (32) 優先日 平成25年11月19日 (2013.11.19)  
 (33) 優先権主張国 米国 (US)

(73) 特許権者 314015767  
 マイクロソフト テクノロジー ライセン  
 シング, エルエルシー  
 アメリカ合衆国 ワシントン州 9805  
 2 レッドモンド ワン マイクロソフト  
 ウェイ  
 (74) 代理人 100107766  
 弁理士 伊東 忠重  
 (74) 代理人 100070150  
 弁理士 伊東 忠彦  
 (74) 代理人 100091214  
 弁理士 大貫 進介

最終頁に続く

(54) 【発明の名称】 ビデオ送信

(57) 【特許請求の範囲】

【請求項 1】

ユーザデバイスであって、

画像キャプチャデバイスから受信された、1人以上のユーザのビデオを、ネットワークを介して、少なくとも別のユーザデバイスに送信するよう構成されているネットワークインタフェースと、

複数の視覚的なユーザ特徴から1以上の特徴を選択し、前記の選択された1以上の特徴を追跡するために、前記の選択された1以上の特徴の検出に基づいて、前記ビデオを制御するよう構成されているビデオコントローラであって、前記の選択された1以上の特徴の前記検出は、前記1人以上のユーザのうちの少なくとも1人のユーザに関して、前記少なくとも1人のユーザの複数の検出された骨格ポイントを受信することを含み、前記ビデオは、前記複数の検出された骨格ポイントに基づいて制御される、ビデオコントローラと、

前記ユーザデバイスと前記別のユーザデバイスとの間の通信チャネルと、前記ユーザデバイス及び/又は前記別のユーザデバイスの1以上のリソースと、のうちの少なくとも一方に関する情報を受信し、前記の受信された情報に基づいて、前記ビデオコントローラによる該選択を制御するよう構成されているリソースマネージャと、

を備えたユーザデバイス。

【請求項 2】

前記複数の視覚的なユーザ特徴は、ユーザの複数の身体パーツである、請求項1記載のユーザデバイス。

10

20

## 【請求項 3】

前記ビデオは、複数のユーザのものであり、前記ビデオコントローラは、前記複数のユーザの各々に関する選択された特徴を追跡するために、前記複数のユーザの各々に関する前記選択された特徴の検出に基づいて、前記ビデオを制御するようさらに構成されている、請求項 1 記載のユーザデバイス。

## 【請求項 4】

前記ネットワークインタフェースは、前記画像キャプチャデバイスから受信された、前記 1 人以上のユーザのさらなるビデオを、前記ネットワークを介して、さらなるユーザデバイスに送信するようさらに構成されており、

前記ビデオコントローラは、前記複数の視覚的なユーザ特徴から 1 以上のさらなる特徴を選択し、前記の選択された 1 以上のさらなる特徴を追跡するために、前記の選択された 1 以上のさらなる特徴の検出に基づいて、前記さらなるビデオを制御するようさらに構成されており、

10

前記リソースマネージャは、前記ユーザデバイスと前記さらなるユーザデバイスとの間の通信チャネルと、前記さらなるユーザデバイスの 1 以上のリソースと、のうちの少なくとも一方に関するさらなる情報を受信し、前記の受信されたさらなる情報に基づいて、前記ビデオコントローラによる該選択を制御するようさらに構成されている、請求項 1 記載のユーザデバイス。

## 【請求項 5】

前記制御は、前記の選択された 1 以上の特徴の前記検出に基づいて境界データを生成することを含む、請求項 1 記載のユーザデバイス。

20

## 【請求項 6】

前記制御は、前記の生成された境界データに基づいて、前記画像キャプチャデバイスから受信されたビデオデータを処理することを含む、請求項 5 記載のユーザデバイス。

## 【請求項 7】

前記ビデオデータを処理することは、前記の生成された境界データに基づいて、前記ビデオデータをクロップすることを含む、請求項 6 記載のユーザデバイス。

## 【請求項 8】

ユーザデバイスから別のユーザデバイスに 1 人以上のユーザのビデオを送信する方法であって、

30

画像キャプチャデバイスから前記ビデオを受信するステップと、

前記ユーザデバイスと前記別のユーザデバイスとの間の通信チャネルと、前記ユーザデバイス及び / 又は前記別のユーザデバイスの 1 以上のリソースと、のうちの少なくとも一方に関する情報を受信するステップと、

前記の受信された情報に基づいて、複数の視覚的なユーザ特徴から 1 以上の特徴を選択するステップと、

前記の受信されたビデオにおいて、前記の選択された 1 以上の特徴を検出するステップと、

前記の選択された 1 以上の特徴を追跡するために、前記の選択された 1 以上の特徴の前記検出に基づいて、前記ビデオを制御するステップであって、前記の選択された 1 以上の特徴の前記検出は、前記 1 人以上のユーザのうちの少なくとも 1 人のユーザに関して、前記少なくとも 1 人のユーザの複数の検出された骨格ポイントを受信することを含み、前記ビデオは、前記複数の検出された骨格ポイントに基づいて制御される、ステップと、

40

前記ビデオを前記別のユーザデバイスに送信するステップと、  
を含む方法。

## 【請求項 9】

ユーザデバイスであって、

ネットワークに接続するためのネットワークインタフェースと、

1 以上のプロセッサであって、

画像キャプチャデバイスから受信された、1 人以上のユーザのビデオを、前記ネット

50

ワークを介して、別のユーザデバイスに送信し、

複数の視覚的なユーザ特徴から特徴を選択し、

第1の時間及び第2の時間のそれぞれにおいて、前記の選択された特徴の検出に基づいて、第1のセットの境界データ及び第2のセットの境界データを生成し、

動的モデルを使用して、前記第1のセットの境界データ及び前記第2のセットの境界データに基づいて、過渡データを生成し、

前記の選択された特徴を追跡するために、前記過渡データに基づいて前記ビデオを制御する

よう構成されている1以上のプロセッサと、

を備えたユーザデバイス。

10

#### 【請求項10】

ユーザデバイスからネットワークを介して別のユーザデバイスに1人以上のユーザのビデオを送信する方法であって、

画像キャプチャデバイスから前記ビデオを受信するステップと、

複数の視覚的なユーザ特徴から特徴を選択するステップと、

第1の時間及び第2の時間のそれぞれにおいて、前記の選択された特徴の検出に基づいて、第1のセットの境界データ及び第2のセットの境界データを生成するステップと、

動的モデルを使用して、前記第1のセットの境界データ及び前記第2のセットの境界データに基づいて、過渡データを生成するステップと、

前記の選択された特徴を追跡するために、前記過渡データに基づいて前記ビデオを制御するステップと、

20

前記ネットワークを介して前記別のユーザデバイスに前記ビデオを送信するステップと、

、

を含む方法。

#### 【発明の詳細な説明】

#### 【背景技術】

#### 【0001】

従来の通信システムは、パーソナルコンピュータやモバイルデバイス等といったデバイスのユーザが、インターネット等のパケットベースのコンピュータネットワークを介して音声通話又はビデオ通話を行うことを可能にしている。そのような通信システムは、ボイスオーバーインターネットプロトコル(VoIP)システム又はビデオオーバーインターネットプロトコル(VoIP)システムを含む。これらのシステムは、従来の固定回線網又はモバイルセルラネットワークよりも、しばしば著しく低コストであるため、ユーザにとって有益である。これは、特に長距離通信の場合に当てはまり得る。VoIPシステムを使用するために、ユーザは、自分のデバイスにクライアントソフトウェアをインストールして実行する。クライアントソフトウェアは、VoIP接続をセットアップするだけでなく、登録及びユーザ認証等の他の機能も提供する。音声通信に加えて、クライアントは、インスタントメッセージング(「IM」)、SMSメッセージング、ファイル転送、及びボイスメール等の他の通信メディアのための接続もセットアップする。

30

#### 【0002】

40

最近では、ネットワーク間(internet)能力及び機能が、テレビジョンセットの他の(例えば大型スクリーンの)表示手段に接続されるよう構成されているゲームコンソール、テレビジョンセット自体(「スマートTV」としばしば呼ばれる)、テレビジョンセットに接続されるよう構成されているセットトップボックス等といったユーザデバイスに統合されている。これは、インターネット等のパケットベースのコンピュータネットワークを介する通信を可能にするための、ゲームコンソール、テレビジョンセット、セットトップボックス(又は同様のもの)へのクライアントソフトウェアの統合を含む。クライアントソフトウェアのこの統合は、遠端ユーザから受信されるビデオ信号を近端ユーザに出力することにより、ビデオ通話のために大型の高解像度スクリーンを利用することを可能にしている。さらに、特に、電気幹線により電力供給される大型のコンシューマ電子デバイス

50

に対する電力要件は、例えばモバイルデバイスよりも厳格ではないので、大きな処理能力が、セッソップボックスやTV等といったユーザデバイス内に提供され得る。これは、ユーザデバイスに接続されているカメラ又は他の同様の画像入力手段から受信されるビデオデータの高品質音声・ビデオ符号化等といったフルレンジの機能を、組み込まれた通信クライアントに含めることを可能にすることができる。

#### 【発明の概要】

##### 【0003】

この発明の概要は、発明を実施するための形態において以下でさらに説明されるコンセプトのうち選択したものを簡略化した形で紹介するために提供される。この発明の概要は、特許請求される主題の主要な特徴又は必要不可欠な特徴を特定することを意図するものではないし、特許請求される主題の範囲を限定するために使用されることを意図するものでもない。

##### 【0004】

ネットワークインタフェース、ビデオコントローラ、及びリソースマネージャを備えたユーザデバイスが開示される。ネットワークインタフェースは、画像キャプチャデバイスから受信された、1人以上のユーザのビデオを、ネットワークを介して、少なくとも別のユーザデバイスに送信するよう構成されている。ビデオコントローラは、複数の視覚的なユーザ特徴から1以上の特徴を選択し、選択された特徴を追跡するために、選択された特徴の検出に基づいてビデオを制御するよう構成されている。リソースマネージャは、ユーザデバイスと別のユーザデバイスとの間の通信チャネル、並びに/又は、ユーザデバイス及び/若しくは別のユーザデバイスの1以上のリソース、に関する情報を受信し、受信された情報に基づいて、ビデオコントローラによる上記選択を制御するよう構成されている。

##### 【0005】

ネットワークに接続するためのネットワークインタフェース及び1以上のプロセッサを備えたユーザデバイスも開示される。1以上のプロセッサは、画像キャプチャデバイスから受信された、1人以上のユーザのビデオを、ネットワークを介して、別のユーザデバイスに送信するよう構成されている。1以上のプロセッサは、複数の視覚的なユーザ特徴から特徴を選択し、第1の時間及び第2の時間のそれぞれにおいて、選択された特徴の検出に基づいて、第1のセットの境界データ及び第2のセットの境界データを生成するようさらに構成されている。1以上のプロセッサは、動的モデルを使用して、第1のセットの境界データ及び第2のセットの境界データに基づいて過渡データ(transition data)を生成し、選択された特徴を追跡するために、過渡データに基づいてビデオを制御するようさらに構成されている。

##### 【0006】

これらに対応する方法、及び、これらに対応する方法の各々に合わせて構成されている実行可能なコードを含むそれぞれのコンピュータプログラム製品も開示される。

#### 【図面の簡単な説明】

##### 【0007】

本主題の理解のために、且つ、同主題がどのように実施され得るかを示すために、例として図面が参照される。

【図1】通信システムの概略図。

【図2】ユーザデバイスの概略ブロック図。

【図3】ビデオ制御技術の機能図。

【図4A】ユーザデバイスとインタラクトしているユーザの概略図。

【図4B】ユーザデバイスとインタラクトしているユーザの概略図。

【図5】放射線パターンの概略図。

【図6A】ユーザの概略正面図。

【図6B】ユーザの概略見下げ図。

【図6C】ユーザの概略側面図。

【図 7 A】ユーザの検出された骨格ポイントの概略図。

【図 7 B】ユーザの推定された身体パーツの概略図。

【図 8】ビデオを制御する方法のフローチャート。

【図 9】時間の経過に伴ってビデオを制御する方法のフローチャート。

【発明を実施するための形態】

【 0 0 0 8 】

ビデオ通話（１対１又はマルチパーティ）中、室内のユーザの位置が、奥行き検出手段（奥行き検出器又は同様のもの）を用いて検出され、その検出に基づいて当該ユーザを追跡するために、他の１人以上のユーザに送信される、ビデオ通話のビデオが制御される技術が、本明細書で開示される。より詳細には、当該ユーザのそれぞれの視覚的な特徴（当該ユーザの身体パーツ等）が、そのようにして検出され、それらの特徴（例えば身体パーツ）から選択された特徴（全てである必要はない）を追跡するために、ビデオが制御される。特定の視覚的な特徴が追跡されるか否かは、通信チャネル状態（例えばチャネル帯域幅）、並びに／又は、近端デバイス及び／若しくは遠端デバイスのリソースに依拠する。例えば、より低いチャネル帯域幅及び／又は遠端デバイスのより小さなスクリーンサイズの場合には、（例えば）より少ない特徴（例えば身体パーツ）が追跡され得、より高いチャネル帯域幅及び／又は遠端デバイスのより大きなスクリーンサイズの場合には、より多くの特徴（例えば身体パーツ）が追跡され得る。

【 0 0 0 9 】

図 1 は、第 1 のユーザデバイス 1 0 4 に関連付けられている第 1 のユーザ 1 0 2 a（「ユーザ A」）、第 2 のユーザデバイス 1 1 0 に関連付けられている第 2 のユーザ 1 0 8（「ユーザ B」）、第 3 のユーザデバイス 1 1 4 に関連付けられている第 3 のユーザ 1 1 2（「ユーザ C」）、及び第 1 のユーザの近くにいる第 4 のユーザ 1 0 2 b（「ユーザ D」）を含む通信システム 1 0 0 を示している。他の実施形態では、通信システム 1 0 0 は、任意の数のユーザ及び関連ユーザデバイスを含み得る。ユーザデバイス 1 0 4、1 1 0、及び 1 1 4 は、通信システム 1 0 0 において、ネットワーク 1 0 6 を介して通信することができ、それにより、ユーザ 1 0 2 a、1 0 2 b、1 0 8、及び 1 1 2 は、ネットワーク 1 0 6 を介して、他のユーザの各々と通信することが可能になる。図 1 に示される通信システム 1 0 0 は、パケットベースの通信システムであるが、他のタイプの通信システムが使用されてもよい。ネットワーク 1 0 6 は、例えば、インターネットであり得る。ユーザデバイス 1 0 4 及び 1 1 0 の各々は、例えば、携帯電話機、タブレット、ラップトップ、パーソナルコンピュータ（「PC」）（例えば、Windows（登録商標）PC、Mac OS（登録商標）PC、及びLinux（登録商標）PCを含む）、ゲームコンソール若しくは他のゲームデバイス、テレビジョン、携帯情報端末（「PDA」）、又はネットワーク 1 0 6 に接続することができる他の組み込みデバイスであり得る。ユーザデバイス 1 0 4 は、ユーザ 1 0 2 a、1 0 2 b から情報を受け取り、ユーザ 1 0 2 a、1 0 2 b に情報を出力するよう構成されている。ユーザデバイス 1 0 4 は、ディスプレイ及びスピーカ等の出力手段を備える。ユーザデバイス 1 0 4 は、キーボード、タッチスクリーン、オーディオ信号を受信するためのマイクロフォン、及び／又は、一連のビデオフレーム（画像）から構成されるビデオ信号の画像をキャプチャするためのカメラ等の入力手段も備える。ユーザデバイス 1 0 4 は、ネットワーク 1 0 6 に接続されている。

【 0 0 1 0 】

ユーザデバイス 1 0 4 は、通信システム 1 0 0 に関連付けられているソフトウェアプロバイダにより提供された通信クライアントのインスタンスを実行する。通信クライアントは、ユーザデバイス 1 0 4 内のローカルプロセッサ上で実行されるソフトウェアプログラムである。通信クライアントは、ユーザデバイス 1 0 4 が通信システム 1 0 0 を介してデータを送受信するためにユーザデバイス 1 0 4 において必要とされる処理を実行する。

【 0 0 1 1 】

ユーザデバイス 1 1 0 は、ユーザデバイス 1 0 4 に対応するものであり、ユーザデバイス 1 0 4 において実行される通信クライアントに対応する通信クライアントを、ローカル

10

20

30

40

50

プロセッサ上で実行する。ユーザデバイス 110 における通信クライアントは、ユーザデバイス 104 における通信クライアントが、ユーザ 102a 及び 102b がネットワーク 106 を介して通信することを可能にするために必要とされる処理を実行するのと同様に、ユーザ 108 がネットワーク 106 を介して通信することを可能にするために必要とされる処理を実行する。ユーザデバイス 114 は、ユーザデバイス 104 に対応するものであり、ユーザデバイス 104 において実行される通信クライアントに対応する通信クライアントを、ローカルプロセッサ上で実行する。ユーザデバイス 114 における通信クライアントは、ユーザデバイス 104 における通信クライアントが、ユーザ 102a 及び 102b がネットワーク 106 を介して通信することを可能にするために必要とされる処理を実行するのと同様に、ユーザ 112 がネットワーク 106 を介して通信することを可能にするために必要とされる処理を実行する。ユーザデバイス 104、110、及び 114 は、通信システム 100 におけるエンドポイントである。図 1 は、明瞭さのために、4 人のユーザ (102a、102b、108、及び 112) 及び 3 つのユーザデバイス (104、110、及び 114) しか示していないが、より多くのユーザ及びユーザデバイスが、通信システム 100 に含まれてもよく、それぞれのユーザデバイス上で実行されるそれぞれの通信クライアントを使用して、通信システム 100 を介して通信してもよい。

#### 【0012】

図 2 は、通信システム 100 を介して通信するための通信クライアントインスタンス 206 が実行されるユーザデバイス 104 の詳細な図を示している。ユーザデバイス 104 は、プロセッサすなわち「処理モジュール」202 を備える。プロセッサ 202 は、1 以上の中央処理装置 (「CPU」) を含む。プロセッサ 202 には、タッチスクリーンとして実装され得るディスプレイ 208、オーディオ信号を出力するためのスピーカ (又は「ラウドスピーカ」) 210、及び、赤外線 (IR) プロジェクタ等の、非可視放射線を出力するための放射線プロジェクタ 224 といった出力デバイス; オーディオ信号を受信するためのマイクロフォン 212、画像データを受信するためのカメラ 216、キーパッド 218、及び、例えば、赤外線信号の画像を受信するための IR カメラといった、非可視放射線信号の非可視画像を受信するための放射線センサアレイ 226 といった入力デバイス; データを記憶するためのメモリ 214; 及び、ネットワーク 106 と通信するためのモデムといったネットワークインタフェース 220 が接続される。スピーカ 210 は、オーディオ信号処理モジュール 209 を介してプロセッサ 202 に接続される。ユーザデバイス 104 は、図 2 に示される要素以外の他の要素を備えてもよい。ディスプレイ 208、スピーカ 210、マイクロフォン 212、メモリ 214、カメラ 216、キーパッド 218、及びネットワークインタフェース 220 は、ユーザデバイス 104 に統合され得る。代替的に、ディスプレイ 208、スピーカ 210、マイクロフォン 212、メモリ 214、カメラ 216、キーパッド 218、及びネットワークインタフェース 220 のうちの 1 以上は、ユーザデバイス 104 に統合されなくてもよく、それぞれのインタフェースを介してプロセッサ 202 に接続されてもよい。そのようなインタフェースの一例が USB インタフェースである。例えば、ゲームコンソールの形態のユーザデバイスの場合、カメラ 216、プロジェクタ 224、センサ 226、ディスプレイ 208、マイクロフォン 212、キーパッド 218 (例えばゲームコントローラ) は、ゲームコンソールに統合されず、それぞれのインタフェースを介してプロセッサ 202 に接続され得るのに対し、ネットワークインタフェース 220、メモリ 214、及びプロセッサ 202 は、ゲームコンソールに統合され得る。ネットワークインタフェース 220 を介するネットワーク 106 へのユーザデバイス 104 の接続が、無線接続である場合、ネットワークインタフェース 220 は、ネットワーク 106 に信号を無線で送信するとともにネットワーク 106 から信号を無線で受信するためのアンテナを含み得る。

#### 【0013】

プロジェクタ 224 及びセンサ 226 は、3 次元 (「3D」) の非可視放射線データをキャプチャするための奥行き検出器 222 を構成する。この実施形態では、プロジェクタ 224 は、センサ 226 の前方に、センサ 226 により検出可能な放射線パターンを投射

10

20

30

40

50

する。センサ 226 からのセンサデータは、(図 5 及び図 6 A ~ 図 6 C を参照して以下でより詳細に説明するように) 検出された放射線パターンにおける歪みに基づいて 3D 画像を構築するために使用される。奥行き検出器 222 及びカメラ 216 は、可能であればユーザデバイス 104 とは別の電源を有する、ユーザデバイス 104 に外付けされる 1 つのユニット内に一緒に収容され、適切なインタフェース (例えば、USB 又は USB ベース) を介してプロセッサ 202 に接続され得る。そのようなユニットの一例が、Microsoft Kinect Sensor (登録商標) である。

#### 【0014】

図 2 はまた、プロセッサ 202 上で実行されるオペレーティングシステム (「OS」) 204 も示している。OS 204 の上で、通信システム 100 のクライアントインスタンス 206 のソフトウェアが実行される。オペレーティングシステム 204 は、ユーザデバイス 104 のハードウェアリソースを管理し、ネットワークインタフェース 220 を介してネットワーク 106 との間で通信されるデータを処理する。クライアント 206 は、オペレーティングシステム 204 と通信し、通信システム上の接続を管理する。クライアント 206 は、情報をユーザに 102 に提示するとともに情報をユーザ 102 から受け取るために使用されるクライアントユーザインタフェースを有する。このように、クライアント 206 は、ユーザ 102 が通信システム 100 を介して通信することを可能にするために必要とされる処理を実行する。

#### 【0015】

次に図 3 を参照して、ネットワークを介してビデオを送信する方法について説明する。図 3 は、ユーザデバイス 104 の一部の機能図である。

#### 【0016】

図 3 に示されるように、ユーザデバイス 104 は、エンコーダ 308 (例えば、H264 エンコーダ) を備える。これは、ネットワーク 106 を介する送信に先立って符号化するためにプロセッサ 202 によりビデオデータが供給されるハードウェアエンコーダであってもよいし、(例えば) プロセッサ 202 上で実行されるソフトウェアにより実装されるソフトウェアエンコーダであってもよいし、これらの組合せであってもよい。ユーザデバイスは、コントローラ 302 (例えば、クライアント 106 のソフトウェアの一部として実装される)、カメラ 216、奥行き検出器 222、及び奥行きプロセッサ 310 をさらに備える。コントローラ 302 は、リソースマネージャ 306 及びビデオ信号プロセッサ 304 を含む。カメラ 216 は、可視スペクトル (すなわち、人間の目に見える) における、ユーザ 102 a、102 b のビデオ (図 3 において「近端ビデオ」とラベリングされている) の画像をキャプチャし、キャプチャした画像を、ビデオ信号プロセッサの第 1 の入力部に供給する。センサ 226 は、非可視スペクトル (すなわち、人間の目には見えない) における画像をキャプチャし、キャプチャした画像を、奥行きプロセッサ 310 の入力部に供給する。奥行きプロセッサ 310 は、ビデオ信号プロセッサの第 2 の入力部に接続されている出力部を有する。プロジェクト 224 は、センサ 226 により検出可能な非可視放射線を、ユーザ 102 a、102 b に向けて、センサ 226 の前方に投射する。コントローラ 302、エンコーダ 308、及びセンサプロセッサ 310 は、ビデオ処理システム 300 を構成する。

#### 【0017】

リソースマネージャ 306 は、第 1 の入力部、第 2 の入力部、第 3 の入力部、及び出力部を有する。リソースマネージャ 306 の第 1 の入力部は、ユーザデバイス 104 とネットワーク 106 の 1 以上の他のユーザデバイス (例えば、110、114) との間の 1 以上の通信チャネルに関する情報を受信するよう構成されている。リソースマネージャ 306 の第 2 の入力部は、ユーザデバイス 104 及び / 又は 1 以上の他のユーザデバイス (例えば、110、114) のリソースに関する情報を受信するよう構成されている。リソースマネージャ 306 の第 3 の入力部は、ビデオ信号プロセッサ 304 の出力部に接続されている。

#### 【0018】

10

20

30

40

50

リソースマネージャ 306 の出力部は、ビデオ信号プロセッサ 304 の第 3 の入力部に接続されている。ビデオ信号プロセッサ 304 は、エンコーダ 308 の入力部に接続されている出力部を有する。エンコーダは、ネットワーク 106 を介して第 2 のユーザデバイス 110 及び第 3 のユーザデバイスのうちの少なくとも 1 つに送信するためにネットワークインタフェース 220 に供給される前にユーザデバイス 104 においてさらなる処理（例えばパケット化）が実行され得る符号化されたビデオデータを供給するよう構成されている。

【0019】

センサプロセッサ 310 は、センサ 226 によりキャプチャされセンサ 226 から受信されたセンサデータを処理して、センサ 226 の視野内の各ユーザについてそれぞれの複数の視覚的なユーザ特徴を検出するよう動作可能である。

10

【0020】

ビデオ信号プロセッサ 304 は、リソースマネージャ 306 から出力された情報に基づいて、複数の視覚的なユーザ特徴から特徴を選択し、センサプロセッサ 310 による選択された特徴の検出に基づいて符号化するためにエンコーダ 308 に供給されるビデオを制御するよう構成されている。ビデオは、選択された特徴を追跡するためにそのように制御される、すなわち、ビデオにより、選択された特徴が経時的に追跡され、最終的に、検出された特徴がビデオ内に見えるようにする。

【0021】

次に、これについて図 4 A 及び図 4 B を参照して説明する。図 4 A 及び図 4 B は、ユーザデバイス 104（この実施形態ではゲームコンソールとして示されている）の近くににいるユーザ 102（例えば、102 a、102 b）を示している。ユーザデバイスは、ディスプレイ 208、カメラ 216、プロジェクタ 224、及びセンサ 226 に接続されており、この実施形態では、これらの全てが、ユーザデバイス 104 に外付けされている。図 4 A において、カメラ 216 によりキャプチャされた、第 1 の可視領域 402 a のビデオ - ユーザ 102 の下側部分（例えば、足、脚等）を含まずユーザ 102 の上側部分（例えば、頭、肩等）を含むよう制限されている - が、符号化して最終的にネットワーク 106 を介して別のユーザ（例えば、112、108）に送信するために、エンコーダ 308 に供給される。図 4 A に示されるように、領域 402 a のビデオのバージョンは、別のユーザに送信され、任意的に、別のユーザから受信されたビデオの上に重ねられてディスプレイ 208 上に表示される。図 4 B において、カメラ 216 によりキャプチャされた、第 2 の可視領域 402 b のビデオ - ユーザ 102 の上側部分及びユーザ 102 の下側部分の両方を含む - が、符号化して最終的にネットワーク 106 を介して別のユーザに送信するために、エンコーダ 308 に供給される。図 4 B に示されるように、第 2 の領域 402 b のビデオのバージョンは、別のユーザに送信され、任意的に、別のユーザから受信されたビデオの上に重ねられてディスプレイ 208 上に表示される。

20

30

【0022】

第 1 の可視領域 402 a 及び第 2 の可視領域 402 b は、以下で説明するように、リソースマネージャ 306 から受信されたデータとセンサプロセッサ 310 から受信されたデータとに基づいてビデオ制御モジュールにより決定されたそれぞれのサイズ及び位置を有する。

40

【0023】

図 5 は、プロジェクタ 224 により放射された放射線パターンを示している。図 5 に示されるように、放射線パターンは、系統的に不均質であり、交互強度（alternating intensity）を有する複数の系統的に配された領域を含む。詳細には、図 5 の放射線パターンは、放射線ドットの実質的に一様なアレイを含む。放射線パターンは、この実施形態では赤外線（IR）放射線パターンであり、さらに、センサ 226 により検出可能である。図 5 の放射線パターンは、例示的なものであり、他の放射線パターンの使用も想定されている。

【0024】

50



この放射線パターン 500 は、プロジェクタ 224 により、センサ 226 の前方に投射される。センサ 226 は、センサ 226 の視野内に投射されたその非可視放射線パターンの画像をキャプチャする。これらの画像は、センサ 226 の視野内のユーザの奥行きを算出する（ユーザの 3 次元表現を実効的に構築する）ためにセンサプロセッサ 310 により処理され、これにより、異なるユーザと、それらの異なるそれぞれの身体パーツと、の認識が可能になる。

#### 【0025】

図 6 B は、ユーザデバイス 104 の近くにいるユーザ 102 の見下げ図を示しており、図 6 A は、（カメラ 216 及び検出器 222 のセンサ 226 により確認される）対応する正面図であり、図 6 C は、対応する側面図である。図示されるように、ユーザ 102 は、左手を検出器 222 に向けて伸ばしている姿勢をとっている（もちろん、ユーザは他の形の姿勢をとることもできる）。ユーザの頭が、ユーザの胴の上に突出しており、ユーザの胴は、ユーザの右腕の前にある。放射線パターン 500 が、プロジェクタ 224 により、ユーザ上に投射されている。

#### 【0026】

図 6 A に示されるように、ユーザ 102 は、したがって、センサ 226 により検出される投射された放射線パターンを歪めるように作用する形状を有する。ここで、プロジェクタからより遠くにあるユーザのパーツ上に投射された放射線パターンの部分（すなわち、この場合には、放射線パターンのドットが、より離れるようになる）は、プロジェクタからより近くにあるユーザのパーツ上に投射された放射線パターンの部分（すなわち、この場合には、放射線パターンのドットが、より離れないようになる）と比べて、実効的に広げられる。広がりスケーリングの程度は、プロジェクタからの距離に応じ、ユーザのかなり後ろのオブジェクト上に投射された放射線の部分は、センサ 226 にとって実効的に不可視である。放射線パターン 500 は、系統的に不均質であるので、ユーザの形状による放射線パターンの歪みを使用して、センサプロセッサ 310 が、センサ 226 によりキャプチャされた歪んだ放射線パターンの画像を処理することにより、その形状を識別することができる（例えば、ユーザの頭、左手、右腕、胴等を識別することができる）。例えば、センサからユーザのある領域までの距離は、ユーザのその領域内における検出された放射線パターンのドット群の距離を測定することにより求めることができる。

#### 【0027】

図 5 及び図 6 A においては、放射線パターンが、人間の目に見えるものとして示されているが、これは理解を助けるために過ぎず、ユーザ 102 上に投射される放射線パターンは、実際には人間の目に見えない。

#### 【0028】

次に図 7 A 及び図 7 B を参照して説明するように、センサ 226 によりキャプチャされた非可視放射線パターン 500 の画像（センサデータ）は、センサプロセッサ 310 により処理される。この処理は、骨格検出処理を含む。骨格検出処理は、当技術分野において知られており、例えば、Microsoft Xbox 360（登録商標）（時として、Microsoft Kinect Sensor（登録商標）とともに使用される）において現在実装されており、その結果が、ソフトウェア開発者による使用のために、アプリケーションプログラミングインタフェース（API）を介して利用可能になっている。

#### 【0029】

センサプロセッサ 310 は、センサ 226 からセンサデータを受信し、センサデータを処理して、センサ 226 の視野内の複数のユーザ（例えば、102 a、102 b）を判別し、当技術分野において知られている骨格検出技術を使用して、各ユーザについて、それぞれの複数の骨格ポイントを識別する。各骨格ポイントは、ビデオ内の対応する人間の関節のおおよその位置を表す。

#### 【0030】

詳細には、この実施形態では、センサプロセッサ 310 は、センサ 226 の視野内の各

10

20

30

40

50

ユーザについて、20個のそれぞれの骨格ポイントを検出する。各骨格ポイントは、人間の20個の認識される関節のうちの1つに対応するものであり、1人以上のユーザは、センサ226の視野内で移動するので、これらの認識される関節の各々は、空間及び時間において変化する。任意の瞬間時におけるこれらの関節の位置が、センサ226により検出された、1以上のユーザのそれぞれの3次元形状に基づいて算出される。骨格ポイントはまた、追跡状態(tracking state)を有する。この状態は、関節がはっきりと可視である場合には「追跡されている」であり、関節がはっきりとは可視ではないがセンサプロセッサ310がその位置を推測している場合には「推定されている」であり、例えば、座位モード追跡(seated-mode tracking)における下側関節の場合には(すなわち、ユーザが座っていると検出され、下側関節はセンサプロセッサ310により追跡されていない場合には)「追跡されていない」であり得る。

10

【0031】

各骨格ポイントは、対応する関節が正しく検出されている尤度を示すそれぞれの信頼値とともに提供され得る。所定の閾値未満の信頼値を有するポイントは、ビデオモジュール304による処理から除外され得る。

【0032】

これらの20個の骨格ポイントが図7Aに示されており、対応する人間の関節が表1に示されている。

## 【表 1】

表 1－骨格ポイント

骨格ポイントの名称：	対応する人間の関節：	ラベル：
AnkleLeft	左足首	722b
AnkleRight	右足首	722a
ElbowLeft	左肘	706b
ElbowRight	右肘	706a
FootLeft	左足	724b
FootRight	右足	724a
HandLeft	左手	702b
HandRight	右手	702a
Head	頭	710
HipCenter	両臀部間の中央	716
HipLeft	左臀部	718b
HipRight	右臀部	718a
KneeLeft	左膝	720b
KneeRight	右膝	720a
ShoulderCenter	両肩間の中央	712
ShoulderLeft	左肩	708b
ShoulderRight	右肩	708a
Spine	脊柱	714
WristLeft	左手首	704b
WristRight	右手首	704a

## 【 0 0 3 3 】

特定の時間における、センサプロセッサによりレポートされた骨格ポイントの位置が、その特定の時間における、ビデオのフレーム（画像）内での対応する人間の関節の位置に対応するように、カメラ 216 からのビデオと骨格ポイントとが関連付けられる。センサプロセッサ 310 は、これらの検出された骨格ポイントを、骨格ポイントデータとして、ビデオコントローラ 304 による使用のためにビデオコントローラ 304 に供給する。ビデオデータのフレームごとに、センサプロセッサ 310 により供給される骨格ポイントデータは、例えば、ビデオフレームサイズの点で制限される座標系のデカルト座標（ $x, y$ ）として表現される位置等の、そのフレーム内での骨格ポイントの位置を含む。

## 【 0 0 3 4 】

10

20

30

40

50

ビデオコントローラ 304 は、1人以上のユーザ（102a、102b）に関する検出された骨格ポイントを受信し、それらの骨格ポイントから、そのユーザの複数の視覚的な特徴（又は、それらのユーザのそれぞれの複数の視覚的な特徴）を判別するよう構成されている。この実施形態では、視覚的なユーザ特徴は、人間の身体パーツの形態をとる。身体パーツは、ビデオコントローラにより検出される。身体パーツの各々は、ビデオプロセッサ 310 により提供された 1 以上の骨格ポイントから推定することを介して検出され、身体パーツの各々は、カメラ 216 からのビデオの対応するビデオフレーム内の領域に対応する（すなわち、前述した座標系内の領域として規定される）。図 7B は、図 7A の骨格ポイントに基づいて検出された身体パーツを示している。

【表 2】

10

表 2－身体パーツ（視覚的な特徴）

身体パーツ（視覚的な特徴）の名前	図 7B における検出された特徴：
頭	750
肩	752
脊柱中心部（脊柱の上側部分）	756
脊柱下部（脊柱の下側部分）	758
脊柱全体（脊柱の全て）	760
臀部	762
肘	754a, 754b
脚	764a, 764b
足	766a, 766b

20

30

【0035】

これらの視覚的な特徴は、実際に見ることができ識別することができる、ユーザの身体の特徴を表すという意味で視覚的であるということに留意されたい。しかしながら、この実施形態では、これらの視覚的な特徴は、ビデオコントローラによっては「確認されない」（例えば、これらの視覚的な特徴は、カメラ 216 によりキャプチャされたビデオデータ内では検出されない）。そうではなく、ビデオコントローラは、（例えばカメラ 216 からのビデオのフレームの処理に基づいてではなく、）センサプロセッサ 310 により提供された 20 個の骨格ポイントの配置から、カメラ 216 からのビデオのフレーム内のこれらの特徴の（おおよその）相対位置、形状、及びサイズを推定する（例えば、各身体パーツを、その身体パーツに関連する骨格ポイントの検出された配置から算出される位置及びサイズ（並びに任意的に向き）を有する矩形（又は同様の形状）として近似することにより）。

40

【0036】

次に図 8 を参照して、検出されたユーザ特徴（この実施形態では身体パーツ）に基づいて、ネットワークを介して送信されるビデオを制御する方法 800 について説明する。図 8 は、方法 800 のフローチャートを左側に示し、各方法ステップの対応する視覚的な表現を右側に示している。

【0037】

この実施形態では、方法 800 は、クライアント 106 の一部として - 具体的にはコントローラ 302 により - アルゴリズム的に実施される。

50

## 【 0 0 3 8 】

方法 8 0 0 は、第 1 のユーザデバイス（近端デバイス）1 0 4 及び第 2 のユーザデバイス（遠端デバイス）1 1 0 を使用して、これらのユーザデバイスのユーザの間で行われるリアルタイムビデオ通話のコンテキストにおいて説明される。

## 【 0 0 3 9 】

方法 8 0 0 は、以下の身体領域を考慮する：

- 領域 1：頭、肩
- 領域 2：頭、肩、脊柱中心部、肘
- 領域 3：頭、肩、脊柱中心部、肘、脊柱下部、臀部
- 領域 4：頭、肩、脊柱全体、肘、臀部、脚、足

10

## 【 0 0 4 0 】

ビデオ通話の開始時に、近端デバイス 1 0 4 のクライアント 1 0 6 のリソースマネージャ 3 0 6 は、近端デバイス 1 0 4 により受信された情報に基づいて、（遠端デバイス 1 1 0 に送信されるビデオのために使用される）ビデオ解像度を決定する。例えば、この情報は、以下のうちの 1 以上に関する情報を含み得る：

- 近端デバイスと遠端デバイスとの間の通信チャネルのチャネル帯域幅といったネットワークリソース（ビデオは、近端デバイスから、そのチャネルを介して、遠端デバイスに送信される）

- 例えば、スクリーンサイズ（より多くの画素を有するスクリーンは、より少ない画素を有するスクリーンよりも多くのスクリーンリソースを有する）；遠端デバイスのスクリーンの横方向の長さ（例えば、4：3、16：9 等）；遠端デバイスのスクリーンのアスペクト比（例えば、4：3、16：9 等）；遠端デバイスの 1 以上のプロセッサの処理リソース（例えば、メガヘルツ、ギガヘルツ等を単位とする利用可能なプロセッササイクル）；遠端デバイスのメモリリソース（例えば、メガバイト、ギガバイト等を単位とする利用可能なメモリ）；ビデオ復号能力といった、遠端デバイスから受信される遠端デバイスのリソース

20

- 例えば、近端デバイスの 1 以上のプロセッサの処理リソース（例えば、メガヘルツ、ギガヘルツ等を単位とする利用可能なプロセッササイクル）；近端デバイスのメモリリソース（例えば、メガバイト、ギガバイト等を単位とする利用可能なメモリ）；ビデオ符号化能力といった、近端デバイスのリソース

30

## 【 0 0 4 1 】

ビデオ解像度は、遠端デバイス 1 1 0 とのネゴシエーションを介して少なくとも部分的に決定され得る（例えば、近端デバイス 1 0 4 のリソースマネージャ 3 0 6 が、遠端デバイスに対して、遠端デバイスのリソースに関する情報を要求することにより）。

## 【 0 0 4 2 】

ビデオ通話中、近端デバイス 1 0 4 のリソースマネージャ 3 0 6 は、利用可能なリソース（例えば利用可能な帯域幅）をモニタリングし、遠端デバイス 1 1 0 に送信されているビデオのための解像度を増大又は低減させる決定を行い、その決定をビデオコントローラ 3 0 4 に通信する。したがって、例えば、チャネル帯域幅の変動が生じていることに起因して、すなわち、近端デバイス及び遠端デバイスのうち的一方又は双方が、信頼できない無線接続（例えば W i F i（登録商標）接続）を介してネットワーク 1 0 6 に接続されていることから、ビデオ解像度はビデオ通話中に動的に変わり得る。

40

## 【 0 0 4 3 】

特徴を選択することには、受信された情報により、より良いチャネル品質及び／又はより多くのデバイスリソースが示されている場合には、より多くの特徴を選択すること、及び、受信された情報により、より悪いチャネル品質及び／又はより少ないデバイスリソースが示されている場合には、より少ない特徴を選択することが含まれる。例えば、より低い帯域幅及び／又はより小さなスクリーンサイズの場合には、より低い解像度が、リソースマネージャ 3 0 6 により決定される（これは、より少ない身体パーツを含む領域の選択をもたらす）。より高い帯域幅及び／又はより大きなスクリーンサイズの場合には、上記

50

のより低い帯域幅及び／又はより小さなスクリーンサイズの場合に決定される解像度よりも高い解像度が、リソースマネージャ 306 により決定される（これは、より多くの身体パーツを含む領域の選択をもたらす）。

#### 【0044】

この実施形態では、ビデオは、1920×1080 画像という固定解像度でカメラ 216 からキャプチャされる。しかしながら、以下で説明するように、キャプチャされたビデオは、送信の前にクロップされ得る（クロッピングにより、ビデオの外側部分の画像が除去される）。すなわち、キャプチャされたビデオデータのうち選択された部分のみ（選択された可視領域からキャプチャされる）が、符号化してその後遠端デバイスに送信するためにエンコーダ 308 に供給される。したがって、クロップされたビデオデータを供給することは、規定された領域の外側のビデオデータはエンコーダ 308 に供給されないの

10

#### 【0045】

ステップ S802 において、骨格ポイントデータが、センサプロセッサ 310 から、ビデオコントローラにより受信される。図 8 の右側の例示図において、位置しているユーザ 102a 及び 102b（センサ 226 の視野内の両者）に関する骨格データが示されている。しかしながら、方法 800 は、明らかなように、任意の数のユーザ（1人以上）に関する骨格データの受信に適用することができる。

#### 【0046】

ステップ S804 において、ビデオコントローラ 304 は、リソースマネージャ 306 から受信された決定された解像度に基づいて、センサ 216 の視野内の各ユーザについて、それぞれの身体領域を選択する。ビデオコントローラは、このビデオ解像度に応じて、以下のように、表 2 の可能な複数の身体領域から、それぞれの身体領域を選択する：

20

- 1920×1080 の解像度：領域 4
- 1280×1024 以上の解像度：領域 3
- 640×480 以上の解像度：領域 2
- 他の解像度：領域 1

#### 【0047】

代替的に以下のものであってもよい：

- 1920 の幅の解像度：領域 4
- 1280 以上の幅の解像度：領域 3
- 640 以上の幅の解像度：領域 2
- 他の解像度：領域 1

30

#### 【0048】

特徴を選択することには、より高い解像度（より良いチャネル状態及び／又はより多くのデバイスリソース）の場合には、より多くの特徴を選択すること、及び、より低い解像度（より悪いチャネル状態及び／又はより少ないデバイスリソース）の場合には、より少ない特徴を選択することが含まれる。

#### 【0049】

センサ 226 の視野内の各ユーザ 102a、102b について、選択された領域が、受信された骨格データに基づいて検出される（すなわち、選択された領域のそれぞれの身体パーツが検出される）。次いで、ビデオが、符号化されて送信される前に、（以下で説明するように）検出された身体パーツに基づいてクロップされる（例えば、領域 1 に関しては、ビデオは、ユーザ 102a の頭、ユーザ 102b の頭、ユーザ 102a の肩、及びユーザ 102b の肩（領域 1 内には頭及び肩だけが含まれるので、他の身体パーツではない）の検出に基づいてクロップされる。したがって、全ての骨格ポイントが、クロッピングに必ずしも寄与するわけではない（なぜならば、これらの一部は、選択された領域の部分ではない身体パーツを規定することだけに役立ち得るからである）。すなわち、ビデオの制御は、複数の検出された骨格ポイントから選択された骨格ポイントに基づき得るものであり、複数の検出された骨格ポイントのうちの他の骨格ポイントには基づかない。

40

50

## 【 0 0 5 0 】

例えば、図 8 の右側の例示図において、領域 2（頭、肩、脊柱中心部、肘）が選択されており、したがって、ビデオ制御は、ユーザ 1 0 2 a に関する、領域 2 における検出された特徴（図 8 の 5 5 0 a）と、ユーザ 1 0 2 b に関する、領域 2 における検出された特徴（図 8 の 5 5 0 b）と、の両方に基づく。

## 【 0 0 5 1 】

（この実施形態では、決定された解像度に従ってビデオを選択的にクロップすることを含む）ビデオ制御の一部として、ステップ S 8 0 6 において、ビデオコントローラは、選択された領域の身体パーツの検出に基づいて、境界データを生成する。境界データは、矩形境界 5 5 2（クロッピング矩形）を規定する。クロッピング矩形は、全ての追跡されるユーザに関するそれぞれの検出された領域の和集合（union）から形成される。

10

## 【 0 0 5 2 】

ステップ S 8 0 8 において、ビデオコントローラ 3 0 4 は、予め定められたアスペクト比（例えば、4 : 3、1 6 : 9 等といった、遠端デバイス 1 1 0 のディスプレイのアスペクト比）に基づいて、生成された境界データを変更する。これにより、クロッピング矩形が、予め定められたアスペクト比に合わせて調整される。調整されたクロッピング矩形（外接矩形）が、図 8 の右側に 5 5 4 として示されている（この実施形態では、1 6 : 9 のアスペクト比を有する）。

## 【 0 0 5 3 】

ステップ S 8 1 0 において、ビデオが、以下でより詳細に説明するように、調整された矩形に基づいてクロップされる。次いで、クロップされたビデオが、エンコーダ 3 0 8 により符号化される前に、出力解像度にスケーリングされる（例えば、送信先の遠端デバイスのスクリーンの解像度に合わせられる）。

20

## 【 0 0 5 4 】

方法 8 0 0 は、ビデオデータのフレームごとに実行される。外接矩形により、フレーム間でユーザ特徴が追跡されるので（ユーザ特徴は、ユーザが移動するにつれて移動し得る、あるいは、ユーザ特徴は、例えば 2 人のユーザが離れるにつれ離れ得る）、クロッピング矩形も移動する。

## 【 0 0 5 5 】

しかしながら、ビデオデータの各フレームを、そのフレームに関して決定された調整された矩形に合わせて単純にクロップする（すなわち、そのフレームのうち、調整された矩形により規定される部分からのビデオデータのみを供給する）以外に、過渡データが、1 以上の以前のフレームに関して算出されたそれぞれの調整された矩形と、現フレームに関して算出された調整された矩形と、に基づいて、生成される。過渡データは、弾性バネモデルに基づいて生成される。

30

## 【 0 0 5 6 】

諸実施形態において、弾性バネモデルは、以下のように定義され得る：

## 【 数 1 】

$$m * \frac{d^2x}{dt^2} = -k * x - D * \frac{dx}{dt}$$

40

ここで、m（「質量」）、k（「剛性（stiffness）」）、及び D（「減衰（damping）」）は、設定可能な定数であり、x（変位）及び t（時間）は、変数である。すなわち、このモデルでは、遷移の加速度が、その遷移の変位と速度との重み付け和に比例する。

## 【 0 0 5 7 】

クロッピング矩形は、弾性バネモデルに従って移動する。これは、フレーム間の遷移を平滑化させ、ぎくしゃくしたビデオを防ぐ。これはまた、次の理由から符号化の効率を増大させる。弾性モデルは、クロッピング矩形の移動を実効的に「抑制させる（dampen）」ので、弾性モデルは、隣接フレーム間の差を少なくさせ、これは、明らかなように、より効率的な差分符号化をもたらす。

50

## 【 0 0 5 8 】

次に、これについて図 9 を参照して説明する。図 9 は、時間の経過に伴って、遠端ユーザに送信されるビデオを制御する方法 9 0 0 のフローチャートと、右側に方法 9 0 0 の例示的な付随表現と、を示している。

## 【 0 0 5 9 】

ステップ S 9 0 2 において、キャプチャされたビデオの一連のビデオフレームにおける現ビデオフレーム 9 5 6 に関して、ビデオコントローラは、第 1 の時間及び 1 以上の第 2 の時間のそれぞれにおいて、選択された視覚的なユーザ特徴（この実施形態では身体パーツ）の検出に基づいて、（例えば、上述したように、予め定められたアスペクト比を有する外接矩形を算出することにより）第 1 のセットの境界データ及び 1 以上の第 2 のセットの境界データを生成する。第 1 の時間は、現フレーム 9 5 6 の時間であり、1 以上の第 2 の時間は、以前のフレームの 1 以上の時間である。以前のフレームに関する境界データは、後のフレームに関する境界データの前に生成され得る（例えば、フレーム単位で）。したがって、ステップ S 9 0 2 は、複数のビデオフレームに及ぶ期間にわたって生じ得る。

## 【 0 0 6 0 】

第 1 の時間  $t_1$  及び第 2 の時間  $t_2$  それぞれにおける 2 つの例示的な外接矩形 9 5 0、9 5 2 が、図 9 の右側に示されている。 $t_2$  の矩形は、 $t_1$  の矩形よりも大きく示されている。これは、例えば、2 人のユーザ（1 0 2 a、1 0 2 b）が追跡されており、これらのユーザが離れるように移動した結果として、又は、利用可能なリソースの変化（例えば、チャンネル帯域幅の増加）が、リソースマネージャ 3 0 6 に、決定されたアスペクト比を増大させた結果として、より多くの身体パーツを含む領域の選択（例えば、領域 1 から領域 2 への切り替え）がもたらされ、外接矩形が、それに応じてビデオコントローラ 3 0 4 により自動的に調整されたことに起因し得る。

## 【 0 0 6 1 】

ステップ S 9 0 4 において、ビデオモジュール 3 0 4 は、上述した動的モデルを使用して、第 1 のセットの境界データ及び第 2 のセットの境界データに基づいて、過渡データを生成する。生成された過渡データは、以前のフレームの外接矩形から現フレーム 9 5 6 の外接矩形への過渡を表す過渡外接矩形 9 5 4（図 9 に示されている）を実効的に規定する。

## 【 0 0 6 2 】

例えば、上記弾性バネモデルに従って、外接矩形は、異なる位置における 1 以上のポイント（境界データの形態）によりパラメータ化され得る。ポイントは、矩形 9 5 2 のパラメータである、時間  $t_2$  における第 2 の位置（「desiredPosition」）と、矩形 9 5 0 のパラメータである、時間  $t_1$  における第 1 の位置（「currentPosition」）と、を有し得る。この場合、過渡データは、以下のように「currentPosition」を更新することにより生成され得る（更新された「currentPosition」が、過渡外接矩形 9 5 4 のパラメータである）：

velocity=0

previousTime=0

currentPosition=<some\_constant\_initial\_value>

UpdatePosition(desiredPosition,time)

{

x=currentPosition-desiredPosition;

force=-stiffness\*x-damping\*m\_velocity;

acceleration=force/mass;

dt=time-previousTime;

velocity+=acceleration\*dt;

currentPosition+=velocity\*dt;

previousTime=time;

10

20

30

40

50



}

#### 【0063】

ステップS908において、ビデオコントローラは、選択された特徴を追跡するために、過渡データに基づいてビデオを制御する。ここで、これは、フレーム956のうち外接矩形954内の画像データのみが、符号化して遠端ユーザ108に送信するためにエンコーダ308に供給される（フレーム956のうち外接矩形954外の画像データは供給されない）ように、ビデオをクロップすることを含む。

#### 【0064】

検出器222及びセンサプロセッサ310は、検出器222から約1メートル以上離れているユーザ（すなわち、センサの視野外にいるユーザ）は追跡されないように構成されているので、そのようなユーザはクロッピング矩形に影響を及ぼさない。したがって、例えば、2人のユーザが、最初には追跡されていたが、1人のユーザが、視野外に移動した場合、方法800及び900に従って、以後に算出されるクロッピング矩形は、残っているユーザの特徴のみに基づくようになり、したがって、残っているユーザに自動的に「ズームインする」ようになる（弾性モデルは、このズームに関して滑らかな遷移を確実にする）。

#### 【0065】

諸実施形態において、1920×1080のビデオに十分な帯域幅があるが、追跡されるべき少なくとも1つの身体があることをリソースマネージャ306が考慮する場合には、動的な追跡を可能にするために、送信されるビデオの解像度が1280×720に切り替えられるように、ビデオコントローラ304は、リソースマネージャ306にフィードバックすることができる（送信されるビデオの解像度が、キャプチャされた時点の解像度と等しい場合には、クロッピング矩形は、各フレームについて、そのフレームの全体を実効的に包含するので、「操作（manoeuvre）の余地」はない - これは、例えば、アップスケーリングが用いられない場合に当てはまる）。

#### 【0066】

示されているように、センサプロセッサ310は、検出された骨格ポイントの各々を識別する情報を供給するだけでなく、センサ226の視野内の1人以上のユーザのうちどのユーザが、それらの骨格ポイントに対応しているかを識別する情報も供給する。したがって、説明した方法800は、（センサ検出器226がユーザを識別することができる限り）任意の数のユーザに対して実施することができ、選択された領域内の特徴ごとに、各領域について特徴を追跡する。したがって、方法800は、外接矩形が、新たなユーザがフレームに入ったときに新たなユーザに対応するために、（弾性モデルにより）滑らかに自動的に遷移する（ビデオ内の選択された領域に関してそのユーザの身体パーツを含むように実効的にズームアウトすることにより）とともに、ユーザがフレームから出たときに調整するために、（弾性モデルにより）滑らかに自動的に遷移する（残っているユーザの、選択された領域に関する身体パーツのみが、ビデオ内に保たれるように、選択された領域に関するそのユーザの身体パーツにより以前に占められていた領域を除外するように実効的にズームインすることにより）ことで、複数のユーザがフレームに入ったとき及びフレームから出たときに自動的に適応する。

#### 【0067】

コントローラ302により実行される方法800、900は、例えば、ユーザデバイス104、110、及び114を使用して（ユーザ102a、102b、108、及び112の間で）、ネットワーク106を介して行われるマルチパーティ通話中に実行されてもよい。この場合、近端デバイス104（第1のユーザデバイス）により、遠端デバイス110（第2のユーザデバイス）及び遠端デバイス114（第3のユーザデバイス）の各々について、個々の外接矩形が決定される。

#### 【0068】

すなわち、上記に加えて、リソースマネージャは、ユーザデバイス104と第3のユー

10

20

30

40

50

ザデバイス 114 との間の通信チャネルと、さらなる第 3 のユーザデバイス 114 のリソースと、のうちの少なくとも 1 つに関するさらなる情報を受信することができる（第 2 のユーザデバイス 110 に関する同様の情報等の上述した情報を受信することに加えて）。次いで、リソースマネージャは、第 3 のユーザデバイス 114 に送信されるビデオを制御するために、受信されたさらなる情報に基づいて、複数の視覚的なユーザ特徴（例えば身体パーツ）からさらなる特徴を選択する（第 2 のユーザデバイス 110 に送信されるビデオを制御するために、前述した特徴を選択することに加えて）。

【0069】

次いで、ビデオコントローラ 304 は、第 2 のユーザデバイス 110 に関して選択された特徴の検出に基づいて、第 2 のユーザデバイス 110 に送信されるビデオを制御するとともに、（第 3 のユーザデバイス 114 のビデオ内の選択されたさらなる特徴を追跡するために、）第 3 のユーザデバイス 114 に関して選択されたさらなる特徴の検出に基づいて、第 3 のユーザデバイス 114 に送信されるビデオを制御する。

【0070】

第 3 のユーザデバイスに関するさらなる特徴の選択は、第 2 のユーザデバイスに関する特徴の選択とは独立した異なるものであり得る。したがって、第 3 のユーザデバイスに送信されるさらなるビデオは、第 2 のユーザデバイスに送信されるビデオとは異なるものであり得、第 3 のユーザデバイスに送信されるさらなるビデオにより、第 2 のユーザデバイスに送信されるビデオよりも多い又は少ないユーザ特徴が追跡され得る。

【0071】

第 2 のユーザデバイスのビデオに関する特徴の選択は、第 3 のユーザデバイスのビデオに関する特徴の選択とは独立したものである。（第 1 のユーザデバイス 104 が遅い接続を介してネットワーク 106 に接続されている等の）状態は、両方に関して同様のクロッピングをもたらし得るのに対し、（第 2 のユーザデバイス及び第 3 のユーザデバイスのうちの一方が、遅い接続を介してネットワーク 106 に接続されている、あるいは、第 2 のユーザデバイス及び第 3 のユーザデバイスのうちの一方が、制限されたリソースを有する等の）他の状態は、異なるクロッピングをもたらし得る。

【0072】

例えば、第 3 のユーザデバイス 114 は、小型スクリーン（例えばスマートフォンスクリーン）を有し得る、且つ／又は、遅い接続を介してネットワーク 106 に接続され得る。一方、第 2 のユーザデバイス 110 は、大型スクリーンを有し得る（例えば、TV スクリーンに接続され得る）、且つ／又は、速い接続を介してネットワーク 106 に接続され得る。この場合、第 2 のユーザデバイス 110 に送信されるビデオは、ユーザ 108 が、ユーザ 102 a 及び 102 b の全身を表示するビデオを受信するように、「領域 4」のクロッピング（表 2 参照）が適用され得る。一方、第 3 のユーザデバイス 114 に送信されるビデオは、ユーザ 102 a、102 b のそれぞれの頭及び肩だけを表示するビデオを受信するように、「領域 1」のクロッピング（表 2 参照）が適用され得る。

【0073】

上記において、ネットワークを介して送信されるビデオは、ビデオ信号処理を介して制御されるが、代替的又は追加的に、ビデオは、選択された特徴の検出に基づいてカメラ自体を操作することにより、（コントローラ 302 の）ビデオコントローラ 304 によって制御されてもよい。例えば、そのような操作には、パン操作、ズーム操作、及びティルト操作のうちの少なくとも 1 つを実行するようにカメラの機構（mechanics）を操作することが含まれる。例えばマルチパーティ通話の場合、光学的ズーム及びデジタルズーム（クロッピング）が、例えば、最高選択数のユーザ特徴を表示するビデオをキャプチャするために使用される機械的ズーム（例えば、機械的ズームは、上記の例においては第 2 のユーザデバイス 110 に対して表示される領域を選択するために使用され得る）とともに使用され得、そのビデオのクロッピングは、より少ないユーザ特徴を含むビデオが送信されるべきユーザ（例えば、上記の例においては第 3 のユーザデバイス 114）のためのビデオを制御するために使用され得る。

## 【 0 0 7 4 】

さらに、上記において、弾性モデルが使用されるが、代替として、過渡データを生成するために、（例えば、1以上の時間に関する微分方程式（differential equations in time）に基づく）任意の動的モデルが使用されてもよい。

## 【 0 0 7 5 】

さらに、上記において、選択された視覚的な特徴（脚、腕等）は、非可視放射線パターンを、当該パターンを検出するよう構成されているセンサの前方に投射する奥行き検出器により供給されたセンサデータに基づいて検出されるが、代替の検出も想定されている。例えば、奥行き検出は、奥行きを測定するために放射線伝搬時間が使用される検出に基づく飛行時間型（time-of-flight）であってもよい。あるいは、（複数の2次元画像から）異なる勾配（inclination）を有するカメラ（3D画像認識アルゴリズムとともに使用されるプレノプティックカメラ又は同様のカメラ等）のアレイを使用して3D画像を構築し、この3D画像から視覚的な特徴が検出されてもよい。

## 【 0 0 7 6 】

一般に、本明細書で説明した機能（例えば、図3に示される機能モジュール及び図8と図9とに示される機能ステップ）のいずれも、ソフトウェア、ファームウェア、ハードウェア（例えば固定論理回路）、又はこれらの組合せを用いて実装することができる。図3において別個に示されているモジュール（ビデオ処理システム300、コントローラ302、ビデオ信号プロセッサ304、リソースマネージャ306、エンコーダ308等）及び図8と図9とにおいて別個に示されているステップは、別個のモジュール及び別個のステップとして実装されてもよいし、そうでなくてもよい。本明細書で使用されている「モジュール」、「機能」、「コンポーネント」、及び「論理」という用語は、一般に、ソフトウェア、ファームウェア、ハードウェア、又はこれらの組合せを表す。ソフトウェア実装の場合、モジュール、機能、又は論理は、プロセッサ（例えば1以上のCPU）上で実行されたときに指定されたタスクを実行するプログラムコードを表す。プログラムコードは、1以上のコンピュータ読み取り可能なメモリデバイスに記憶することができる。本明細書で説明した技術の特徴は、プラットフォーム非依存であり、これは、本技術が、多様なプロセッサを有する多様な商用コンピューティングプラットフォーム上で実装され得ることを意味する。例えば、ユーザデバイスは、例えば、プロセッサや機能ブロック等といった、ユーザデバイスのハードウェアに処理を実行させるエンティティ（例えばソフトウェア）も含み得る。例えば、ユーザデバイスは、ユーザデバイス（より詳細には、ユーザデバイスのオペレーティングシステム及び関連ハードウェア）に処理を実行させる命令を保持するよう構成され得るコンピュータ読み取り可能な媒体を含み得る。例えば、図3のモジュールの一部又は全ては、1以上のプロセッサ上で実行されるクライアントアプリケーションのソフトウェアにより実装され得る。したがって、命令は、処理を実行するようオペレーティングシステム及び関連ハードウェアを構成するように機能し、そのようにして、機能を実行するオペレーティングシステム及び関連ハードウェアの変換をもたらす。命令は、多種多様な構成を通じて、コンピュータ読み取り可能な媒体により、ユーザデバイスに提供され得る。

## 【 0 0 7 7 】

コンピュータ読み取り可能な媒体の1つのそのような構成は、信号担持媒体であり、したがって、ネットワークを介して等、（例えば搬送波として）命令を伝送するよう構成される。コンピュータ読み取り可能な媒体はまた、コンピュータ読み取り可能な記憶媒体として構成され得、したがって、信号担持媒体ではない。コンピュータ読み取り可能な記憶媒体の例は、ランダムアクセスメモリ（RAM）と、読み取り専用メモリ（ROM）と、光ディスクと、フラッシュメモリと、ハードディスクメモリと、磁気技術、光技術、及び他の技術を使用して命令及び他のデータを記憶することができる他のメモリデバイスと、を含む。

## 【 0 0 7 8 】

構造的特徴及び/又は方法的動作に特有の言葉で主題について説明したが、添付の特許

10

20

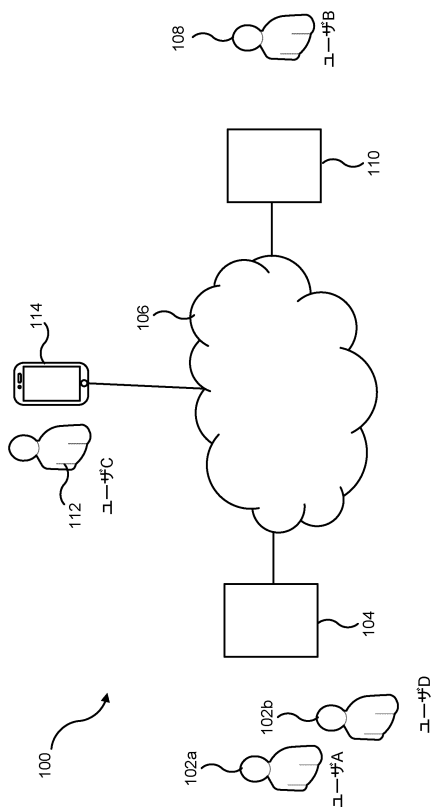
30

40

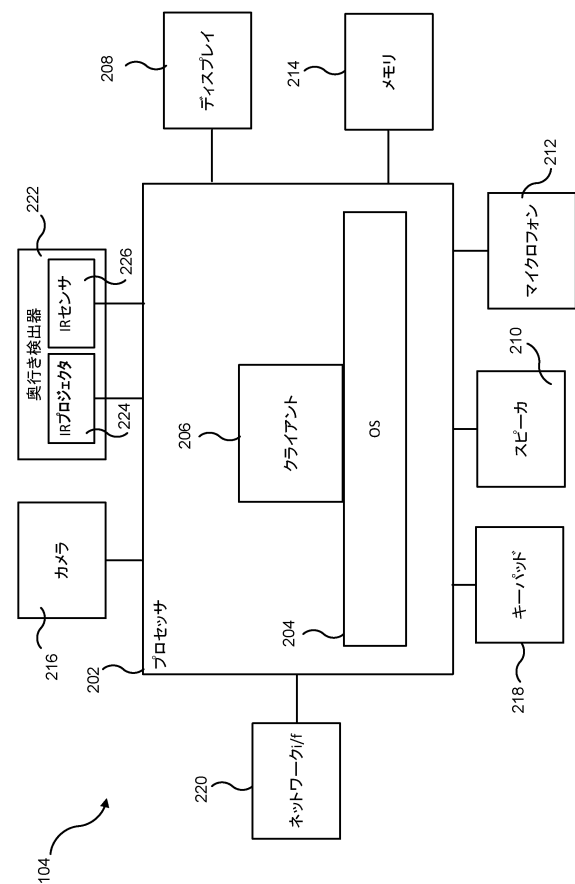
50

請求の範囲において定められる主題は、上述した特定の特徴又は動作に必ずしも限定されるものではないことを理解されたい。そうではなく、上述した特定の特徴及び動作は、請求項を実施する例示的な形態として開示されている。

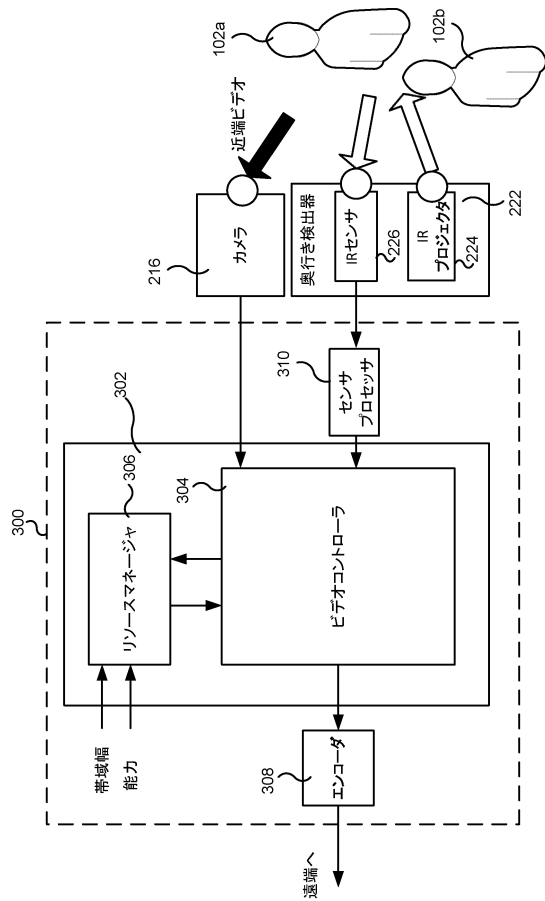
【図 1】



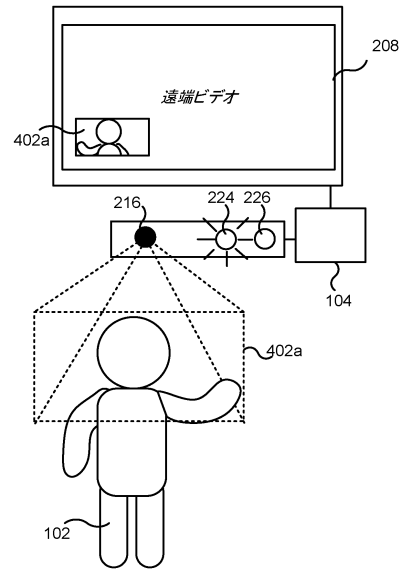
【図 2】



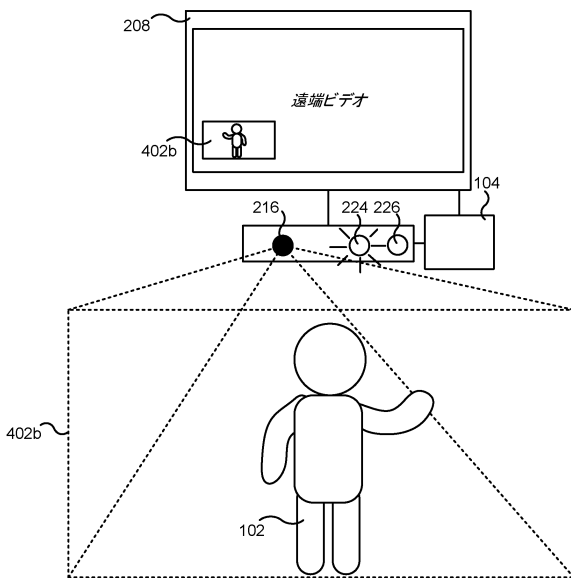
【 図 3 】



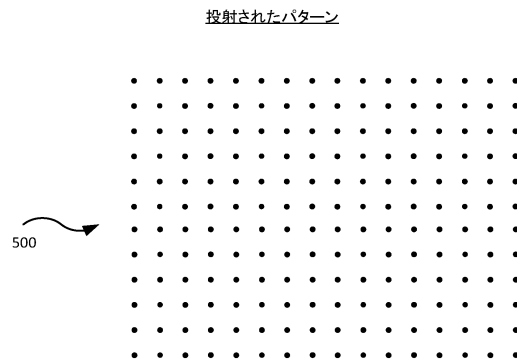
【 図 4 A 】



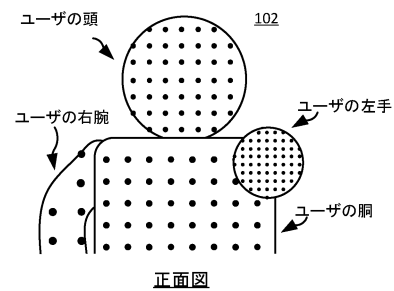
【 図 4 B 】



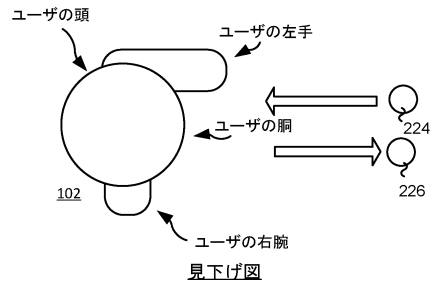
【 図 5 】



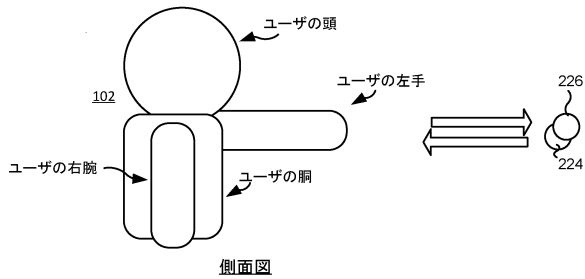
【 図 6 A 】



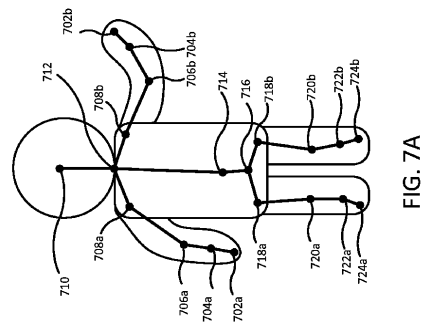
【図 6 B】



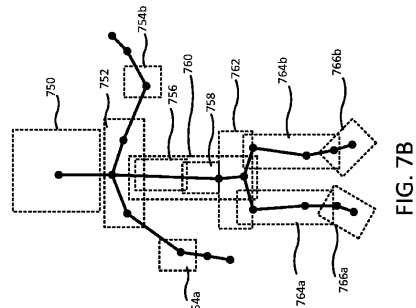
【図 6 C】



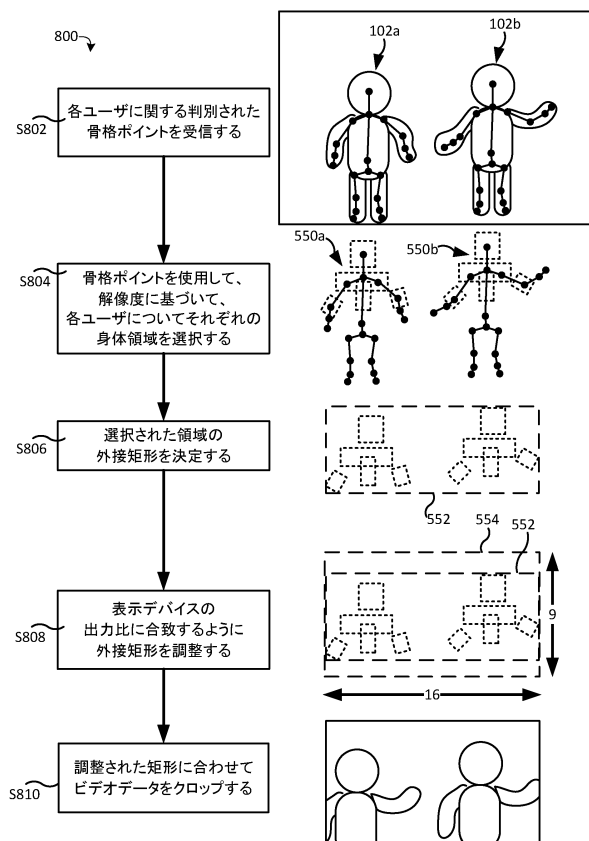
【図 7 A】



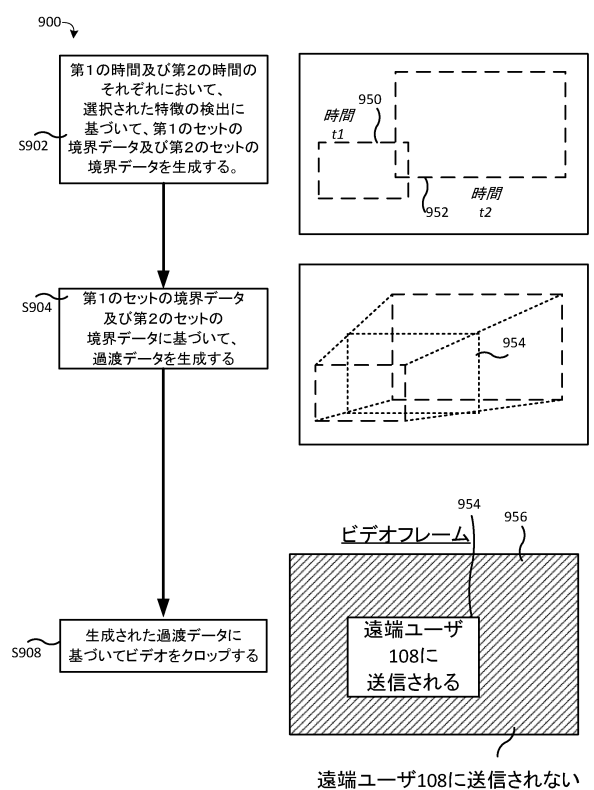
【図 7 B】



【図 8】



【図 9】



## フロントページの続き

- (72)発明者 ベレット, ダグラス ディー .  
アメリカ合衆国 98052 - 6399 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 エルシーエー - インターナショナル パテンツ ( 8 / 1172 ) 内
- (72)発明者 ドラーニュ, ルシアン  
アメリカ合衆国 98052 - 6399 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 エルシーエー - インターナショナル パテンツ ( 8 / 1172 ) 内
- (72)発明者 コレンカーク, ショーン イー .  
アメリカ合衆国 98052 - 6399 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 エルシーエー - インターナショナル パテンツ ( 8 / 1172 ) 内

審査官 富樫 明

- (56)参考文献 米国特許出願公開第2012/0092445 (US, A1)  
特開2006-128997 (JP, A)  
特表2013-524684 (JP, A)  
米国特許出願公開第2012/0314077 (US, A1)

(58)調査した分野(Int.Cl., DB名)

H04N 7/14 - 7/15  
H04N 21/00 - 21/858  
G06F 13/00  
G06T 1/00