

(19) 日本国特許庁(JP)

## (12) 特 許 公 報(B2)

(11) 特許番号

特許第4625074号  
(P4625074)

(45) 発行日 平成23年2月2日(2011.2.2)

(24) 登録日 平成22年11月12日(2010.11.12)

(51) Int.Cl.

F 1

GO6T	7/20	(2006.01)	GO6T	7/20	300A
GO6F	3/01	(2006.01)	GO6F	3/01	310C
GO1B	11/00	(2006.01)	GO1B	11/00	H
GO1C	3/06	(2006.01)	GO1C	3/06	120Q
GO1B	11/24	(2006.01)	GO1C	3/06	140

請求項の数 19 (全 30 頁) 最終頁に続く

(21) 出願番号	特願2007-513417 (P2007-513417)
(86) (22) 出願日	平成17年5月13日 (2005.5.13)
(65) 公表番号	特表2007-538318 (P2007-538318A)
(43) 公表日	平成19年12月27日 (2007.12.27)
(86) 國際出願番号	PCT/US2005/016828
(87) 國際公開番号	W02005/114556
(87) 國際公開日	平成17年12月1日 (2005.12.1)
審査請求日	平成20年5月13日 (2008.5.13)
(31) 優先権主張番号	60/571,319
(32) 優先日	平成16年5月14日 (2004.5.14)
(33) 優先権主張國	米国(US)
(31) 優先権主張番号	60/635,146
(32) 優先日	平成16年12月10日 (2004.12.10)
(33) 優先権主張國	米国(US)

(73) 特許権者	000005326 本田技研工業株式会社 東京都港区南青山二丁目1番1号
(73) 特許権者	506381360 オハイオ ステート ユニバーシティ リ サーチ ファウンデーション アメリカ合衆国、オハイオ州 43210 -1063, コロンバス, ケニー ロード 1960
(74) 代理人	100064414 弁理士 磯野 道造

最終頁に続く

(54) 【発明の名称】 サインに基づく人間-機械相互作用

## (57) 【特許請求の範囲】

## 【請求項 1】

コンピュータ処理に基づきマンマシンインタラクションを実現するための方法であって、  
画像ピクセルに関連し、デジタル画像を取得するカメラへの距離に対応する深度データ  
を含む対象のデジタル画像を受信するステップと、

前記画像ピクセルのそれぞれ及び前記対象の形状に関する前記深度データに基づいて  
前記画像ピクセル内の対象の手の位置を決定するステップと、

記憶された複数の形状画像の一以上と候補画像とを比較することにより、前記デジタル  
画像で取得された前記対象の前記手の候補画像と前記複数の形状画像からの記憶された形  
状画像とをマッチングすることによって前記対象のデジタル画像を前処理し、前記比較は  
、前記候補画像に関連する前記画像ピクセルのそれぞれに関連する前記深度データを有す  
るベクトルを形成し、記憶された複数の形状画像を表す射影係数ベクトルの行列に基づい  
てマッチングを算出することを含むステップと、

前記手の位置の画像ピクセルデータから前記手の骨格表現を生成し、前記骨格表現から  
少なくとも指の数を含む一以上の手の特性を決定することにより、前記手の候補画像をマ  
ッチングするステップと、

を含むことを特徴とするコンピュータ処理に基づきマンマシンインタラクションを実現  
するための方法。

## 【請求項 2】

前記手の候補画像をマッチングするステップは、

10

20

前記画像ピクセルの決定された位置における前記手のエッジを決定するステップと、  
前記手のエッジに関する前記画像ピクセルにピクセル距離変換を適用して候補エッジ画像を作成するステップと、

をさらに含み、

前記候補エッジ画像と記憶された画像の距離変換との間の相関関係に基づいて類似度値を決定するために、記憶された複数の形状画像の一以上は、記憶された形状画像への前記候補画像の射影に基づいて前記候補画像と比較するための画像のエッジピクセル変換を含む

ことを特徴とする請求項1に記載の方法。

**【請求項3】**

10

前記対象の手の位置を決定するステップは、

前記デジタル画像を複数の所定の領域に分割するステップと、

所定の領域と前記手の位置とを関連付けるステップと、

対象の頭の位置及び対象の胴体の位置の少なくとも一つと関連する一以上の領域を決定するステップと、

をさらに含むことを特徴とする請求項1に記載の方法。

**【請求項4】**

20

コンピュータ処理に基づきマンマシンインタラクションを実現するための方法であって各画像ピクセルに関連し、デジタル画像を取得するカメラまでの距離を表す深度データを備える対象のデジタル画像を受信するステップと、

前記デジタル画像を前処理して、前記各画像ピクセルに関連する前記深度データに基づいて、少なくとも手の位置と、前記対象の頭及び胴体の少なくとも一つの位置と、を決定するステップと、

前記手の形状と記憶された複数の手の形状画像の一つとをマッチングして、前記デジタル画像に関連するマッチングされた形状を提供するステップであって、前記手の位置に関連する前記画像ピクセルのそれぞれに関連する前記深度データを有するベクトルを形成し、記憶された複数の手の形状画像を表す射影係数ベクトルの行列に基づいてマッチングを算出することを含むステップと、

前記対象の頭及び胴体の一つに対する前記手の位置に関連する情報と、前記デジタル画像に関連するマッチングされた形状を示す情報と、を含む候補画像データオブジェクトを生成するステップと、

前記候補画像データオブジェクトとテンプレートジェスチャに関して記憶された一以上のジェスチャプロファイルとを比較することにより、一以上のデジタル画像により取得されたジェスチャとテンプレートジェスチャとをマッチングし、前記ジェスチャプロファイルは、ボディパーティに対する手の位置情報と、前記候補画像データオブジェクトと比較するための手の形状情報と、を含むステップと、

前記手の位置の画像ピクセルデータから手の骨格表現を生成し、前記骨格表現から少なくとも指の数を含む一以上の手の特性を決定することにより、前記手の候補形状をマッチングするステップと、

を含むことを特徴とするコンピュータ処理に基づきマンマシンインタラクションを実現するための方法。

**【請求項5】**

40

トレーニング処理を含む前処理が、

前記対象の既知のポーズを有する画像のトレーニングセットの主成分分析を実行して射影行列を得るステップと、

前記射影行列に基づく前記対象の候補ポーズに対応するベクトルを射影することにより前記デジタル画像内の前記対象のポーズを決定し、既知のポーズを有する前記トレーニングセット内の最近傍のポーズを得るステップと、

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】**

50

前記候補画像データオブジェクトは、形状情報、向き情報及び軌道情報を含む掌情報セクションと、前記指の数及び前記指の方向を含む指情報セクションと、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】**

前記対象の前記デジタル画像は、ハイブリッドな深度・画像情報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】**

前処理が、前記ハイブリッドな深度・画像情報を分析して、前記対象の画像取得システムまでの距離に基づいて位置を決定するステップを含むことを特徴とする請求項7に記載の方法。 10

**【請求項9】**

一以上のデジタル画像により取得されたジェスチャとテンプレートジェスチャとをマッチングするステップは、

用いられる手の数、用いられる指の数、指間の距離、手の位置、手の位置の変化、手の形状の情報及び指情報からなるグループのうちの一以上の特性に関連する値を比較するステップ

をさらに含むことを特徴とする請求項4に記載の方法。

**【請求項10】**

前記手の形状の情報は、指の数、指の識別番号、指の形状及び指の向きからなるグループのうちの一以上の特性をさらに含むことを特徴とする請求項9の記載の方法。 20

**【請求項11】**

コンピュータ処理に基づきマンマシンインタラクションを実現するための方法であって  
デジタル画像を取得するカメラまでの距離を表す各画像ピクセルに関連する深度データを備える一以上のデジタル画像で取得された手の形状と複数の手の形状パターンとをマッチングして候補ジェスチャ手形状を決定するステップであって、前記画像ピクセルのそれに関連する前記深度データを有するベクトルを形成し、記憶された複数の手の形状パターンを表す射影係数ベクトルの行列に基づいてマッチングを算出することを含むステップと、

一以上のデジタル画像から前記手の軌道曲線を決定するステップと、

前記手の軌道曲線と複数の軌道曲線テンプレートとをマッチングして候補ジェスチャ動きを決定するステップと、 30

前記候補ジェスチャ手形状及び前記候補ジェスチャ動きとジェスチャ手形状及びジェスチャ動きに関連する複数のジェスチャプロファイルとを比較することにより、前記候補ジェスチャ手形状及び前記候補ジェスチャ動きに対応するジェスチャを決定するステップと、

前記手の位置の画像ピクセルデータから前記手の骨格表現を生成し、前記骨格表現から少なくとも指の数を含む一以上の手の特性を決定することにより、前記手の候補形状をマッチングするステップと、

を含むことを特徴とするコンピュータ処理に基づきマンマシンインタラクションを実現するための方法。 40

**【請求項12】**

一以上のデジタル画像で取得された前記手の形状が静止しているか否かを判定するステップと、

前記手の形状が静止していると判定された場合に、前記手の形状と静止したジェスチャに対応する複数の手の形状パターンとを比較してジェスチャを決定するステップと、

をさらに含むことを特徴とする請求項11に記載の方法。

**【請求項13】**

人間画像情報を含む画像データのフレームを受信し、前記フレーム内の人間の一以上のボディパーツの位置を決定するように構成されており、前記画像データのフレームのそれぞれは、デジタル画像を取得するカメラまでの距離を表す各画像ピクセルに関連する深度 50

データを備える画像データ前処理モジュールと、

決定された前記位置を受信できるように前記画像データ前処理モジュールに接続され、前記画像データに関する情報と形状データベースに記憶された形状プロファイルとを比較することにより前記一以上のボディパーツの形状をマッチングするように構成されており、前記比較は、前記一以上のボディパーツの決定された位置に関する前記画像ピクセルのそれぞれに関する前記深度データを有するベクトルを形成し、記憶された複数のボディパーツ形状画像を表す射影係数ベクトルの行列に基づいてマッチングを算出することを含む形状マッチングモジュールと、

形状及び位置情報を受信できるように前記画像データ前処理モジュール及び前記形状マッチングモジュールに接続され、前記一以上のボディパーツのジェスチャ特性とジェスチャデータベースに記憶されたジェスチャプロファイルとを比較することにより前記一以上のボディパーツのジェスチャをマッチングするジェスチャマッチングモジュールと、

を備え、

前記ジェスチャ特性は、ボディパーツの位置に関する値と形状情報を含み、

前記ジェスチャマッチングモジュールは、前記手の位置の画像ピクセルデータから前記手の骨格表現を生成し、前記骨格表現から少なくとも指の数を含む一以上の手の特性を決定することにより、前記手の候補形状をマッチングするように構成された手形状マッチングモジュールをさらに備える

ことを特徴とするコンピュータ処理に基づきマンマシンインタラクションを実現するシステム。

#### 【請求項 1 4】

画像データ取得システムをさらに備え、

前記画像データ取得システムは、当該画像データ取得システムからの深度の範囲内のシーンのハイブリッドな画像 - 深度情報を取得するように構成されている

ことを特徴とする請求項 1 3 に記載のコンピュータに基づくシステム。

#### 【請求項 1 5】

前記画像データ取得システムは、単一の光軸に沿って前記ハイブリッドな画像 - 深度情報を取得するように構成された飛行時間カメラを備えていることを特徴とする請求項 1 4 に記載のコンピュータに基づくシステム。

#### 【請求項 1 6】

前記形状マッチングモジュール及び前記ジェスチャマッチングモジュールに接続され、前記形状データベース及び前記ジェスチャデータベースを備えたデータ記憶システムをさらに備えることを特徴とする請求項 1 3 に記載のコンピュータに基づくシステム。

#### 【請求項 1 7】

前記画像データ前処理モジュール、前記形状マッチングモジュール及び前記ジェスチャマッチングモジュール間の接続は、RAM装置を備えていることを特徴とする請求項 1 3 に記載のコンピュータに基づくシステム。

#### 【請求項 1 8】

デジタル画像を取得するカメラまでの距離を表す各画像ピクセルに関する深度データを備える対象のデジタル画像を受信するためのプログラム命令と、

前記デジタル画像に前処理を施して、各画像ピクセルに関する前記深度データに基づいて少なくとも手の位置と前記対象の頭及び胴体の少なくとも一つの位置とを決定するためのプログラム命令と、

前記手の形状と記憶された複数の手の形状画像の一つとをマッチングして、前記デジタル画像に関するマッチングされた形状を提供するためのプログラム命令であって、前記手の位置に関する画像ピクセルのそれぞれに関する深度データを有するベクトルを形成するためのプログラム命令と、記憶された複数の手形状画像を表現する射影係数ベクトルの行列に基づいてマッチングを算出するためのプログラム命令と、を備えるプログラム命令と、

前記対象の頭及び胴体の一つに対する前記手の位置に関する情報と、前記デジタル画

10

20

30

40

50

像に関するマッチングされた形状を示す情報と、を含む候補画像データオブジェクトを生成するためのプログラム命令と、

前記候補画像データオブジェクトとテンプレートジェスチャに関して記憶された一以上のジェスチャプロファイルとを比較することにより、一以上のデジタル画像により取得されたジェスチャとテンプレートジェスチャとをマッチングするためのプログラム命令と、

前記手の位置の画像ピクセルデータから前記手の骨格表現を生成し、前記骨格表現から少なくとも指の数を含む一以上の手の特性を決定することにより、前記手の候補形状をマッチングするためのプログラム命令と、

を含み、

前記ジェスチャプロファイルは、ボディパーティに対する手の位置情報と、前記候補画像データオブジェクトと比較するための手の形状情報と、を含む 10

ことを特徴とするコンピュータ処理に基づきマンマシンインタラクションを実現するためのプログラムを記録したコンピュータ読取可能な媒体。

#### 【請求項 19】

デジタル画像を取得するカメラまでの距離を表す各画像ピクセルに関する深度データを備える対象のデジタル画像を受信するための手段と、

前記デジタル画像に前処理を施して、各画像ピクセルに関する前記深度データに基づいて少なくとも手の位置と前記対象の頭及び胴体の少なくとも一つの位置とを決定するための手段と、

前記手の形状と記憶された複数の手の形状画像の一つとをマッチングして、前記デジタル画像に関するマッチングされた形状を提供するための手段であって、前記手の位置に関する画像ピクセルのそれぞれに関する深度データを有するベクトルを形成するための手段と、記憶された複数の手形状画像を表現する射影係数ベクトルの行列に基づいてマッチングを算出するための手段と、を備える手段と、 20

前記対象の頭及び胴体の一つに対する前記手の位置に関する情報と、前記デジタル画像に関するマッチングされた形状を示す情報と、を含む候補画像データオブジェクトを生成するための手段と、

前記候補画像データオブジェクトとテンプレートジェスチャに関して記憶された一以上のジェスチャプロファイルとを比較することにより、一以上のデジタル画像により取得されたジェスチャとテンプレートジェスチャとをマッチングするための手段と、 30

前記手の位置の画像ピクセルデータから前記手の骨格表現を生成し、前記骨格表現から少なくとも指の数を含む一以上の手の特性を決定することにより、前記手の候補形状をマッチングするための手段と、

を備え、

前記ジェスチャプロファイルは、ボディパーティに対する手の位置情報と、前記候補画像データオブジェクトと比較するための手の形状情報と、を含む

ことを特徴とするコンピュータ処理に基づきマンマシンインタラクションを実現するシステム。

#### 【発明の詳細な説明】

##### 【技術分野】

##### 【0001】

本出願は、Liuらにより2004年5月14日に出願された米国仮特許出願第60/571,319号「深度データを用いた手のジェスチャ認識」に基いて優先権を主張するものであり、参照により本明細書中に組み込むものとともに、フジムラらにより2004年12月10日に出願された米国仮特許出願第60/635,146号「サインに基づくマンマシンインタラクション」に基いて優先権を主張するものであり、参照により本明細書中に組み込むものとする。

##### 【0002】

本発明は、一般的にマンマシンインタラクションに関し、より詳細には、人間のサイン及びジェスチャのコンピュータ視覚認識に基づくヒューマンコンピュータインタラクション 50

ン（H C I）に関する。

【背景技術】

【0003】

超高速計算システム及び高効率デジタル撮像システムの出現により、マンマシンインタラクションに基づくコンピュータビジョンの分野は有意な技術的進歩の時代を経験している。動きが機械からの応答を誘導する単純な動き検出システム（例えば、監視システム）から高度に複雑な3次元（「3D」）撮像サイン認識システムまで、最近数年で急激に発展したテーマである。例えば、マンマシンコミュニケーションに基づくサインの領域において、人間の手話の認識は、マンマシンコミュニケーションの前途有望な技術として最近最も研究されたテーマである。他のサイン認識システム及びより複雑なジェスチャ認識システムは、手及び手の他のボディパーツ（例えば、腕、胴体、頭等）に対する動きを位置決めし追跡するための様々な手法に基づいて開発されてきた。

10

【0004】

これらのサイン及びジェスチャのための従来の技術は、一般的に、サイン又はジェスチャのソースを見つけ出すときにマシンビジョンシステムを助けるために、マーク、特定の色、背景又はグローブを必要としている。例えば、手検出に関するいくつかの従来のアプローチは、手又はシステムへの手のジェスチャに対応する画像領域を決定するために色又は動き情報を用いている。これらのアプローチにおいて、手の動きを追跡することは、様々な照明条件下では非常に信頼できない。いくつかのシステムは、グローブのような特別な装置を用いており、他のいくつかは、タスクが実現可能なように特定の色を有する背景を用いている。

20

【0005】

これらの従来のシステムにおいて、手の位置が画像フレームを横切るように決定されて追跡された場合には、提供されたサインを決定するための手の形状の分析は、非常に困難なタスクである。画像データのリアルタイム処理に基づく形状を分析しようとする場合には、形状認識タスクは、さらに困難になる。手形状分析処理の効率を改善するために、手領域の拡大された視野が、画像取得装置をシーンの好適な領域にフォーカシングすることにより用いられる。しかし、ボディ、腕、頭又は他のポスチャ情報が、撮像フレームから外れることにより失われてしまうおそれがある。

30

【0006】

画像データにおける全体的なボディのポスチャ情報を取得するために、いくつかの深度取得システムが複数のカメラを深度情報を決定するための手段として用いている。異なる角度から取得された画像は、サブジェクトの3D情報を抽出するために有効なデータを提供する。得られた3Dボリュームを理解するために、3Dモデル適合が提案されているが、複雑かつ集中的な計算を要する処理であり、しばしば不安定であり一般的にはリアルタイムアプリケーションに適合しない。深度取得システムに一般的に選択されるステレオビジョンは、大抵の場合サブジェクトのテクスチャの欠乏により手形状分析のために有効な画像解像度を提供することができない。このように、深度は、第一に、大きな指示の動きのような単純なジェスチャの制限されたセットを検出するために用いられる。いくつかのアプローチは、静的な手ポスチャ分析のためにコード化された照明を用いることにより得られた高解像度3Dデータを用いる。しかし、これらの高解像度システムは、一般的に入手不可能又はリアルタイム実行が不能な、大規模な計算及び記憶リソースを必要とする。

40

【発明の開示】

【発明が解決しようとする課題】

【0007】

したがって、（1）手の位置、手の形状、動き、向き、速さ等を同時に含む複数の特徴を分析することができ、（2）特別な背景、マーク又はグローブを用いることなくジェスチャを認識することができ、（3）計算的に安価な計算モジュールに基づいてリアルタイムジェスチャ認識を達成することができ、（4）明るさの変化にほとんど影響を受けない、マンマシンインタラクションのためのジェスチャ及びサイン認識に基づくコンピュータ

50

へのアプローチが必要とされている。

【課題を解決するための手段】

【0008】

本発明によると、マンマシンインタラクションのためのシステム、方法、装置及びコンピュータ読取可能な媒体が提供される。一実施形態において、システムは、マンマシンインタラクションのためのコマンド及び他のコミュニケーションとしてのジェスチャ及びサインの認識及び解釈に基づくコンピュータビジョンを提供する。

【0009】

一実施形態によると、マンマシンインタラクションのためのシステム及び方法は、対象のデジタル画像における深度及び画像情報に基づき動作する。対象のデジタル画像が受信される。デジタル画像は、対象（又は他の背景）からカメラまでの距離に対応する画像ピクセルに関連する深度データを含む。画像のピクセル内の対象の手の位置は、ピクセルに関連する深度データに基づいて決定される。デジタル画像内で取得された対象の手の候補形状は、記憶された形状画像と候補形状とを比較することにより、形状画像のセットからの記憶された形状画像とマッチングされる。

10

【0010】

他の実施形態によると、マンマシンインタラクションのための方法及びシステムは、受信した対象のデジタル画像に頼っている。デジタル画像は、対象の少なくとも手及び頭又は胴体の位置を決定するために前処理を施される。手の形状は、記憶された手の形状とマッチングされて、デジタル画像に関連するマッチングされた形状を提供する。候補画像データオブジェクトは、対象の頭又は胴体に対する手の位置に関連する情報を含んで生成される。候補画像データオブジェクトは、デジタル画像に関連してマッチングされた形状を示す情報を含む。デジタル画像により取得されたジェスチャは、候補画像データオブジェクトと記憶されたジェスチャプロファイルとを比較することにより記憶されたジェスチャとマッチングされる。記憶されたジェスチャプロファイルは、同様に、ボディパーツに対する手の位置情報と、候補画像データオブジェクトと比較するための手の形状情報と、を含む。

20

【0011】

さらに他の実施形態によると、マンマシンインタラクションのためのシステムは、画像データ前処理モジュールと、形状マッチングモジュールと、ジェスチャマッチングモジュールと、を備えている。画像データ前処理モジュールは、人間画像情報を含む画像データのフレームを受信するためのものである。さらに、画像データ前処理モジュールは、フレームにおいて人間の一以上のボディパーツの位置を決定することができるよう構成されている。この及び他の情報が、形状マッチングモジュールにより用いられ、画像データに関連する情報と形状データベース内に記憶された形状プロファイルとを比較することによりボディパーツの形状をマッチングする。この目的のために、形状マッチングモジュールは、画像データ前処理モジュールに接続されており、決定された位置を受信することができる。さらに、ジェスチャマッチングモジュールも、画像データ前処理モジュール及び形状マッチングモジュールに接続されており、形状及び位置情報を受信することができる。ジェスチャマッチングモジュールは、ボディパーツのジェスチャ特性とジェスチャデータベースに記憶されたジェスチャプロファイルとを比較することによりボディパーツのジェスチャをマッチングすることができる。ジェスチャ特性は、ボディパーツの位置に関連する値と形状情報を含む。

30

【0012】

本明細書に記載された特徴及び利点は、全てを包含しているものではなく、特に、当業者にとっては、図面、明細書及び特許請求の範囲を参照することにより、多くの追加的な特徴及び利点が明らかとなるであろう。さらに、明細書で用いられる言語は、原則として読みやすさと説明の目的のために選択されたものであり、発明の主題を線引き又は制限するためには選択されたものではないことに留意されたい。

【発明を実施するための最良の形態】

40

50

## 【0013】

符号及び以下の説明は、説明のみによって本発明の好適な実施形態に関連している。以下の説明から、本明細書に開示された構造及び方法の他の実施形態が特許請求の範囲に記載された本発明の原則を逸脱しない範囲で用いられる実行可能な代替案として容易に認識されることに留意されたい。

## 【0014】

図1(A)を参照すると、パーソナルコンピューティングマンマシンインタラクションシステムの一実施形態が示されている。この実施形態において、システム100aは、例えば、一以上のデジタルビデオカメラ、デジタル静止画カメラ、飛行時間カメラ、赤外線カメラ等の撮像装置102を備えている。撮像装置102は、パーソナルコンピューティング装置103に接続されている。パーソナルコンピューティング装置103は、ディスプレイ108に接続されている。10

## 【0015】

パーソナルコンピューティング装置103は、有線又は無線接続を介して撮像装置102に接続されている。パーソナルコンピューティング装置103は、コンピュータプロセッサ及びRAM、ROM、フラッシュメモリ、ディスクドライブ等を含むコンピュータ読取可能な記録媒体のような従来のデータ処理及び記憶装置を備えている。コンピュータ読取可能な媒体は、以下に説明するように、コンピュータシステムで実効される場合に、ある機能を実行するためのステップを実行するコンピュータプログラム命令を記憶することができる。好ましくは、コンピュータプロセッサは、実行可能なソフトウェアプログラムから得られるプログラム命令を実行するときに画像処理機能を実行可能な従来の多目的プロセッサである。一実施形態において、パーソナルコンピューティング装置103は、命令をパーソナルコンピューティング装置103に提供するためのジェスチャ認識特徴を含み、コンピュータディスプレイ108において関連する変化を達成する一以上のコンピュータプログラムにアクセスするコンピュータ読取可能な媒体を備えている。例えば、コンピュータゲーミングアプリケーションは、コンピュータシステム103にユーザのジェスチャに基づくコマンドの入力を受け、スクリーン108に表示されるようにゲームを制御させることができる。他の実施形態において、手話認識アプリケーションは、ヒトに従来の手話を用いてパーソナルコンピュータ103とコミュニケーションをとらせることができる。パーソナルコンピューティング装置103は、限定されることなく、デスクトップコンピュータ、ポータブル又はラップトップコンピュータ、携帯情報端末(PDA)、携帯電話、ウェアラブルコンピュータシステム等を含んでもよいことに留意されたい。20 例えれば、一実施形態において、パーソナルコンピューティング装置103は、例えば、カリフオルニア州サンタクララのインテル株式会社製のインテル画像処理ライブラリのような画像処理ライブラリを備えており、インテル株式会社製の1.2GHzペンティアムIIIプロセッサ上で一秒につき14フレーム(320×240ピクセル)で動作する。30

## 【0016】

図1(B)を参照すると、ロボットのマンマシンインタラクションシステムが示されている。一実施形態において、人型ロボットシステム100bは、撮像装置102と、バックパック101内に設けられた一以上のデータ処理ユニット及びデータ記憶装置と、を備えている。このようなシステムの一例が、日本国東京都の本田技研工業株式会社製の先進的人型ロボットASIMO(登録商標)である。この実施形態において、撮像装置102は、人間を識別してジェスチャを制御するための画像データを取得することができるように構成されている。一実施形態において、深度検出カメラは、ロボットの視覚ユニット及びロボットの制御ロジック内のモジュールを含むデータ処理ユニットの一部である。処理ユニットは、以下に説明する技術を実行して人間のジェスチャを分離して動作用のコマンドを認識する。40

## 【0017】

この説明から分かるように、他の多くのアプリケーションが、多くの人間-コンピュータジェスチャに基づくインタラクションシステム、例えば、ドライバ又は同乗者が車両50

に命令を提供する車両アプリケーション、手話翻訳システム等を作成可能である。例えば、さらに他の実施形態において、深度検出カメラは、例えば、運転者アシスタント特徴又は他の車両安全特徴を制御するために、車両内のデータ処理制御ユニット内の計算モジュールに接続された車両のキャビンに設置される。このように、一般的に、データ処理システムは、マイクロチップコントローラ、搭載型プロセッサ等からワークステーション、パーソナルコンピュータ型システム又はネットワーク分散型処理システムまでを含むあらゆる処理システムであってもよい。

#### 【0018】

図2を参照すると、マンマシンインタラクションシステムの一実施形態の機能ブロック図が示されている。この実施形態において、画像データ取得システム102は、データ処理システム104に接続されている。データ処理システムは、データ記憶装置147を備えていることが望ましい。しかし、データ記憶装置147は、処理システム104に対して外付け又はリモートであってもよい。処理システム104は、出力装置106にも接続されている。一実施形態において、出力装置106は、データ処理システム104に外付け又は内装された、追加的なデータ処理モジュールである。例えば、出力装置106は、グラフィカルユーザインターフェイス、ディスプレイ、ジェスチャ - コマンド変換モジュール等であってもよい。

#### 【0019】

一実施形態において、画像データ取得装置102は、深度検出技術に基づいている。深度検出技術は、飛行時間原則に基づいている。この技術を用いたカメラは、飛行時間カメラとして知られている。本実施形態において、カメラ102の画像取得方法は、飛行時間カメラを用いた動的検出に基づいている。動的な深度検出は、赤外線照明のパルスを用いて実行される。パルスがターゲットに照射され、センサがターゲットから反射したエコー信号を読み取る。さらに、カメラ102は、信号がセンサに入るターゲットに対する検出領域を制御するための高速シャッタを有する。遠いオブジェクトに関して、エコーは、シャッタが閉じた後にのみカメラ102内のセンサに到達する。したがって、対応するオブジェクトに関しては信号が記録されない（したがって、ほとんどの背景のオブジェクトは記録されたシーンから消える）。深度検出カメラ102のこの特徴は、深度ウインドウとして呼ばれる動作の幅（例えば、Dmin - Dmax）を提供する。一実施形態において、深度情報は、この動作の幅すなわち深度ウインドウ内のオブジェクト（例えば、人間）に関してのみ記録される。深度ウインドウを超えた他のオブジェクトは、深度データを生成しない。すなわち、シャッタが閉じた後に反射した信号がカメラに到達する。他の実施形態は、Dmax距離を超えたオブジェクトに関する深度値がゼロ深度から始まる周期的な深度値を用いる。この実施形態において、深度値の繰り返しに基づく画像データの前処理は、不要な背景関連データが消えることを保証する。この幅の特徴は、Dmaxを撮像された対象のすぐ後ろに設定することにより、背景雑音を消している（すなわち、Dmax深度を超えたオブジェクトの存在が効果的に消されている）。

#### 【0020】

深度カメラ102は、一般的に深度画像データとして呼ばれるハイブリッドな画像データを提供する。このハイブリッドな深度画像データすなわち深度映像は、リアルタイムの同一の光軸を用いて深度及びカラー / グレースケール情報を同時に取得することにより生成される。画像レートは変化するが、システム100の一実施形態は、1秒につき14フレームでビデオストリームを作動させる。この考察の文脈において、カラー情報は、「黒及び白」（グレースケール）データと、一般的にカラーデータとして呼ばれるもの（例えば、RGB / カラー）と、を含むものとする。例えば、図4を参照すると、一実施形態に係るハイブリッドな深度 - カラーシーン画像が示されている。画像400は、カラー情報サブフレーム400aと、深度情報サブフレーム400bと、を含んでいる。ハイブリッドな画像400は、従来のデジタル画像、すなわち、「カラー」（又はグレースケール / 黒及び白）と同様であり、関連情報が全てのピクセルに関して提供される。ハイブリッドな画像400の深度情報パート（サブフレーム400bとして図示）は、各ピクセルに

10

20

30

40

50

関する深度値を含んでいる。深度データを視覚的に表すために、画像のネガバージョンは、深度の各値（例えば、カメラまでの距離）に対応する明るさの各レベルを用いて生成され、深度サブフレーム 400b により示されるように、近いオブジェクトほど明るく表され、背景のオブジェクトは暗くなる。

#### 【0021】

立体システムと比較して、深度カメラシステム 102 は、いくつかの改善された特性を備えており、例えば、環境がセンサのパルスを用いて同一波長の光を得られない限り、画像が室内環境で明るさ不变であり、深度のエラーは、深度ウインドウ 0.5 ~ 3m に関してはほぼ 5 ~ 15 nm であり、深度情報は、分けられた深度画像パート 400b (1ピクセルにつき 8 ビット) に記録される。このことが、ジェスチャ認識に用いられる手とボディとの間の有意な分離を提供し、ハイブリッドな検出、すなわち、深度及び色画像の両方がリアルタイムスピード（例えば、30 フレーム / 秒）で取得され、（壁のような）テクスチャを有しないオブジェクトの深度が得られる。一実施形態において、これらの改善された特徴及び他の特徴の全ては、複数のカメラのステレオ視システムに反して單一カメラシステム 102 に提供される。單一カメラシステム 102 は、従来のステレオ視複数カメラシステムの二以上のカメラ間の較正を不要にするという追加的な利点も提供する。10

#### 【0022】

一実施形態において、画像データ取得システム 102 は、例えば、イスラエルの Yokneam 又はカリフォルニア州サンホセの Canesta, Inc. の 3DV システムにより製造されたカメラのような飛行時間カメラを備えていることが望ましい。飛行時間カメラは、追跡される人間が動き回る範囲内のエリアを含む深度ウインドウに設定されている。非対象オブジェクト、例えば、背景オブジェクトは、深度画像情報を構成しない。20

#### 【0023】

図 2 に戻って参考すると、一実施形態において、データ処理システム 104 は、画像データ取得システム 102 から画像データの入力を受けるように構成された入力バッファ 141 を備えている。例えば、入力バッファ 141 は、周辺インターフェースハードウェア、ネットワークインターフェースハードウェア（有線又は無線）等に備えられていてもよい。好ましくは、画像データは、入力バッファ 141 を介して入力を受け、内部データ記憶装置 142、例えば、フラッシュ、RAM 等に記憶される。30

#### 【0024】

さらに、データ処理システム 104 は、いくつかのコンピューティングモジュールを備えている。例えば、一実施形態において、コンピュータプロセッサは、一以上のソフトウェアプログラムからの命令の実行を通じていくつかの機能モジュールを実行することができる。一実施形態において、画像データ前処理モジュール 143 は、内部データ記憶装置 142 の画像データにアクセスする。画像データ前処理モジュール 143 は、画像処理アルゴリズムを画像データに適用し、結果的に得られた候補画像データオブジェクトを内部データ記憶装置 142 に記憶させる。画像処理アルゴリズムは、形状決定、動き決定等を含んでいてもよい。例えば、画像データ前処理モジュール 143 は、未加工の画像データを処理してデータ処理システム 104 内の他のモジュールが関心を持つ画像の領域を決定する。一実施形態において、画像データ前処理モジュール 143 は、取得された画像内において、人間の形状、例えば、ボディ、頭、腕、手等に対応するピクセル領域を決定し、他のモジュールが処理することができるように標識決定領域を提供する。40

#### 【0025】

データ処理システム 104 内の形状マッチングモジュール 144 は、画像データ及び候補データオブジェクトにアクセス可能となるように、内部データ記憶装置 142 に接続されている。形状マッチングモジュール 144 は、形状データベース 148 にも接続されている。形状データベース 148 は、データ記憶システム 147 内に記憶されている。例えば、一実施形態において、データ記憶システム 147 は、一以上のディスクドライブ内に実装されていてもよい。

#### 【0026】

50

形状マッチングモジュール 144 は、ユーザにより与えられたジェスチャを決定するために画像データ内の形状を識別することができるよう構成されている。例えば、画像前処理モジュール 143 を欠く実施形態において、形状マッチングモジュール 144 は、人間のボディ、頭、腕、手及び他の形状を検出するアルゴリズムを備えていてもよい。他の実施形態において、形状マッチングモジュール 144 は、特定の形状を検出し、検出された形状と、形状データベース 148 の形状又は形状プロファイルと、をマッチングすることができるよう構成されていてもよい。例えば、一実施形態において、形状マッチングモジュール 144 は、画像内の人間の手を検出し、人間の手の形状と、手の形状のデータベースとを比較し、標準的な手話、手のコマンドサインデータベース等と関連する手のサインに対応する所定の形状を見つけ出す。形状マッチングモジュール 144 は、追加的な入力データを記憶された候補画像データオブジェクトに提供し、追加された一つを生成し、又は、そうでなければ、他の処理システムモジュールにより用いられた出力を提供する。  
。

#### 【0027】

好ましくは、データ処理ユニット 104 は、ジェスチャマッチングモジュール 145 も備えている。ジェスチャマッチングモジュール 145 は、入力を受けた画像データ及び候補画像データオブジェクトを求めて内部データ記憶装置 142 にアクセスする。また、ジェスチャマッチングモジュール 145 は、記憶されたジェスチャ情報プロファイルと候補画像データオブジェクトとを比較するためにジェスチャデータベース 149 にアクセスする。ジェスチャマッチングモジュール 145 は、出力されるジェスチャ認識結果を出力バッファ 146 に提供してもよい。一実施形態において、ジェスチャマッチングモジュール 145 は、データ処理システム 104 の他のモジュールによってアクセス可能となるように、出力を内部データ記憶装置 142 に記憶させる。例えば、一実施形態において、ジェスチャ - コマンドマッチングモジュールは、ジェスチャ認識出力にアクセスし、それを対応するマシンコマンド、例えば、ゲーミングアプリケーションのゲーム制御コマンド、人型ロボットシステムに対する動きコマンド、サウンドシステム（例えば、車両内）に対するボリューム制御コマンド等にマッチングする。  
。

#### 【0028】

図 3 を参照すると、ジェスチャ認識方法の一実施形態のフローチャートが示されている。この実施形態において、シーンのスナップショットを表す画像データが入力される（302）。好ましくは、シーンは、ジェスチャコマンドを提供する人間のサブジェクトを含んでいる。ジェスチャは、いくつかの関連する形状と、シーン内における位置と、動きと、を含んでいる。例えば、手のサインは、最終的な手のサインまでの変化の動きに関する初期の手のサインを含む。さらに、手のサインは、サインに独特の意味を与える他のボディパート（例えば、頭又は胴体）に対する相対位置を含む。  
。

#### 【0029】

##### 1. 画像前処理

この実施形態において、入力された画像データは、前処理されてシーン内の関心のある対象、例えば、人間を検出する（304）。一実施形態において、人間の胴体及び人間の頭の位置が識別される。いくつかの人間検出技術が用いられてもよい。例えば、一実施形態において、テンプレートを用いたグレー画像に基づく人間検出が用いられてもよい。テンプレートは、人間の形状に対応する、与えられたシーン画像のサブセット又は領域を抽出するために用いられる。  
。

#### 【0030】

他の実施形態において、人間検出技術は、人間の上半身の 3 次元特徴を用い、Karhunen - Loeve（「K - L」）分解を用いてそれらを表す。例えば、この実施形態において、ハイブリッドな深度画像ビデオにおける深度情報は、対象である人間の位置を決定するために用いられる。前記した Dmin / Dmax 技術がシーンの範囲を制限するので、特別なカラー、グローブ等が対象を識別するために必要とされない。  
。

#### 【0031】

10

20

30

40

50

一実施形態において、K-L分解を実行するために、入力された画像データは、マルチトレースの入力データセットとして表される。K-L分解技術は、マルチトレースの入力データセットの遅延ゼロの相互共分散の行列の固有値分解をトレースする非相関の（直交する）主成分のセットを算出する。入力データにおける最もコヒーレントなイベントは、いくつかの最も強い成分に対応する、（主成分分析すなわち「PCA」に基づく）主成分トレースを合計することにより較正される。K-L分解は、コンパクトな形態で複数のデータ点を表すことができる。

### 【0032】

以下のPCA手順は、本発明の一実施形態に係る見かけ上の固有空間を算出する一例である。 $m$ 次元の列ベクトル $x$ を定義する。ここで、 $m = p \times q$ であり、 $p, q$ は、上半身の画像サイズである。サンプル画像のシーケンスから観測ベクトル $\{x_1, x_2, \dots, x_n\}$ を形成する。

### 【数1】

$$\begin{aligned} u &= \frac{1}{n} \sum_{i=1}^n x_i \\ \Sigma &= \frac{1}{n} \sum_{i=1}^n (x_i - u)(x_i - u)^T \end{aligned}$$

に基づいて、平均 $u$ 及び共分散行列 $\Sigma$ を算出する。共分散行列 $\Sigma$ の最大の $K$ 、固有値 $\lambda_j$ 及び対応する固有ベクトル $v_j$ （ $j = 1, 2, \dots, K$ ）を計算する。

画像ベクトル $x$ に関して、 $K$ 次元射影計数ベクトルが、 $v_j = v_j^T(x - u)$ により計算される。

### 【0033】

異なる説明によると、このPCA手順は、トレーニングセットを抽出して射影行列を得るために一実施形態において用いられる。図5は、一実施形態に係る横から前までの視覚の全範囲にわたる対象のトレーニング画像のサンプルセットを示す。例えば、トレーニングセットは、ユーザに命令を提供して様々なポーズ角度でポーズをとらせるトレーニング期間において取得される。このように、ジェスチャ認識システムは、ユーザに対してカスタマイズされており、例えば、画像認識は認証目的のために用いられる。一実施形態において、グレースケール及び深度データ画像は、K-L分解が実行されるトレーニングセットとして用いられる。トレーニングアルゴリズムは、（図5に示される角度の任意の範囲内に割り当てられる）既知のポーズ角度を有するM個のトレーニング観測結果 $\{x_1, x_2, \dots, x_M\}$ に基づいており、各観測結果画像 $x_i$ は、 $N = K \times L$ の長さの列ベクトルとして表される。トレーニングアルゴリズムは、以下のステップを含む。

#### 1. トレーニングフェーズ

入力：M次元画像空間における画像データ点 $\{x_1, x_2, \dots, x_M\}$

出力：

### 【数2】

$$\bar{U}_{N \times P}$$

（推定フェーズで用いられる射影行列）

ステップ1：平均ベクトル

### 【数3】

$$\bar{x}$$

及び共分散行列 $(C)$ を構成する。

10

20

30

40

【数4】

$$\bar{x} = \frac{1}{M} \sum_{i=1}^M x_i, \quad C = \frac{1}{M} \sum_{j=1}^M (x_j - \bar{x})(x_j - \bar{x})^t$$

ステップ2：PCA分析を実行する。

【数5】

$$C = \frac{U}{N \times N} A \frac{U^t}{N \times N} \approx \frac{U}{N \times P} A \frac{U^t}{P \times P}$$

ここで、Cの最大のP固有値に対応する第一のP固有ベクトルは、固有空間の最適な線形近似として選択される。

10

ステップ3：推定

各観測結果  $x_i$  に関し、P次元固有空間に射影することができ、すなわち、

【数6】

$$w_i = \frac{U}{P \times N} \left( x_i - \bar{x} \right)$$

2. 推定フェーズ

入力： $x_i$

【数7】

$$U_{N \times P}$$

20

及び新規の観測結果  $y$

出力： $y$  のポーズ

ステップ4：新規の観測結果  $y$  に関するポーズ推定

同様に、新規の観測結果は、P次元固有空間に射影される。

【数8】

$$z = \frac{U}{P \times N} \left( y - \bar{x} \right)$$

最近傍のポーズ

【数9】

$$w = \min_w \|z - w\|.$$

30

を得ることにより決定する。

【0034】

前記した処理は、シーン内の人間のボディの位置を決定するためのポーズ推定における次に利用されるデータベースに記録される対象特異性の画像のセットを生成する。この処理は、既知のポーズ又は向きを用いた対象特異性の画像の取得を必要とする。例えば、トレーニング期間中に、図5に例示するように、対象である人間は、周期一杯である極端な向き（例えば、左向き）から他（例えば、右向き）まで回転する。一実施形態において、このことは、人間に対して、画像データ取得システム102が既知の向きを用いてサンプルトレーニング画像を取得しているシステムの初期化中に、カメラの前で左から右を向くように命令することにより達成される。

40

【0035】

このアプローチを用いて、K-L分解に基づき、一実施形態において、96の深度画像（大きさ  $50 \times 50$ 、16視野、6スケールにわたって撮影）は、深度テンプレートとして用いられる10の固有ベクトルにより表される。様々な人間の深度画像は、これらの深度テンプレートの線形結合により表される。与えられた画像に関して、システムは、まず、y方向において全てのピクセル値を射影し、人間のボディの検出のための前記したK-L分解のステップを予め設定された閾値を超える位置に適用する。

【0036】

50

一実施形態において、与えられたシーン 400において一旦人間が検出されると、画像データに関するさらなる分析が実行される。次のステップは、主なボディパーツの検出を含む。人間が立っており画像データ取得装置 102に向いていると仮定すると、このように立っているポスチャの人間の頭（すなわち顔）は、例えば、図 6 に示されるような射影手法を適用することにより識別される。サインコマンドがボディ及び頭に対してどこに与えられているかによっていくつかのサインは異なる意味を有するので、ユーザの頭の検出は重要である。

#### 【0037】

図 7 は、一実施形態に係る領域のセットにおける画像シーン分割の一例を示す。画像シーン 400 は、8 つの領域 702 に分割される。各領域 702 は、ジェスチャ又はサインに関するいくつかの特定の意味を有している。例えば、手話に関し、「考える」ジェスチャに関する始めと終わりの位置は、領域 2 702b の範囲内に收まり、「空腹」に関するジェスチャは、領域 4 702d に制限される。さらに、頭のボックス 704 及び胴体のボックス 706 は、フレーム 400 に関して定義され、胴体及び頭のゾーンを決定するのをアシストする。すなわち、システムは、与えられたアプリケーションによって、頭及び胴体が定義されたボックス内に收まると予測する。

10

#### 【0038】

したがって、一実施形態において、頭、胴体、腕及び手の一以上の位置は、次の処理に関する画像シーン内で見つけられる。一実施形態において、候補画像データオブジェクトは、位置フィールドのセットを用いて生成される。例えば、データオブジェクトは、頭及び/又は胴体に対する左右の手の位置を含んでいる。一実施形態において、手が位置する領域の番号すなわち識別子は、候補画像データオブジェクト内のフィールドに入れられる。

20

#### 【0039】

領域 702 は、ジェスチャに関連する一般化された動きを定義するために用いられる。例えば、一実施形態において、フレーム 400 のセットにわたって取得されたジェスチャは、特定の手の始めの領域と終わりの領域とを用いて候補画像データオブジェクト内に表される。例えば、右手は、領域 2 702b で始まり領域 5 702e で終わり、右上部から左下部への移行すなわち動きを表す。さらに、ジェスチャに関して、この一般的な動きすなわち移行は、ジェスチャ検出システムが識別可能な特別な意味を有する。

30

#### 【0040】

##### A. 手の位置の検出

手検出手法の一実施形態は、前記した人間検出の結果を利用する。他の実施形態は、カメラまでの距離に関連するヒューリスティックに基づいてシーンの各フレーム内の手の位置を決定するために深度画像処理モジュールに依存しており、例えば、サインを行う手は、カメラに向かう人間のパーツに最接近する。

#### 【0041】

一実施形態において、画像データの前処理中に得られた候補画像データオブジェクト内に記憶されたデータは、例えば、前処理モジュール 143 が手が存在すると決定した領域 702 を分析することにより、手を検出するのに用いられる。他の実施形態において、前処理は、ボディ及び頭を含む画像内の領域を簡単に決定する。与えられた画像から、ボディに対応する位置が手検出手法中に引かれる。この実施形態において、サインを行う手は、大抵の場合ボディの前又は横に位置するので、有意なサイズの残余の塊を分析することにより検出される。残余画像の平均ピクセル強度は、手の部分を選択するためのカットオフ点として用いられる。このヒューリスティックにより生成された画像が手と同様に腕部分を有する場合がある。このような場合には、腕に対応する部分は、領域占有に基づいてピクセルを選択することによりさらに消される。

40

#### 【0042】

図 8 を参照すると、一実施形態に係る手のジェスチャを与えるユーザの深度画像が示されている。フレーム 802 は、システム画像データ取得システム 102 の深度画像データ

50

により取得された全体的なシーンを示す。Dma×領域を超えた背景は、画像データ、すなわち、近似的にDma×の深度値を有するピクセルの部分とならない。ユーザにより提供されたサインを認識するための第一のステップは、画像804により示されるように腕を分離することである。ジェスチャ認識システムの一実施形態において、腕画像抽出は、以下を含む。

#### 【0043】

ステップ1：腕抽出

##### 粗抽出

手が胴体よりもカメラに近い距離ある場合には、腕が深度差により抽出される。手が伸びている場合には、手は、顔及び胴体領域を消去した後に抽出される。このステップの結果が腕画像804である。同時に、顔及び胴体領域ボックスと当該領域情報とが、図7で前記したように識別されて記録される。例えば、図8に示されるシーンにおいて、候補画像データオブジェクトは、右手が図7の領域4 702dに位置することを示す。

10

#### 【0044】

正確な腕抽出

続いて、粗抽出ステップで得られた腕画像から指先が検出される。そして、腕の不要なパートは、アスペクト比により削除される。例えば、腕画像の幅又は高さが顔幅の1.5倍よりも大きい場合には、画像はこの点でトリミングされる。このステップの結果が掌画像806である。二つの手のジェスチャ画像に関して、二つの掌画像が同様に得られる。

#### 【0045】

##### 2. 手形状分析

一実施形態によると、本システムは、記憶された形状データに未知の対象候補形状にマッチングさせる方法を適用して、対象候補形状を識別する。考察の明確性のために、人間の手の形状は好ましい実施形態に関して考察されるが、同様の技術は、HCI、例えば、握手サイン又はオブジェクト、全体的なボディの位置、キック、パンチ等に関して用いられるあらゆる形状に適用可能である。

20

#### 【0046】

手形状分析の複雑さのレベルは、アプリケーションによって異なる。いくつかのアプリケーション、例えば、マウス操作の物真似は、いくつかの手の形状、例えば、右／左クリックのみを必要とする。他方、他のアプリケーションは、もっと複雑な形状及び動きの分析及び認識、例えば、手話を必要とする。一例によると、本発明の実施形態の記述のいくつかは、日本手話（「JSL」）で用いられる基本的な手のパターンに関して記載されている。JSLは、繰り返し用いられる基本的な手の形状を多く有している。一実施形態に係る日本手話の手のサインのいくつかの例に関し、図9を参照する。形状901, 902は、基本的に異なる向きの同一形状であることに留意されたい。

30

#### 【0047】

手の形状は、ボディに対してどこに見えるか、向き及び軌道に応じて異なる意味を有する。さらに、他の手と結合されている場合には、さらに他の異なる意味を有する。このように、言葉の組み合わせの数は、様々な要因に基づいて生成される。例えば、図9における形状903は、「JSLではそれだけで「私」「あなた」「来る」「行く」「明日」「昨日」「どちらの」「何の」「分」「これ」「考える」「時々」「次」を含む多くの意味を有する。

40

#### 【0048】

意味の多様性に対応するために、一実施形態において、要因は、位置、向き、手の形状のタイプ（指を数える）及び軌道（初期及び最終位置）を含んで分析される。位置に関して、一実施形態において、本システムは、図7に示すように空間のラフな区切りを用いる。さらに、一実施形態において、例えば、後記する骨格マッチング技術を用いて、取得された手の様々な特性が分析される。例えば、図11（C）に示すように、指の分析は、幅、向き及び長さを含む。このような分析で収集される指情報プロファイルは、以下のサンプルを含んでいてもよい。

50

始点の x / y 座標 : ( 7 , 19 )  
 終点の x / y 座標 : ( 44 , 18 )  
 全体の幅 : 9.7642  
 スリムな幅 : 7.8002  
 指の幅 : 30

#### 【 0049 】

さらに、一実施形態において、サイン認識に関して記憶された形状同士を区別するために、指の角度のような追加的な要因が測定される。指の骨格の二つの枝間の角度が決定される。例えば、図9のサイン905, 908に関する枝の角度は、サイン「V」「L」であると簡単に区別することができる。同様に、904, 911は、骨格の二つの枝間の距離を測定することにより区別される。さらに、他のサインは、分類に関するさらに詳細な分析を必要とする。詳細な分析が増加するのと同様に、認識セットの手の形状の数も増加する。

#### 【 0050 】

本明細書に記載された様々な実施形態に応じて用いられる形状マッチングのためのいくつかの方法が存在する。一般的に、候補形状及びその関連情報は、候補形状データオブジェクトに構成されていて、形状データベース内に記憶された形状プロファイルのセットと比較されてもよい。好ましくは、記憶された形状プロファイルは、候補形状画像に対する比較、マッチング及びフィルタリングに関連する記憶された形状についての情報を含んでいる。

#### 【 0051 】

一実施形態において、二つの画像間のChamfer距離（「CD」）が用いられ、これは形状の類似度を測定するためにも用いられる。CDは、形状比較のための距離変換を用いて定義される。距離変換の結果が、エッジに対応する各ピクセルがゼロ（0）を含み、一方で、他のピクセルが最近傍のエッジ点までのCD距離値を含む勾配画像である。

#### 【 0052 】

他の実施形態によると、K-L分解が形状マッチングのために用いられる。K-L分解アルゴリズムは、CD法よりも優れたいくつかの利点を有している。両手法は、画像プロファイル内のテンプレートとしての多くの画像を含む。ボディパーツ検出アルゴリズムに関する前記した一つと同様のアプローチに基づく同様のテンプレートは、形状画像テンプレートを生成するために用いられてもよい。前記したように、K-Lアルゴリズムは、必要な固有ベクトルの種類を記録するために用いられる。逆に、CD法は、K-Lアルゴリズムで用いられるベクトル形式と反対に画像フォーマットで各テンプレート画像を記録する。両技術は、一実施形態において適用されているが、K-Lアプローチは比較データに関して必要とされる記憶容量を効果的に低減することができる。形状データベース148内のトレーニングシーケンスは、K-Lに基づく実施形態においてより小さくすることができる。

#### 【 0053 】

候補形状画像のトレーニングシーケンスプロファイルとの比較のための類似度測定に関して、K-L法は、入力された画像を固有ベクトルに射影するのに対し、CD法は、テンプレート画像が好適に整理されない場合に、入力された画像と全てのテンプレート画像とを比較する。このように、どの方法が用いられても、CD法の計算負荷は一般的に重い。

#### 【 0054 】

##### A. 距離マッチング技術

一実施形態において、手の形状は、異なるジェスチャに関する様々な手の形状を含む記憶されたデータセットを基準として分析される。例えば、図10は、一実施形態に係るサインの記録すなわちプロファイルのセットと対応する画像の手のパターンとを示す。サインプロファイルは、サインの定義と追加的なサイン識別データ、例えば、開始及び終了領域、動き等とを有するデータレコード1002を含んでいる。さらに、サインレコード1002は、手話に関連する一以上の手の形状の画像を含む画像ファイル1004を含む又

10

20

30

40

50

は提示する。例えば、サインレコード 1002a ~ 1002e は、対応する画像 1004a ~ 1004e を提示する。しかし、数字の 2 を表すサインに関して、サインレコード 1002b は、二つの画像パターン 1004b(1), 1004b(2) を提示しており、これらは、同等なものとして許容可能なマッチング画像パターンを提供する。さらに、動き及び形状変化を必要とする他のもっと複雑なサインは、サインに関連して初期、中間及び最終画像パターンに関する画像のセットを提示し、例えば、「ついてこい」サイン 1002f は画像レコード 1004fにおいて 3 つの画像を提示し、画像レコード 1004f は、1004f(1), 1004f(2), 1004f(3) を含んでいる。

#### 【0055】

記憶された手の形状データに対して未知の手の形状をマッチングするために、二つの画像間の Chamfer 距離（「CD」）が形状の類似度を測定するために用いられる。CD は、形状比較のための距離変換を用いて定義される。距離変換の結果は、エッジのペアに応する各ピクセルが 0 を得、他のピクセルが最も近いエッジ点までの距離を得る勾配画像である。未知の画像 X とテンプレート Y との間の CD を算出するために、我々は、X のエッジ画像と Y の DT との間の相関関係を得る。

#### 【0056】

その値が小さいほど、類似度は高まる。我々は、このような類似度測定値 S を  
【数 10】

$$S = \sum_i \sum_j E(i,j) DT(i,j)$$

として定義する。ここで、E は X のエッジ画像を表し、DT はテンプレート Y の勾配画像を表す。我々は、テンプレートデータセット内の全ての形状に関する距離変換された画像を記憶する。未知の手の形状に関しては、前記した類似度計算が行われ、最も小さい値を有するパターンがこの手に関するパターンとして選択される。一時的な投票が、シーケンスの特定の画像により生成されるバイアスを低減させるための決定に用いられる。他の実施形態において、CD 手形状マッチング技術は、マッチングを迅速に処理して簡略化するために、指計数アルゴリズムを候補画像の前フィルタリングとして用いて補強される。

例えば、サインレコード 1102 は、指の数と関連している。指計数アルゴリズムが候補画像における指の数を決定した後、指の数に関連したレコードとのみ比較される。

#### 【0057】

#### B. 骨格マッチング技術

図 11(A) ~ 図 11(C) に参照すると、一実施形態に係る画像と手形状マッチング技術に基づいて骨格を表す、対応する処理された画像データとが示されている。一実施形態において、前処理が対象の手に対応する画像データを提供した後、人間 - 機械ジェスチャ認識システムは、手の画像データを分析する。この実施形態において、骨格マッチング技術に基づく手形状分析は、以下のステップを含んでいる。

#### 【0058】

##### 骨格生成

図 11(A) に示す掌画像は、バイナリ画像を得るために閾値であり、バイナリ画像のエッジは、従来の骨格生成技術により平滑化されて小さい枝に低減される。このステップの結果が、図 11(B) に示される骨格である。

#### 【0059】

骨格が得られた後、骨格画像の終点及び交点が骨格の指の全ての線セグメントについて識別され、終点及び交点によって定義される。終点 / 交点対が画像内で決定される。各点対の点間のユークリッド距離が記録される。ユークリッド距離が非常に小さい場合には（例えば、1 ~ 2 ピクセル、画像サイズに応じた閾値）終点及び交点対間の線セグメントは消去される（すなわち、小さすぎて手の実際のエレメントとしては扱われない）。全ての消去動作の後、終点及び交点を再び識別する。このステップの結果が、図 11(B) に示される指骨格を生成する閾値よりも長いユークリッド距離を有する終点及び交点対である。

**【 0 0 6 0 】**

全ての終点及び交点対に関して、それらの間の骨格画像における経路を得る。そして、図11(C)に示されるバイナリ画像におけるこの経路の幅、長さ及び向きの角度を計算する。各経路の二つの幅パラメータが存在し、一つは、経路の平均幅(width-1)であり、他は、幅が(掌の幅により決定された)幅閾値よりも小さい副経路の平均(width-2)である。一実施形態において、このステップの結果が、例えば、終点座標、交点座標、向き、幅、長さ等のような各経路についての特性を記録するデータ構造である。

**【 0 0 6 1 】****指認識**

一実施形態によると、骨格における全ての経路に関して、経路の長さ、width-1、width-2が決定され、width-1及びwidth-2は、例えば、指候補のそれぞれの全体の幅及び細い幅である。さらに、一実施形態において、測定された特性に基づいて、経路が指に相当するか以下かについて判定される。このステップの結果が、指の数を含み、データ構造が各指についての有意なデータを記録するために用いられる。例えば、候補指又はジェスチャに関する候補画像データ構造における指情報セクションは、指情報を記録するために用いられる。10

**【 0 0 6 2 】**

骨格マッチング技術は、回転不变である。したがって、異なる角度の二つの手は、指計数及び手の相対角度に基づいて同様に認識可能である。一方、CD法を用いると、異なる角度の手は同様には決定されない。しかし、骨格に基づく方法は、計算負荷のコスト面でCD法よりも優れている。一実施形態において、手形状分析に関して、複数の指先の実時間追跡も、従来の骨格アルゴリズムの指計数パートを補強するために用いられる。20

**【 0 0 6 3 】****C. 形状情報統合**

一実施形態において、初期に取得されたシーンからの候補対象手の画像を用いて。本システムは、手すなわち掌に関する基本情報を抽出する。この情報は、他のモノ、領域、領域ボックス、向きの角度、顔ボックス、胴体ボックス、掌中心、平均深度値、掌から胴体中心までの距離、掌から顔中心までの距離を含む。この情報は、画像処理技術を通じて決定され、候補画像データオブジェクトに収集される。候補画像データオブジェクトは、前ステップで収集された追加的な情報、例えば、ボディ、頭、手の前処理の位置等を含む。一実施形態において、候補画像データオブジェクトは、サイン又はジェスチャマッチングのために続いて用いられる、決定された掌データを記録するための掌情報セクションを含む。30

**【 0 0 6 4 】**

この実施形態によると、骨格処理から認識された指を用いて、本システムは、指構造を掌の特性として記録する。したがって、表1に示されるマルチレベル構造が、候補画像データオブジェクトに対して定義される。

**【 0 0 6 5 】**

さらに、一実施形態において、いくつかの形状を含む、より複雑なジェスチャをマッチングするために、前記ステップが、ジェスチャに関する全ての画像を通じて繰り返される。この実施形態において、全ての画像における手すなわち掌についての全ての情報は、公報ジェスチャデータオブジェクトに関連する掌の構造に記録される。40

【表1】

表1. 候補画像データオブジェクトの掌セクション			
掌	領域		
	領域ボックス		
	中心		
	型(左又は右手)		
	向き		
	顔ボックス		
	胴体ボックス		
	平均深度値		
	顔までの距離		
	胴体までの距離		
指	指番号		
	Finger_1	始点	
		終点	
		Width_1	
		Width_2	
		指の長さ	
		向き	
	Finger_2	...	
	Finger_3	...	
	...	...	

10

20

## 【0066】

## D. 軌道決定

一実施形態において、一以上の形状が、ジェスチャを定義するための軌道に関連している。例えば、一実施形態において、手は、手に関連する時間情報を有する3D空間における曲線すなわち軌道を生成するように空間内で動く。図12は、一実施形態に係る手の3次元画像の線チャートのサンプルを示す。一実施形態において、ジェスチャマッチング処理は、軌道情報を含む。しかし、マッチングを簡略化するために、軌道情報は、X方向の動き、Y方向の動き及びZ方向の動きの三つのパートに分割される。X及びYは、画像平面における手の水平及び鉛直座標であり、Zはカメラシステムまでの距離を示す手の平均深度値により表される。

30

## 【0067】

一実施形態において、軌道情報を記憶されたジェスチャ軌道情報にマッチングするために、以下の3ステップが実行される。

## 【0068】

1. 各手の動きのX, Y及びZ成分が、(均一のタイムインターバル[0, N]、ここでNはサンプル点、に正規化された)テンプレート曲線にマッチングされる。図13は、一実施形態に係るX, Y及びZ曲線のテンプレートパターンのいくつかの例を示す。曲線パターン1301は、「押して戻るパターン」1301a、「押して止まって戻るパターン」1301b、「押して戻って押して戻るパターン」1301cのようなパターンに対応している。曲線1301は、FFTを用いて円滑化され、円滑化前よりも局所的なノイズを低減されている。

40

## 【0069】

2. 曲線マッチングのために、各次元における有意な動きのみが考慮され、小さい方向変化によるエラーを回避する。例えば、「後退する」ジェスチャに関して、手は、主に胴体から前にカメラ方向に向けて動く。しかし、画像平面における手の水平及び鉛直のいずれも存在し、このようなずれはごくわずかであるので、無視されて考慮から除外される。

## 【0070】

3. 無視できない動きに関して、本システムにより取得された候補曲線は、曲線データセット内の全てのテンプレートと比較される。一実施形態において、最小二乗エラー計算が、候補及びテンプレート曲線間の類似性を算出するために用いられる。例えば、未知の候

50

補曲線  $Y(t)$ ,  $t = 0, 1, \dots, N$  及びテンプレート曲線  $S(t)$ ,  $t = 0, 1, \dots, N$  について、最小二乗エラー（「MSE」）は、以下のように計算される。

#### 【数11】

$$MSE = \sum_{t=1}^N [Y(t) - S(t)]^2$$

計算された MSE が小さいほど、候補及びテンプレート曲線間の類似性の度合いは高くなる。一旦全てのテンプレートが候補と比較されると、最小の MSE に対応するテンプレートが対応するジェスチャの識別のための候補に選択されて指定される。

#### 【0071】

例えば、静的ではない手話の画像の一連のフレームを示す図 14 (A) を参照すると、一実施形態において、静的ではないサイン又はジェスチャについて、本システムは、一連のフレームにおける特定の指の始点及び終点（例えば、ストロークの極点）を記録する。手の軌道は、図 14 (B) に示すように、一連のフレームの始点から終点までが X, Y 及び Z 方向において記録される。10

#### 【0072】

##### E. 手の向きの決定

一実施形態において、手の向きも取得される。例えば、図 14 (C) に示すように、指の向きは、いくつかのジェスチャ又はサイン、例えば、サムアップとサムダウンのジェスチャ、について関連性を有している。したがって、一実施形態において、一旦関連性を有する手の形状、例えば、スタートの手の形状とサムアップジェスチャ、が決定されると、対象候補画像形状に対する手の向きが取得される。図 15 は、一実施形態に係る「サムアップ」の手のパターンに関する手の向きの分析結果の曲線 / 画像図を示す。手の向きの情報は、手の形状に関する領域の主軸を分析することにより取得される。例えば、前記した骨格分析に基づいて、腕及び手の主軸が用いられ、頭、胴体又は初期位置について測定された角度変化が画像データから得られる。20

#### 【0073】

##### ジェスチャマッチング

候補画像すなわち対象のフレームの一連の情報が一以上のデータオブジェクト内に収集された後、ジェスチャマッチング処理が認識システムに提供されたサイン又はジェスチャを決定するために適用される。一実施形態において、手の形状の情報及び軌道情報は、画像データにより表されるサイン又はジェスチャを分類するために用いられる候補画像データオブジェクト内に記録される。候補データオブジェクト内に収集された特徴情報は、分析され、記憶されたパターンに関連する情報を含む記憶されたジェスチャ / サインと比較される。ジェスチャ又はサインプロファイル内の情報は、候補データオブジェクト内の情報に対して比較、マッチング又はフィルタリングするために用いられる。例えば、一実施形態において、ジェスチャプロファイルは、データオブジェクト内の n 個の特性のセットとして表され、例えば、5 個の特性値のペアは、 $G = <\text{形状}, \text{位置}, \text{軌道}, \text{向き}, \text{速さ}>$  の形態を有するジェスチャ「G」を表すために用いられ、各ジェスチャについて、これらの特性に関する値のユニークな組み合わせがジェスチャを定義する。30

#### 【0074】

各特性がいくつかの記憶されたトレーニングパターンを区別するので、n 次元空間 G（例えば、 $n = 5$ ）は、数ダースのジェスチャパターンを容易に含むことができる。40

#### 【0075】

一実施形態において、マシンとのインタラクションを意図するジェスチャの認識を容易にするため、ジェスチャ区切りサインがジェスチャの始めと終わりに用いられ、マシンに対して、提供されたジェスチャがジェスチャ認識システムとのコミュニケーションを意図していることを示す。他の実施形態において、ジェスチャを行う手は、ジェスチャ終了後にマシンの視野領域から外れる。このアプローチは、ジェスチャの区切りとしても用いられる。例えば、画像データ取得システム 102 により取得された一連の 300 フレーム（10 秒）において、フレームの最初のセット、例えば 1 ~ 130 の間、手は、Z 軸におい50

て前向き及び後ろ向きに動く。続いて、フレームの次のセット、130～180の間、手は検出されない。フレーム180の後、手は、再び前向きに動き、フレーム200～260で固定された深度値で静止する。このように、フレーム130～180は二つのジェスチャ間の区切りとして解釈され、画像データはそれに応じて処理される。

#### 【0076】

自然で連続的な人間のジェスチャを理解するために、隠れマルコフモデル（「HMM」）のような手法が考慮されるべきである。図16は、一実施形態に係るジェスチャ認識のためのステップを説明するためのフロー図である。画像データが深度ビデオ源から入力を受ける間、手の位置がチェックされ、手が静止しているか動いているかが判定される（1602）。動いている手に関して、動きを含むジェスチャに関する形状が、候補ビデオフレーム内の形状とマッチングするために用いられる（1604）。例えば、手ジェスチャデータベース内の手形状テンプレートが、手の形状をマッチングする（1604）ときに用いられる。手の軌道が、動きの方向（X, Y及びZ）を検証することにより取得される（1608）。一旦候補ジェスチャフレームの軌道が決定されると、前記したように記憶された軌道パターンとマッチングされる（1610）。一旦形状、位置及び軌道が決定されると、ジェスチャマッチング処理が、さらに以下に説明されるように、候補画像データオブジェクトに基づいて行われる（1612）。出力されるジェスチャが、ジェスチャデータベースから決定される（1614）。

10

#### 【0077】

手が動いていない場合には、動きを含まないジェスチャ又はサインが用いられ、前記したように、例えばCD、K-L、骨格または類似した手法で、静止した手の形状の候補画像とマッチングする（1606）。形状マッチング処理（1606）から、サインが記憶されたサインから選択されて決定される（1616）。

20

#### 【0078】

一実施形態において、ジェスチャの軌道は、例えばジェスチャプロファイル内の多くの予め決定されたパターンに分類される。様々な特徴が収集された後、例えば以下に示す表2のような認識テーブルが、対応するジェスチャ又はサインを識別するために用いられる。

#### 【表2】

30

表2. JSL手話情報のサンプル							
ジェスチャ	メッセージ	2日以内	ビット	月曜日	飛行機	タクシー	再び
両手	YES	NO	NO	NO	NO	NO	NO
指先の距離	NO	NO	YES 減少	YES 増加	NO	NO	NO
X方向の動き	NO	NO	NO	NO	NO	NO	YES 増加
Y方向の動き	NO	NO	NO	YES 増加	YES 減少	NO	NO
Z方向の動き	YES 減少	YES 増加	YES 減少	NO	NO	NO	NO
手の形状	L形状	V形状	親指+人差し指	L形状	親指+小指	親指+小指	水平なV
向き	鉛直	鉛直	鉛直	鉛直	鉛直	鉛直～45 -60度	鉛直～水平
形状の変化	NO	NO	NO	NO	NO	NO	拳から水平 なV

#### 【0079】

40

一実施形態によると、本システムは、前記したように掌の構造内的情報をチェックして、このジェスチャが片手のジェスチャであるか両手のジェスチャであるかを判定し、この特性を二つの掌コード変数（0 - 片手のジェスチャ、1 - 両手のジェスチャ）として記録する。同時に、掌の構造における重要な指の数を判定し、例えば各掌の指の数が一である場合には、一本指のジェスチャグループに分類する。指の数に関する変数がここでは用いられる。

#### 【0080】

この実施形態において、ジェスチャがどの指のグループに入るかに關係なく、本システムは、以下の掌についての特性をチェックする。画像の掌における掌のX, Y, Z（深度）、向きの角度を記録する。動きの型をチェックする。

50

0 - 静止 ; 1 - 減少 ; 2 - 増加

1 \* - まず減少し、周期的に増加 / 減少の動き

2 \* - まず増加し、周期的に減少 / 増加の動き

掌の位置の軌道、始めの位置、終わりの位置及び最もよくある位置をチェックする。位置の定義を見るために図1を参照されたい。いくつかのジェスチャは指の長さの変化、例えば40のみを有するので、平均指長さの変化をチェックする。

#### 【0081】

一実施形態において、これらの特性は、座標、座標の変化、位置、指の長さ等のいくつかに分けられた構造に記録される。

#### 【0082】

この実施形態によると、ジェスチャ認識システムは、始めの手の形状、終わりの手の形状、最もよくある手の形状、手の形状の変化コード等を含む手の形状特性を得る。これらの特性は、他のデータ構造である手の形状に記録される。表3は、手の形状の認識のために用いられる特性を示す。

#### 【表3】

表3. 手形状認識		
指の数	手の形状	基準
1	人差し指	$[\text{指の長さ}] / [\text{顔の幅}] > 0.45$
	小指	$[\text{指の長さ}] / [\text{顔の幅}] < 0.45$ , 指の向きは掌の向きに近い
	親指のみ	$[\text{指の長さ}] / [\text{顔の幅}] < 0.45$ , 指と掌との間の向きの差は無視できない
	親指+開いた掌	$[\text{指の長さ}] / [\text{顔の幅}] < 0.45$ , $[\text{掌の面積}] / [\text{顔の面積}] > 1$
2	V形状	$25 < [\text{二本の指間の角度差}] < 50$
	水平なV	水平な指の向き(~0)
	鉛直なV	鉛直な指の向き(~90)
	反転V	鉛直な指の向き(~-90)
	親指+小指	$[\text{二本の指間の角度差}] > 90$
	人差し指+小指	$50 < [\text{二本の指間の角度差}] < 60$
	L形状	$60 < [\text{二本の指間の角度差}] < 90$
3	3本の長い指	$[\text{平均(指の長さ)} / \text{顔の幅}] > 0.45$
	水平	水平な指の向き(~0)
	鉛直	鉛直な指の向き(~90)
	反転	鉛直な指の向き(~-90)
4	人差し指+中指+親指	$[\text{平均(指の長さ)} / \text{顔の幅}] < 0.45$
	親指以外の指	$[\text{平均(指の長さ)} / \text{顔の幅}] > 0.42$
	小指以外の指	$[\text{平均(指の長さ)} / \text{顔の幅}] < 0.42$
5	星形状	_____

#### 【0083】

二本指グループの特別な特性は、二本指間の距離が変化するか否かを含む。これは、例えば、0 - 静止 ; 1 - 減少 ; 3 - 増加として指間距離構造に記録される。

#### 【0084】

一実施形態において、結合されたジェスチャ情報は、例えば表4に示すようなジェスチャデータ構造に記録される。ジェスチャデータ構造は、ジェスチャデータベースに記憶されたジェスチャに関連するジェスチャプロファイルの一部である。

10

20

30

40

【表4】

表4. ジェスチャ情報		
ジェスチャ	二つの掌コード	可能な値: 0, 1
	指の数	可能な値: 1, 2, 3, 4, 5
距離	指間の距離	各掌の二本の指間の距離を記録する配列
	変化コードの距離	可能な値: 0, 1, 2
座標	掌のX座標	X座標を記録する配列
	掌のY座標	Y座標を記録する配列
	掌のZ座標	Z座標を記録する配列
	掌の角度	掌の向きを記録する配列
座標変化	掌のX座標変化コード	可能な値: 0, 1, 2, 1*, 2*
	掌のY座標変化コード	可能な値: 0, 1, 2, 1*, 2*
	掌のZ座標変化コード	可能な値: 0, 1, 2, 1*, 2*
	掌の角度変化コード	可能な値: 0, 1, 2
手の形状	手の形状のリスト	各掌の手の形状を記録する配列
	始めの手の形状	最初のいくつかのフレームで優位な手の形状
	終わりの手の形状	最後のいくつかのフレームで優位な手の形状
	最もよくある手の形状	全てのフレームで優位な手の形状
	手の形状の変化コード	可能な値: 0, 1
位置	始点	最初のいくつかのフレームで優位な位置
	終点	最後のいくつかのフレームで優位な位置
	最もよくある位置	全てのフレームで優位な位置
指の長さ	指の長さの配列	各掌の平均指長さを記録する配列
	指の長さの変化コード	可能な値: 0, 1

## 【0085】

10

## ロボットへの命令

発話の使用は、ロボットへ命令を与える良い手法である。実際にはしかし、現在の発話認識は、ロボットとのコミュニケーションをとるための手法としては望ましくない制限を有している。環境及びロボット由来のノイズにより、（ユーザがロボットから数メートルだけ離れている場合であっても、）人間であるユーザにより生成された音声信号がロボット認識のために有効な程度にクリアではないことがある。

## 【0086】

20

一方、サインは多くの場合で有効である。オフィス環境で従事するロボットに関して、「3歩下がる」のような単純な動きの命令に加えて、我々は以下のフレーズを手で示すことができる。「ミーティングはいつ？」「彼をスミス氏のオフィスに連れて行きなさい」「私の命令を繰り返しなさい」「私のスーツケースを持って私についてきなさい」「ボスの伝言は何？」「お客様にあいさつしなさい」「2階に行きなさい」。一実施形態によると、JSLに基づくサインの語彙は、この種の命令に用いられる基本的な言語をロボットに解釈されることが可能のように提供される。さらに、「もし」「それから」「間」のような任意の記憶された言語を追加することにより、「もしスミス氏が来たら、それから私を呼びなさい」のような文を作ることができる。

30

## 【0087】

例えば、経験的な一実施形態において、動きジェスチャのコマンドセットのセットが、人型ロボットであるASIMOのために開発されている。これらを含むジェスチャが、表5に列記されている。

40

【表 5】

表5. ASIMOの基本的な動きのジェスチャコマンドセット	
3歩前に進む	急ぐ／ゆっくり／レギュラー
5歩後退する	40度回転
動作7	あそこを見る
ついていく	彼の言うことを聞く
繰り返す	2階に行く
5分休憩する	1分後に始める
2分で戻る	もう一度
挨拶する	静かに
捕まえる	彼を思い出す
OK／バイバイ	ミーティングはどこですか？
メッセージは何ですか？	何か新しいことがありますか？
上／下／2日前／2日で	電話をかける

10

## 【0088】

この実験的な実施形態において、（動きのないタイムインターバルのような）区切りで始まるサインコマンドを用いて、表5のジェスチャコマンドが、たったのエラー率5%以下である実験に基づくソフトウェアにより正確に解釈される。

## 【0089】

本明細書では本発明の特定の実施形態及びアプリケーションが説明されたが、本発明が本明細書に開示された厳格な構造及び要素に限定されず、特許請求の範囲で定義される本発明の精神及び要旨を逸脱しない範囲で、本発明の方法及び装置の配置、動作及び詳細において様々な修正、変化及び改変が可能であることが理解されるべきである。

20

## 【図面の簡単な説明】

## 【0090】

【図1】(A)は、パーソナルコンピューティングマンマシンインタラクションシステムの一実施形態を示す図であり、(B)は、ロボットのマンマシンインタラクションシステムの一実施形態を示す図である。

【図2】マンマシンインタラクションシステムの一実施形態を示す機能ブロック図である。

30

【図3】ジェスチャ認識方法の一実施形態を示すフロー チャートである。

【図4】一実施形態に係るハイブリッドな深度 - 色シーン画像を示す図である。

【図5】一実施形態に係る、横から前までの視野の範囲をフルに通るトレーニング画像のセットのサンプルを示す図である。

【図6】一実施形態に係る、前向きの頭位置を決定するために適用される射影方法からの深度成分フレーム及び対応する閾値画像を示す図である。

【図7】一実施形態に係る領域のセット内の画像シーン分割の一例を示す図である。

【図8】一実施形態に係る手のジェスチャを与えるユーザの深度画像を示す図である。

【図9】一実施形態に係る日本手話の手のサインの例を示す図である。

【図10】一実施形態に係るサインレコード又はプロファイルのセットと対応する画像の手のパターンとを示す図である。

40

【図11】(A)～(C)は、一実施形態に係る手形状マッチング技術に基づく骨格を説明するための、画像のセットと対応する処理された画像データとを示す図である。

【図12】一実施形態に係る手の3次元軌道のラインチャートのサンプルを示す図である。

【図13】一実施形態に係るX, Y及びZ曲線のためのテンプレートパターンの例を示す図である。

【図14】(A)は、一実施形態に係る静的ではない手のサインの画像の一連のフレームを示す図であり、(B)は、(A)の一連のフレームに関する始点から終点までのX, Y及びZ方向において記録された手の軌道を説明するピクセルチャートを示す図であり、(

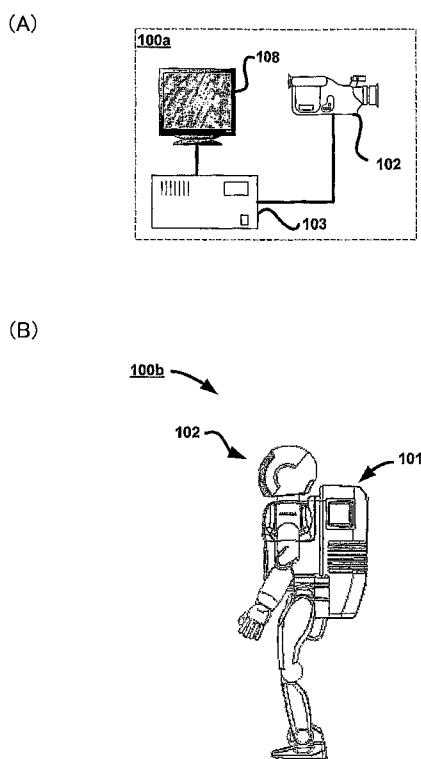
50

C)は、(A)の一連のフレームで追跡する指の向きの角度を説明するための、フレームにわたる向きの角度変化のチャートを示す図である。

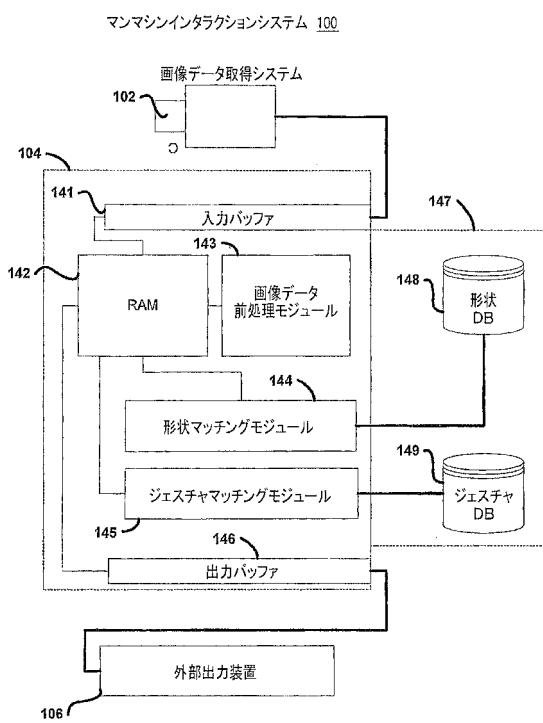
【図15】一実施形態に係る「サムアップ」の手の形状に関する手のむきの分析結果を示す曲線／画像図である。

【図16】一実施形態に係るジェスチャ認識のためのステップを説明するためのフロー図である。

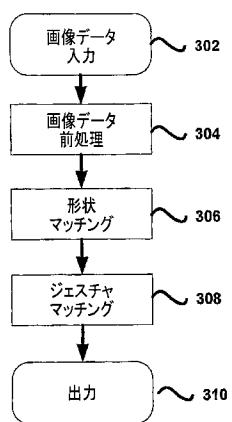
【図1】



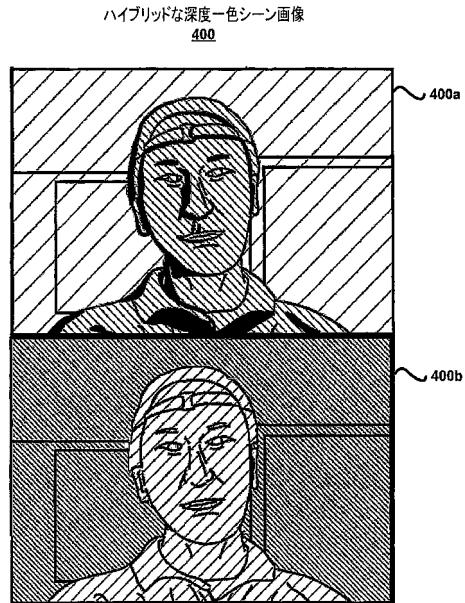
【図2】



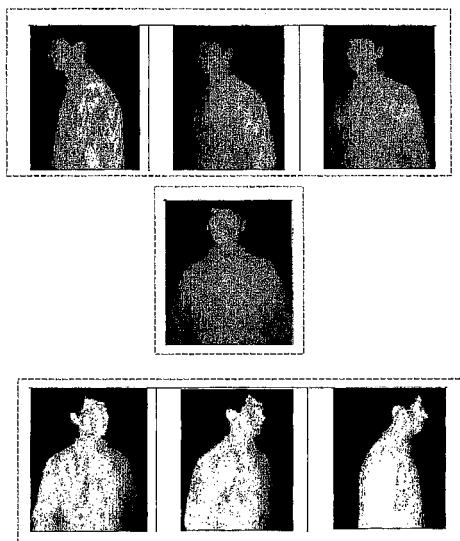
【図3】



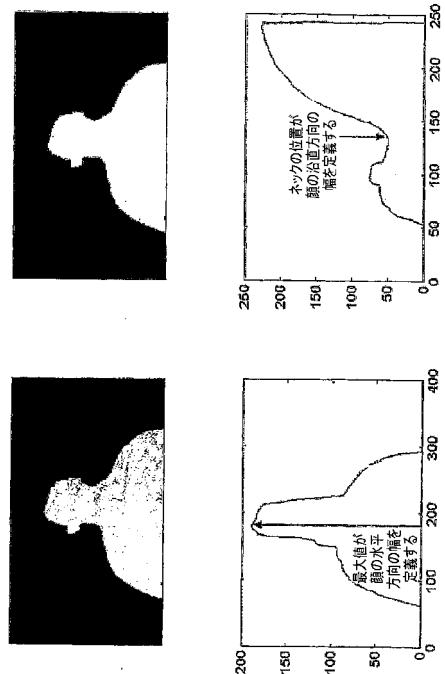
【図4】



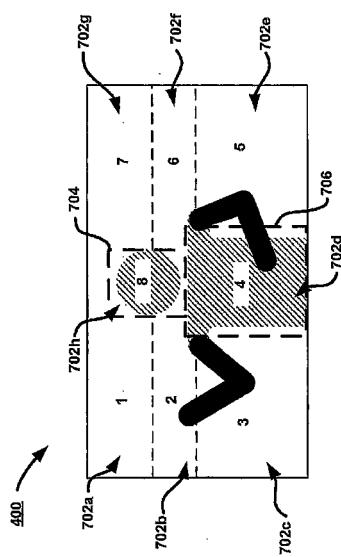
【図5】



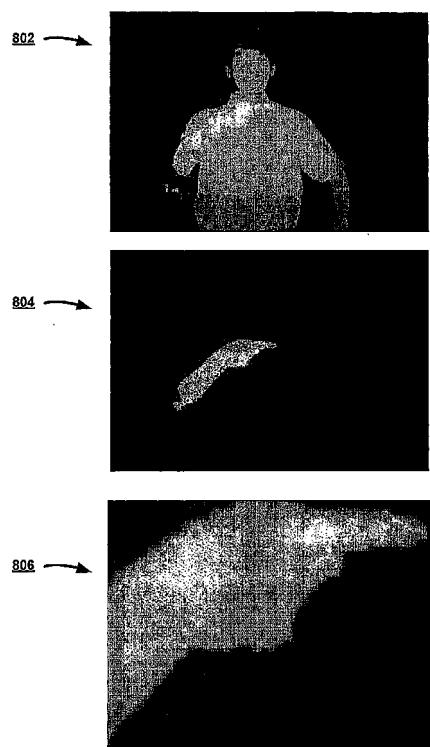
【図6】



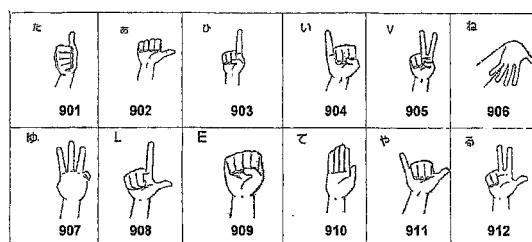
【図7】



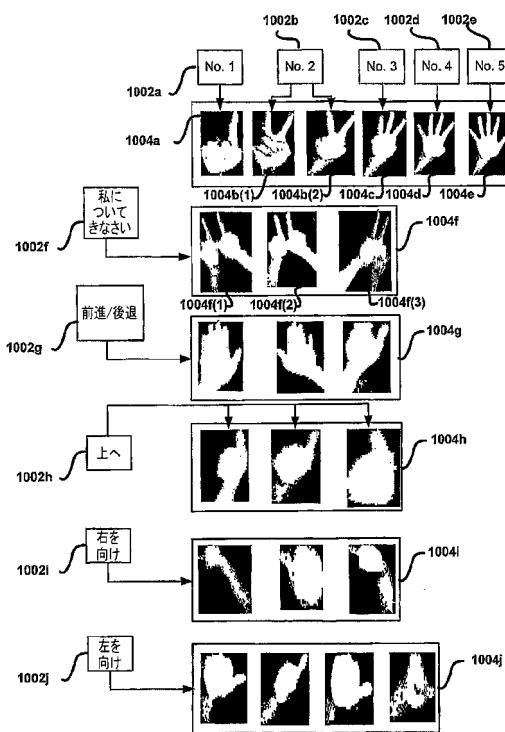
【図8】



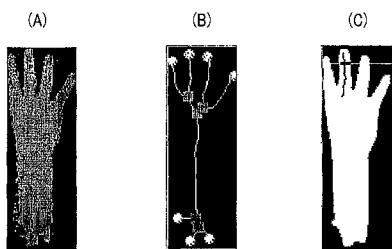
【図9】



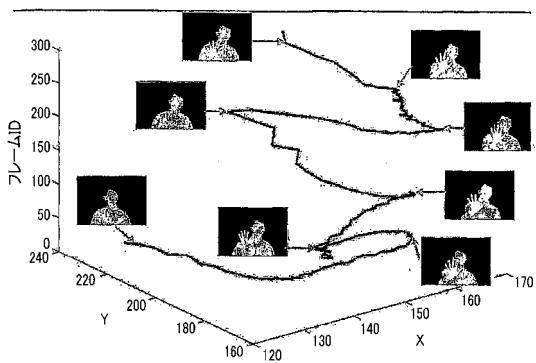
【図10】



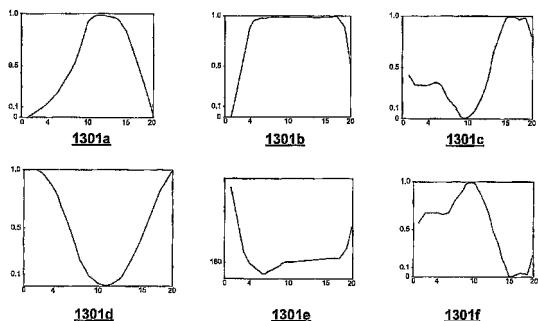
【図 1 1】



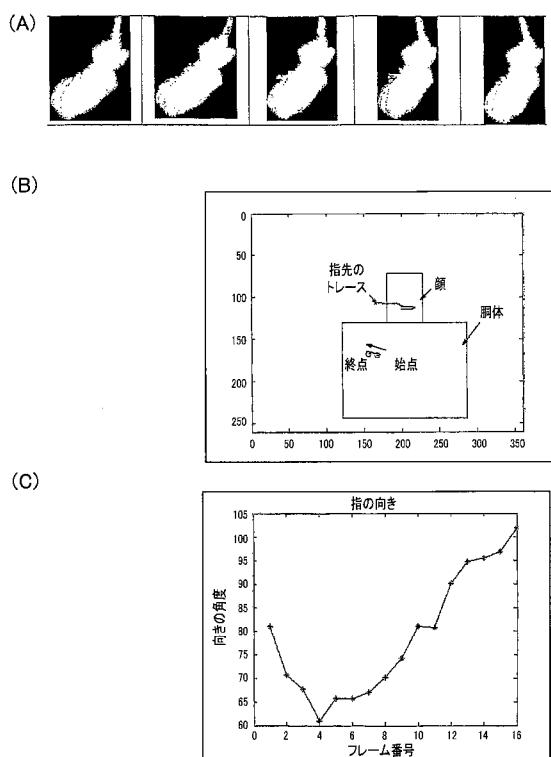
【図 1 2】



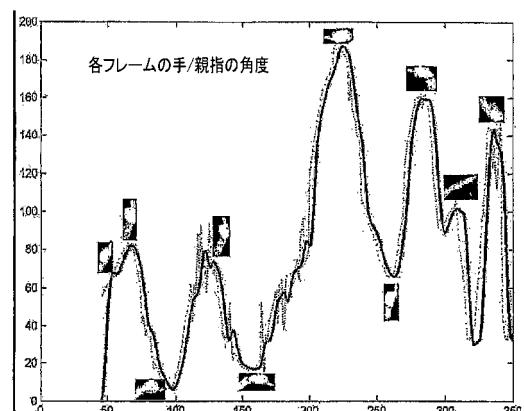
【図 1 3】



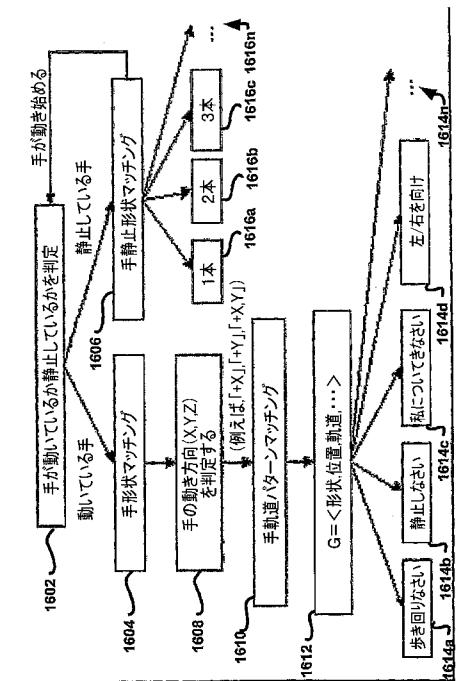
【図 1 4】



【図 1 5】



【図16】



---

フロントページの続き

(51)Int.Cl.

F I

G 0 1 S 17/89 (2006.01)

G 0 1 B 11/24

K

G 0 1 S 17/89

(31)優先権主張番号 11/129,164

(32)優先日 平成17年5月12日(2005.5.12)

(33)優先権主張国 米国(US)

(72)発明者 藤村 希久雄

アメリカ合衆国, カリフォルニア州 94041, マウンテン ビュー, スイート 300, カリフォルニア ストリート 800, ホンダ リサーチ インスティチュート ユーエスエー, インク. 内

(72)発明者 リュー, シア

アメリカ合衆国, オハイオ州 43202, コロンバス, エーピーティ. 1, スティンチコーム ディーアール. 520

審査官 松尾 俊介

(56)参考文献 特開2001-188897 (JP, A)

特開2003-346162 (JP, A)

特開平05-334432 (JP, A)

特開2004-094288 (JP, A)

(58)調査した分野(Int.Cl., DB名)

G06T 7/00~7/60

G01B 11/00

G01B 11/24

G01C 3/06

G01S 17/89

G06F 3/01