(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7679197号
(P7679197)

(45)発行日　令和7年5月19日(2025.5.19)

(24)登録日　令和7年5月9日(2025.5.9)

(51)国際特許分類

Ｇ０６Ｔ	*7/20*	*(2017.01)*	Ｇ０６Ｔ	7/20　３００Ｚ
Ｇ０８Ｇ	*1/16*	*(2006.01)*	Ｇ０８Ｇ	1/16　Ｃ

請求項の数　16　（全23頁）

(21)出願番号　特願2020-552164(P2020-552164)
(86)(22)出願日　平成30年12月13日(2018.12.13)
(65)公表番号　特表2021-507434(P2021-507434A)
(43)公表日　令和3年2月22日(2021.2.22)
(86)国際出願番号　PCT/IB2018/001557
(87)国際公開番号　WO2019/116099
(87)国際公開日　令和1年6月20日(2019.6.20)
審査請求日　令和3年12月2日(2021.12.2)
審判番号　不服2023-12470(P2023-12470/J1)
審判請求日　令和5年7月25日(2023.7.25)
(31)優先権主張番号　62/598,359
(32)優先日　平成29年12月13日(2017.12.13)
(33)優先権主張国・地域又は機関
最終頁に続く

(73)特許権者　520211786
ヒューマニシング オートノミー リミテッド
イギリス ダブリュシー２アール １エル エー ロンドン ストランド サマセット ハウス（番地なし）
(74)代理人　110001243
弁理士法人谷・阿部特許事務所
(72)発明者　マヤ オードリー ララ ピンデウス
イギリス ダブリュシー２アール １エル エー ロンドン ストランド サマセット ハウス（番地なし） ヒューマニシング オートノミー リミテッド内
(72)発明者　ラウナック ボーズ
イギリス ダブリュシー２アール １エル
最終頁に続く

(54)【発明の名称】　歩行者の意図を予測するためのシステムおよび方法

(57)【特許請求の範囲】

【請求項１】

一つまたは複数のプロセッサを備えるコンピュータシステムによって実行される方法であって、

前記一つまたは複数のプロセッサが、ビデオフィードから複数の連続画像を取得するステップと、

前記一つまたは複数のプロセッサが、前記複数の連続画像における各画像内の人間に対応する各キーポイントをそれぞれ決定するステップであって、

前記一つまたは複数のプロセッサが、前記人間における身体輪郭内の各領域を識別するステップと、

前記一つまたは複数のプロセッサが、既定ポイントが人体に関する輪郭内の各領域に含まれる各種部位の各々を指示するテンプレートに、前記身体輪郭内の各領域を比較するステップと、

前記一つまたは複数のプロセッサが、前記身体輪郭内の各領域に対して、前記テンプレートに含まれる前記各種部位内においてマクロレベルで配置された前記既定ポイントに対応する前記各キーポイントを適用するステップと、

前記一つまたは複数のプロセッサが、前記身体輪郭内の各領域の少なくとも１つに対して、前記テンプレートに含まれる前記各種部位内における詳細部位に沿って粒状レベルで配置された前記既定ポイントに対応する前記各キーポイントを適用するステップと、

を含む、ステップと、

前記一つまたは複数のプロセッサが、各画像に関する前記各キーポイントを前記人間のポーズへそれぞれ集約するステップであって、
前記一つまたは複数のプロセッサが、キーポイントの複数のセットを決定するステップであって、キーポイントの各セットは、前記人間の個別の身体部分に対応する、ステップと、
前記一つまたは複数のプロセッサが、前記キーポイントの各セットにおける一端のキーポイントが他のキーポイントの各セットにおける一端のキーポイントに順次接続される関係をそれぞれ判定して、接続関係にある前記キーポイントの各二つのセットにおける接続された一端のキーポイント間における相対的位置および相対的距離を記憶するステップと、
前記一つまたは複数のプロセッサが、前記複数の連続画像における前記キーポイントの各二つのセットにおける前記接続された一端のキーポイント間における相対的位置および相対的距離の経時的な変化として、前記キーポイントの各セットにおける前記一端のキーポイントが他のキーポイントの各セットにおける前記接続された一端のキーポイントに対してどのように移動するかに基づいて、相対的なキーポイント移動のベクトルを決定するステップと、
前記一つまたは複数のプロセッサが、前記ポーズに、相対的なキーポイント移動の前記ベクトルをマッピングするステップと
を含む、ステップと、
前記一つまたは複数のプロセッサが、候補のポーズから意図に変換する複数のテンプレートポーズと前記ポーズを比較するために、データベースへクエリを送信するステップと、
前記一つまたは複数のプロセッサが、前記人間の意図、またはマッチングテンプレートを特定することができないことのいずれかを示す前記データベースから応答メッセージを受信するステップと、
前記一つまたは複数のプロセッサが、前記人間の前記意図を示す前記応答メッセージに応答して、前記意図に対応する命令を出力するステップと
を備える方法。
【請求項２】
前記マッチングテンプレートを特定することができないことを示す前記応答メッセージに応答して、
前記一つまたは複数のプロセッサが、前記ビデオフィードをキャプチャしている車両の通常動作を停止させ、安全動作モードに入るためのコマンドを出力するステップと、
前記一つまたは複数のプロセッサが、前記ビデオフィードを監視して、前記人間が前記ビデオの画像内に存在していない場合を決定するステップと、
前記人間が前記ビデオの画像内に存在していないと決定したことに応答して、前記一つまたは複数のプロセッサが、前記安全動作モードの動作を停止させ、前記車両の通常動作を再開させるコマンドを出力するステップと
をさらに備える、
請求項１に記載の方法。
【請求項３】
前記安全動作モードは、複数のモードのいずれかであり得て、前記コマンドは、前記複数のモードの特定の１つの表示を含み、前記複数のモードの前記特定の１つは、前記車両に対する前記人間、および他の障害物の位置に基づいて選択される、
請求項２に記載の方法。
【請求項４】
前記安全動作モードは、入った場合に、前記車両に旋回、加速、移動の停止、オペレータへの制御の提供、視覚的、聴覚的、またはマルチメディアメッセージの出力、およびホーンの鳴動のうちの少なくとも１つを実行するようコマンドすることを含む、
請求項２に記載の方法。
【請求項５】

一つまたは複数のプロセッサを備えるコンピュータシステムによって実行される方法であって、
前記一つまたは複数のプロセッサが、ビデオフィードから複数の連続画像を取得するステップと、
前記一つまたは複数のプロセッサが、前記複数の連続画像における各画像内の人間に対応する各キーポイントをそれぞれ決定するステップと、
前記一つまたは複数のプロセッサが、各画像に関する前記各キーポイントを前記人間のポーズへそれぞれ集約するステップであって、
前記一つまたは複数のプロセッサが、キーポイントの複数のセットを決定するステップであって、キーポイントの各セットは、前記人間の個別の身体部分に対応する、ステップと、
前記一つまたは複数のプロセッサが、キーポイントの各セットにおける一端のキーポイントが他のキーポイントの各セットにおける接続された一端のキーポイントに対してどのように移動するかに基づいて、相対的なキーポイント移動のベクトルを決定するステップと、
前記一つまたは複数のプロセッサが、前記ポーズに、相対的なキーポイント移動の前記ベクトルをマッピングするステップと
を含む、ステップと、
前記一つまたは複数のプロセッサが、候補のポーズからそれぞれの将来の行動を示す意図に変換する複数のテンプレートポーズと前記ポーズを比較するために、文化的ポーズ意図データベースへクエリを送信するステップであって、
前記一つまたは複数のプロセッサが、前記ビデオフィードをキャプチャしている車両の地理的位置を決定するステップと、
前記一つまたは複数のプロセッサが、複数の文化的ポーズ意図データベースのインデックスにアクセスするステップであって、前記インデックスは、前記複数の文化的ポーズ意図データベースのうちの各文化的ポーズ意図データベースを各データベース位置および各アドレスに対応させるエントリを含む、ステップと、
前記一つまたは複数のプロセッサが、前記地理的位置を前記エントリに対して比較して、どの文化的ポーズ意図データベースにクエリを行うかを決定するステップであって、前記複数の文化的ポーズ意図データベースのうちのマッチング文化的ポーズ意図データベースは、
前記一つまたは複数のプロセッサが、前記複数の文化的ポーズ意図データベースから、前記ポーズをそれぞれの将来の行動を示す意図に変換することに関連して使用される前記文化的ポーズ意図データベースを選択するステップであって、該選択は、前記ビデオフィードが撮影された場所に基づいて実行され、前記ポーズが、第一の場所に対応する前記複数の文化的ポーズ意図データベースのうちの第一のデータベースによって第一の意図に変換されることが示され、前記ポーズが、前記第一の場所とは異なる第二の場所に対応する前記複数の文化的ポーズ意図データベースのうちの第二のデータベースによって前記第一の意図とは異なる第二の意図に変換されることが示される、ステップ
によって、前記地理的位置とマッチングする前記マッチング文化的ポーズ意図データベースの前記各データベース位置に基づいて選択される、ステップと、
前記一つまたは複数のプロセッサが、前記クエリが送信される前記データベースとして、前記マッチング文化的ポーズ意図データベースを割り当てるステップと
を含む、ステップと、
前記一つまたは複数のプロセッサが、将来の行動を示す前記人間の意図、またはマッチングテンプレートを特定することができないことの何れかを示す、前記選択された文化的ポーズ意図データベースから応答メッセージを受信するステップと、
前記一つまたは複数のプロセッサが、前記将来の行動を示す前記人間の前記意図を示す前記応答メッセージに応答して、前記意図に対応する命令を出力するステップと
を備える方法。

【請求項６】
前記一つまたは複数のプロセッサが、前記複数の連続画像が人間を含むかを決定するステップと、
前記複数の連続画像が前記人間を含むという決定に応答して、前記一つまたは複数のプロセッサが、前記複数の連続画像のうちの各画像内の前記人間に対応する各キーポイントを決定する前記ステップを実行するステップと、
前記複数の連続画像が前記人間を含まないという決定に応答して、前記一つまたは複数のプロセッサが、前記複数の連続画像を破棄し、前記ビデオフィードから次の複数の連続画像を取得するステップと
をさらに備える、
請求項１の方法。
【請求項７】
前記ビデオフィードから前記複数の連続画像を取得するステップは、
前記一つまたは複数のプロセッサが、現在の時間より前の時間の長さ、または現在キャプチャされた画像より前の連続画像の量のいずれかを表す量を決定するステップと、
前記一つまたは複数のプロセッサが、現在時刻、または現在キャプチャされた画像のいずれかに対する前記連続画像の量を前記ビデオフィードから取得するステップと
をさらに含む、
請求項１に記載の方法。
【請求項８】
前記意図は、機械学習、統計分析、および心理的行動モデルを前記複数の連続画像のうちの各画像に適用することの少なくとも一つを使用して決定される、
請求項１に記載の方法。
【請求項９】
エンコードされた命令を備えるコンピュータ可読記憶媒体であって、クライアントデバイスのプロセッサによって前記命令が実行されると、前記プロセッサに、
ビデオフィードから複数の連続画像を取得することと、
前記複数の連続画像における各画像において、人間に対応する各キーポイントをそれぞれ決定することであって、
前記人間における身体輪郭内の各領域を識別することと、
既定ポイントが人体に関する輪郭内の各領域に含まれる各種部位の各々を指示するテンプレートに、前記身体輪郭内の各領域を比較することと、
前記身体輪郭内の各領域に含まれる前記各種部位に対して、前記テンプレートに含まれる前記各種部位内においてマクロレベルで配置された前記既定ポイントに対応する前記各キーポイントを適用することと、
前記身体輪郭内の各領域に含まれる前記各種部位の少なくとも１つに対して、前記テンプレートに含まれる前記各種部位内における詳細部位に沿って粒状レベルで配置された前記既定ポイントに対応する前記各キーポイントを適用することと
を含む、ことと、
各画像に関する前記各キーポイントを前記人間のポーズへそれぞれ集約することであって、
キーポイントの複数のセットを決定することであって、キーポイントの各セットは、前記人間の個別の身体部分に対応する、ことと、
前記キーポイントの各セットにおける一端のキーポイントが他のキーポイントの各セットにおける一端のキーポイントに順次接続される関係をそれぞれ判定して、接続関係にある前記キーポイントの各二つのセットにおける接続された一端のキーポイント間における相対的位置および相対的距離を記憶することと、
前記複数の連続画像における前記キーポイントの各二つのセットにおける前記接続された一端のキーポイント間における相対的位置および相対的距離の経時的な変化として、前記キーポイントの各セットにおける前記一端のキーポイントが他のキーポイントの各セ

ットにおける前記接続された一端のキーポイントに対してどのように移動するかに基づいて、相対的なキーポイント移動のベクトルを決定することと、
前記ポーズに、相対的なキーポイント移動の前記ベクトルをマッピングする、ことと、を含む、ことと、
候補のポーズを意図に変換する複数のテンプレートポーズと前記ポーズを比較することにより、前記ポーズとマッチングするテンプレートを特定するために、データベースへクエリを送信することであって、各テンプレートは関連する意図に対応している、ことと、
マッチングテンプレートに基づく前記人間の意図、または前記マッチングテンプレートを特定することができないことの何れかを示す応答メッセージを前記データベースから受信することと、
前記人間の意図を示す前記応答メッセージに応答して、前記意図を出力することと
を行わせる、コンピュータ可読記憶媒体。
【請求項１０】
前記命令は、前記プロセッサに、前記マッチングテンプレートを特定することができないことを示す前記応答メッセージに応答して、
前記ビデオフィードをキャプチャしている車両の通常動作を停止させ、安全動作モードに入るためのコマンドを出力することと、
前記ビデオフィードを監視して、前記人間が前記ビデオの画像内に存在していない場合を決定することと、
前記人間が前記ビデオの画像内に存在していないと決定したことに応答して、前記安全動作モードの動作を停止させ、前記車両の通常動作を再開させるコマンドを出力することと、
をさらに行わせる、
請求項９に記載のコンピュータ可読記憶媒体。
【請求項１１】
前記安全動作モードは、複数のモードのいずれかであり得て、前記コマンドは、前記複数のモードの特定の１つの表示を含み、前記複数のモードの前記特定の１つは、前記車両に対する前記人間、および他の障害物の位置に基づいて選択される、
請求項１０に記載のコンピュータ可読記憶媒体。
【請求項１２】
前記安全動作モードは、入った場合に、前記車両に旋回、加速、移動の停止、オペレータへの制御の提供、およびホーンの鳴動の少なくとも１つを実行するようコマンドすることを含む、
請求項１０に記載のコンピュータ可読記憶媒体。
【請求項１３】
前記命令は、前記プロセッサに、前記クエリを前記データベースへ送信する場合に、
前記ビデオフィードをキャプチャしている車両の地理的位置を決定することと、
複数の候補データベースのインデックスにアクセスすることであって、前記インデックスは、前記複数の候補データベースのうちの各候補データベースを各データベース位置および各アドレスに対応させるエントリを含む、ことと、
前記地理的位置を前記エントリに対して比較して、クエリを行う候補データベースを決定することであって、前記複数の候補データベースのうちのマッチング候補データベースは、前記地理的位置とマッチングする前記マッチング候補データベースの前記各データベース位置に基づいて選択される、ことと、
前記クエリが送信される前記データベースとして、前記マッチング候補データベースを割り当てることと、
をさらに行わせる、
請求項９に記載のコンピュータ可読記憶媒体。
【請求項１４】
前記命令は、前記プロセッサに、

前記複数の連続画像が人間を含むかを決定することと、
前記複数の連続画像が前記人間を含むという決定に応答して、前記複数の連続画像のうちの各画像内の前記人間に対応する各キーポイントの前記決定を実行することと、
前記複数の連続画像が前記人間を含まないという決定に応答して、前記複数の連続画像を破棄し、前記ビデオフィードから次の複数の連続画像を取得することと、
をさらに行わせる、
請求項９に記載のコンピュータ可読記憶媒体。
【請求項１５】
前記命令は、前記プロセッサに、前記ビデオフィードから前記複数の連続画像を取得した場合に、
現在の時間より前の時間の長さ、または現在キャプチャされた画像より前の連続画像の量のいずれかを表す量を決定することと、
現在の時間、または現在キャプチャされた画像のいずれかに対する前記連続画像の量を前記ビデオフィードから取得することと、
をさらに行わせる、
請求項１４に記載のコンピュータ可読記憶媒体。
【請求項１６】
前記意図は、機械学習、統計分析、および心理的行動モデルを前記複数の連続画像のうちの各画像に適用することの少なくとも１つを使用して決定される、
請求項９に記載のコンピュータ可読記憶媒体。
【発明の詳細な説明】
【技術分野】
【０００１】
本開示は、概して、自動化、または自動運転車を誘導するコンピュータビジョンの分野に関するものであり、より具体的には、歩行者の意図を予測するようコンピュータビジョンを適用することに関する。
【背景技術】
【０００２】
この出願は、２０１７年１２月１３日に提出された米国仮出願第６２／５９８，３５９号の利益を主張し、その全体が参照により本明細書に組み込まれる。
【０００３】
歩行者の意図を判断する関連する従来のシステムは、歩行者が移動している方向、または速度を分析すること、および道路のエッジからの歩行者の距離を分析したりするなどの意図判断を実行するための粗雑で原始的な手段を用いている。従来技術のシステムは、これらの変数を静的モデルへと取り込み、ユーザの最も起こり得る意図を大まかに判断する。このアプローチは、過度に広範囲であり、明示的、または暗示的なジェスチャーの形態における人間のボディーランゲージ、および本人の周囲に対する人間の認識という他の事物の間で、アクセスに失敗する画一的なアプローチである。技術的な観点から、従来技術のシステムは、人間の一連の画像に基づく人間のポーズを理解する技術的な精巧さに欠けていることで、ジェスチャー、または認識情報をより正確な意図予測に変換することが可能ではない。
【図面の簡単な説明】
【０００４】
開示された実施形態は、詳細な説明、添付の特許請求の範囲、および添付の図（または図面）からより容易に明らかとなる他の利点、および特徴を有する。図の簡単な紹介を以下に示す。
【０００５】
【図１】本開示のいくつかの実施形態によって、意図判断サービスを用いて人間の意図が判断されるシステムの一実施形態を図示する。
【図２】本開示のいくつかの実施形態によって、機械可読媒体から命令を読み取り、それ

らをプロセッサ（またはコントローラ）において実行することが可能である例示的な機械のコンポーネントを図示するブロック図の一実施形態を図示する。
【図３】本開示のいくつかの実施形態によって、サービスをサポートするモジュールと、データベースとを含む意図判断サービスの一実施形態を図示する。
【図４】本開示のいくつかの実施形態によって、さまざまな人間活動に関連するポーズの検出のためにキーポイントを描写する。
【図５】本開示のいくつかの実施形態によって、カメラに対して人間が対面している方向に関係なく、同じポーズを検出するためにキーポイントを描写する。
【図６】本開示のいくつかの実施形態によって、ユーザが注意深いかを検出すること、および注意散漫のタイプを検出することのためにキーポイントを描写する。
【図７】本開示のいくつかの実施形態によって、地理別のジェスチャーの差の図を描写する。
【図８】本開示のいくつかの実施形態によって、いくつかの例示的なポーズを予測される意図へマッピングするためにフローダイヤグラムを描写する。
【図９】本開示のいくつかの実施形態によって、目印を参照して、その目印を参照して人間の方向を判断するために用いられるキーポイントを描写する。
【図１０】本開示のいくつかの実施形態によって、人間の身体のさまざまな部分を特定するためにキーポイントクラスターを描写する。
【図１１】本開示のいくつかの実施形態によって、ビデオフィードから受信された画像を意図判断へ変換するためにフローダイヤグラムを描写する。
【発明を実施するための形態】
【０００６】
図面（図）、および以下の説明は、例示のみによる好ましい実施形態に関する。以下の議論から、本明細書に開示される構造、および方法の代替的な実施形態は、主張されるものの原理から逸脱することなく、用いられ得る実行可能な代替として容易に認識されることとなることが留意されたい。
【０００７】
次に、ここでいくつかの実施形態を詳細に参照し、その例は、添付図面に図示される。実施可能な場合はいつでも、同様または類似の参照番号が図で用いられ、同様または類似の機能を示す場合があることに留意されたい。図は、例示のみを目的として、開示されたシステム（または方法）の実施形態を描写する。当業者は、本明細書に図示される構造、および方法の代替的な実施形態が、本明細書に記載される本発明の原理から逸脱することなく用いられ得ることを以下の説明から容易に認識することであろう。
【０００８】
（構成の概説）
開示されたシステム、方法、およびコンピュータ可読記憶媒体の一実施形態は、（例えば、ビデオから）人間の一連の画像をキーポイントの集積へと変換し、人間のポーズを判断することによって、人間の意図を判断することを含む。次に、判断されたポーズは、既知のポーズから意図へのマッピングに基づいて、人間の意図へマッピングされる。これらのポーズから意図へのマッピングは、人間の文化に応じて異なる意図を現示する可能性があるポーズを教えるさまざまなジェスチャー、および癖を説明するためにビデオがキャプチャされる場所の地理に基づいて異なる場合がある。この目的および他の目的のために、本開示のいくつかの実施形態において、プロセッサは、カメラ（例えば、ビデオフィードを提供するカメラ）、遠赤外線センサー、およびＬＩＤＡＲなどの視覚および／または深度センサーによって提供されるフィードから複数の一連の画像を取得する。カメラがビデオフィードを提供する状況において、カメラは自動運転車のダッシュボード上もしくはその近く、または車両上もしくは車両内の任意の場所（例えば、車両の車体構造へと組み込まれる）に取り付けることができる。プロセッサは、一連の画像の画像から画像へと変わるように、人間の頭、腕、および脚に対応するポイント、およびそれらの互いの相対位置など、複数の一連の画像のそれぞれの画像における人間に対応するそれぞれのキーポイン

トを判断する。

【０００９】

プロセッサは、それぞれの画像に対してそれぞれのキーポイントを人間のポーズへと集約する。例えば、キーポイントの分析から判断されるように、一連の画像にわたる人間の脚の移動に基づいて、プロセッサはキーポイントを画像に渡るキーポイントの移動を示すベクトルへと集約する。次に、プロセッサは、ポーズが通りを横断する意図などの所与の意図へマッピングすることが既知であるかを判断すべく、データベースへクエリを伝送し、意図候補のポーズを意図に変換する複数のテンプレートポーズと比較する。プロセッサは、データベースから、人間の意図、またはマッチングするテンプレートの場所を特定することができないことのいずれかを示す応答メッセージを受信する。人間の意図を示す応答メッセージに応答して、プロセッサは、意図に対応するコマンド（例えば、車両を停止するために、または車両のドライバーに警告するために）を出力する。

【００１０】

（システムの概説）

図１は、本開示のいくつかの実施形態によって、意図判断サービスを用いて人間の意図を判断するシステムの一実施形態を図示する。システム１００は、車両１１０を含む。車両１１０は、自動車として描写されている一方で、車両１１０は、自動または半自動で人間の近くを通行するよう構成される任意の電動装置であってよい。本明細書で用いられる「自動」という用語、およびその変形例は、完全に自動化された動作、または他の機能ではなく（例えば、「半自動」）、いくつかの機能を動作するよう人間のインプットに依存する半自動動作を参照する。例えば、車両１１０は、任意の方向、または任意の速さで航行するよう、命令されることなく、（または少なくともいくつかの機能を実行するよう命令されている間に）人間の付近を飛行、または歩くよう構成されるドローン、または二足歩行のロボットであってよい。車両１１０は、他の物の中の人間の周りで安全に動作するよう構成され、人間１１４などの近くの人間の意図を判断し（または言われ）、人間１１４の意図を考慮して、安全に調和した行動をとる。

【００１１】

カメラ１１２は、車両１１０に動作可能に連結され、人間１１４の意図を判断することに用いられる。マイクロフォンなどの他のセンサー１１３は、車両１１０に動作可能に結合することができ、人間１１４の意図判断（例えば、人間１１４が「私は今、通りを横断するつもりだ」と話した場合、車両１１０は、マイクロフォンインプットに基づいて、車両１１０が衝突を回避するよう車両を減速、または停止させる安全モードを起動させることができる）を補足する。本明細書で用いられるように、「動作可能に結合される」という用語は、直接取り付け（例えば、同じ回路基板へと配線するか、または組み込まれる）、間接装着（例えば、ワイヤ、または無線通信プロトコルを通じて接続される２つのコンポーネント）などを参照する。一連の画像は、カメラ１１２によってキャプチャされ、ネットワーク１２０を介して意図判断サービス１３０に与えられる。ネットワーク１２０の技術的態様は、図２に関してさらに詳細に説明されることとなる。以下の開示において、図３などの他の図に関してさらに詳細に説明されることとなるように、意図判断サービス１３０は、人間１１４の意図を判断する。ネットワーク１２０、および意図判断サービス１３０が、車両１１０から離れているように描写されている一方で、ネットワーク１２０および／または意図判断サービス１３０は、全体的、または部分的に、車両１１０内に実装されることができる。

【００１２】

（コンピュータ機器の構成）

図２は、機械可読媒体から命令を読み取り、それらをプロセッサ（またはコントローラ）において実行することが可能である例示的な機械のコンポーネントを図示するブロック図である。具体的には、図２は、本明細書で説明される任意の１つまたは複数の方法を機械に実行させるために、プログラムコード（例えば、ソフトウェア）を中に有するコンピュータシステム２００の例示的な形態における機械の図式表現を図示する。プログラムコ

ードは、１つまたは複数のプロセッサ２０２によって実行可能な命令２２４から構成されてよい。代替の実施形態において、機械は、スタンドアローンデバイスとして動作するか、または他の機械に接続（例：ネットワーク化）されてよい。ネットワーク化された展開において、機械は、サーバクライアントネットワーク環境の中のサーバマシン、またはクライアントマシンの能力で動作するか、またはピアツーピア（または分散型）ネットワーク環境のピアマシンとして動作することができる。

【００１３】

マシンは、サーバコンピュータ、クライアントコンピュータ、パーソナルコンピュータ（ＰＣ）、タブレットＰＣ、セットトップボックス（ＳＴＢ）、携帯情報端末（ＰＤＡ）、携帯電話、スマートフォン、ｗｅｂアプライアンス、ネットワークルータ、スイッチまたはブリッジ、車載用コンピュータ（例えば、車両１１０が動作するための車両１１０内に組み込まれたコンピュータ）、インフラストラクチャコンピュータ、またはその機械によって取られるアクションを特定する（一連のまたはその他の）命令２２４を実行することが可能な任意の機械であってよい。さらに、単一の機械のみが図示されている一方で、「機械」という用語は、また、命令１２４を個別に、または協働して実行する機械の任意の集合を含むものとされ、本明細書で説明される方法のいずれか１つまたは複数を実行するものである。

【００１４】

例示的なコンピュータシステム２００は、プロセッサ２０２（例えば、中央処理装置（ＣＰＵ）、グラフィック処理装置（ＧＰＵ）、デジタル信号プロセッサ（ＤＳＰ）、１つまたは複数の特定用途向け集積回路（ＡＳＩＣｓ）、１つまたは複数の無線周波数集積回路（ＲＦＩＣｓ）、またはこれらの任意の組み合わせ）と、メインメモリ２０４と、スタティックメモリ２０６とを含み、バス２０８を介して互いに通信するよう構成される。コンピュータシステム２００は、ビジュアルディスプレイインターフェース２１０をさらに含んでよい。ビジュアルインターフェースは、ユーザインターフェースをスクリーン（またはディスプレイ）上に表示することを可能とするソフトウェアドライバを含んでよい。ビジュアルインターフェースは、直接的に（例えば、スクリーン上に）、または間接的に表面上またはウィンドウ上などに（例えば、ビジュアルプロジェクションユニットを介して）間接的にユーザインターフェースを表示することができる。説明を容易にするために、ビジュアルインターフェースは、スクリーンとして説明されてよい。ビジュアルインターフェース２１０は、タッチ可能スクリーンを含んでよく、またはタッチ可能スクリーンとインターフェースで接続してよい。また、コンピュータシステム２００は、英数字入力デバイス２１２（例えば、キーボードまたはタッチスクリーンキーボード）と、カーソル制御デバイス２１４（例えば、マウス、トラックボール、ジョイスティック、人感センサー、または他のポインティング機器）と、記憶ユニット２１６と、信号生成デバイス２１８（例えば、スピーカ）と、ネットワークインターフェースデバイス２２０とを含み、また、バス２０８を介して通信するよう構成される。信号生成デバイス２１８は、ビジュアルインターフェース２１０へ出力するか、または振動発生器、およびロボットアームなどの他のデバイスへ信号を出力することができ、車両１１０の人間の運転者に歩行者への危険を警告（例えば、運転席を振動させることによって運転者の注意を引く、警告の表示、およびプロセッサ２０２は、運転者が眠ってしまったと判断した場合に、ロボットアームを用いて運転者を軽く叩くなど）する。

【００１５】

記憶ユニット ２１６は、本明細書で説明される任意の１つまたは複数の方法、または機能を具現化する命令２２４（例えば、ソフトウェア）が記憶される機械可読媒体２２２を含む。また、命令２２４（例えば、ソフトウェア）は、コンピュータシステム２００によるその実行の間に、メインメモリ２０４内、またはプロセッサ２０２内（例えば、プロセッサのキャッシュメモリ内）に完全に、または少なくとも部分的に存在してよく、メインメモリ２０４、およびプロセッサ２０２はまた、機械可読媒体を構成する。命令２２４（例えば、ソフトウェア）は、ネットワークインターフェースデバイス２２０を介してネッ

トワーク２２６を超えて伝送、または受信されることができる。
【００１６】
機械可読媒体２２２が例示的な実施形態において単一の媒体であると示されている一方で、「機械可読媒体」という用語は、命令（例えば、命令２２４）を記憶することができる単一の媒体、または複数の媒体（例えば、集中もしくは分散データベース、または関連するキャッシュおよびサーバ）を含むことを理解されたい。また、「機械可読媒体」という用語は、機械による実行のために命令（例えば、命令２２４）を記憶することが可能である任意の媒体を含むことができ、それが機械に本明細書に開示される任意の１つまたは複数の方法を実行させることを理解されたい。「機械可読媒体」という用語は、固体メモリ、光学媒体、および磁気媒体の形態におけるデータリポジトリを含むが、これに限定されない。
【００１７】
（意図判断サービスの構成）
上述のように、本明細書に説明されるシステム、および方法は、一般に、自動運転車（例えば、車両１１０のカメラ１１２）に搭載、または統合されるカメラの画像から判断されるように、人間のポーズに基づく人間の意図（例えば、人間１１４）を判断することを対象としている。画像は、意図判断サービス１３０のプロセッサ（例えば、プロセッサ２０２）によって取り込まれる。図３は、本開示のいくつかの実施形態によって、サービスをサポートするモジュール、およびデータベースを含み、意図判断サービスの一実施形態を図示する。意図判断サービス３３０は、意図判断サービス１３０を詳細に見たものであり、本明細書に記載されているものと同じ機能を有する。意図判断サービス３３０は、そこに描写されているさまざまなモジュール、およびデータベースを含む単一のサーバであってよく、または複数のサーバ、およびデータベース渡って分散されてよい。上述のように、意図判断サービス３３０は、全体的、または部分的に、車両１１０内に実装されてよい。分散サーバ、およびデータベースは、ネットワーク１２０によってアクセス可能なファーストパーティまたはサードパーティのサービス、およびデータベースであってよい。
【００１８】
意図判断サービス３３０は、キーポイント判断モジュール３３７を含む。カメラ１１２からの画像を取り込んだ後に、意図判断サービス３３０のプロセッサ２０２は、人間検出モジュール３３３を実行する。人間検出モジュール３３３は、画像の構成を分析することにおいて、コンピュータビジョンおよび／またはオブジェクト認識技術を用いることによって、人間が画像の中にいるかを判断し、画像内のエッジ、または他の特徴が、人間の既知の形態とマッチングするかを判断する。画像の中に人間がいると判断すると、意図判断サービス３３０のプロセッサ２０２は、キーポイント判断モジュール３３７を実行し、人間のキーポイントを判断する。この判断は、図４について説明されることとなる。図４は、本開示のいくつかの実施形態によって、さまざまな人間活動に関連するポーズの検出のためにキーポイントを描写する。人間４１０は、ジェスチャーを行わずに中立の位置に立っている。キーポイントは、人間４１０の上に重ね合わせられた円で示される。線は、ポーズの特徴を図示するために用いられ、以下でさらに詳細に説明されることとなる。
【００１９】
キーポイント判断モジュール３３７は、キーポイントが人間４１０のどこにあるかを判断する。キーポイント判断モジュール３３７は、例えば、最初に人体の輪郭を識別し、キーポイントがどこにあるかを判断し、次に輪郭、またはキーポイント／ポーズ判断の他の方法に基づいて、人体の既定のポイントをキーポイントとしてマッチングする。例えば、キーポイント判断モジュール３３７は、それぞれの人間の足首、膝、大腿、胴体の中心、手首、肘、肩、首、および額に適用されるキーポイントを指示する人間画像のテンプレートを参照してよい。キーポイント判断モジュール３３７は、人間画像を人間画像のテンプレートと比較することによって、これらのポイントそれぞれを識別することができ、テンプレートは、キーポイントが存在するそれぞれの領域に対応する特徴を指し示す。あるいは、キーポイント判断モジュール３３７は、コンピュータビジョンを用いて、キーポイン

トが適用されるそれぞれの身体部分を区別、および識別するよう構成されることができ、これらの身体部分の位置を特定し、キーポイントを適用することができる。人間４２０、人間４３０、人間４４０、人間４５０、および人間４６０によって描写されるように、キーポイント判断モジュール３３７は、人間の異なるポーズに関係なく、人間のキーポイントを識別するよう構成される。人間のポーズの挙動が検討され、図３の他のモジュールを参照して以下に説明される。

【００２０】

人間（例えば、人間１１４）が存在するそれぞれの画像に対して、どのキーポイントが用いられるかを判断した後に、意図判断サービス３３０のプロセッサ２０２は、キーポイント集約モジュール３３１を実行する。キーポイント集約モジュール３３１は、単一の画像、または一連の連続した画像に対するキーポイントを集約することができる。単一の画像に対するキーポイントを集約することが、特定の時点でのポーズの表現を作成する。一連の連続した画像のキーポイントを集約することは、キーポイントの移動のベクトルが経時的に作成され、既知の動き、またはジェスチャーへマッピングすることができる。キーポイント集約モジュール３３１が単一の画像に対してキーポイントを集約している場合、再び図４を参照すると、キーポイント集約モジュール３３１は、直線コネクタを用いてキーポイントをマッピングすることによってキーポイントを集約することができる。

【００２１】

キーポイント集約モジュール３３１は、既定のスキーマに基づいて、どのキーポイントをそのような方法で接続するかを判断する。例えば、上述のように、キーポイントモジュール３３１は、それぞれのキーポイントが特定の身体部分（例えば、足首、膝、大腿など）に対応することが分かっている。キーポイント集約モジュール３３１は、既定のスキーマにアクセスし、足首のキーポイントが膝のキーポイントに接続され、大腿のキーポイントに接続されることなどを判断する。これらのキーポイントの直線コネクタは、図４に描写されているそれぞれの人間４１０、人間４２０、人間４３０、人間４４０、人間４５０、および人間４６０について表現されている。記憶すること、または直線コネクタを生成することではなく、またはそれに加えて、キーポイント集約モジュール３３１は、既定のスキーマのそれぞれのキーポイント間の距離を記憶することができる。後で説明されるように、これらの線の相対的な位置、またはキーポイント間の距離は、ユーザのポーズを判断することに用いられる。例えば、人間４５０の手と頭との間の距離は、人間４１０の手と頭との間の距離よりもはるかに短く、これは、人間４５０に対して、人間４５０が電話を頭の方へ持っている可能性が高いと判断することに用いられることになる。

【００２２】

キーポイント集約モジュール３３１が一連の画像のキーポイントを集約している場合、キーポイント集約モジュール３３１は、個々の画像のそれぞれに対してキーポイントを集約し、次に、それぞれの画像から一連の次のそれぞれの画像へのそれぞれのキーポイントに対する運動ベクトルを計算する。例えば、最初の画像が人間４１０を含み、次の画像が人間４１０を含み、人間４１０と同じ人間であるが異なるポーズである状況において、キーポイント集約モジュール３３１は、運動ベクトルを計算し、２つの画像間の人間の左足と膝との動きの変換を描写する。この運動ベクトルは、左足のキーポイントが以前あった場所とそれが移動した場所との間の距離を示す、人間４３０の下部の線で図示されている。

【００２３】

また、キーポイント集約モジュール３３１は、画像間のキーポイントの移動に基づき、且つ画像の中の他のオブジェクトを参照し、キーポイントの位置に基づいて、ユーザが移動、または対面している方向を示すことができる。図５は、本開示のいくつかの実施形態によって、カメラに対して人間が対面している方向に関係なく、同じポーズを検出するためのキーポイントを描写する。キーポイント集約モジュール３３１は、二次元空間において画像を見る場合に、キーポイントの近さに基づいて、人間が対面している方向がどこかを判断することができる。例えば、人間１１４が図５に描写され、位置５１０、位置５２０、位置５３０、位置５４０、および位置５５０などのさまざまな回転位置にある。画像

５４０において、画像５１０と比較すると、肩と首とのキーポイントが互いに近くなっている。キーポイント間のこれらの相対距離、および画像間におけるキーポイント間のこれらの相対距離の変化は、キーポイント集約モジュール３３１によって分析され、三次元空間において人間１１４の回転を判断する。キーポイント集約モジュール３３１は、回転を判断することにおいて、人間１１４が回転してくる方向を検出することが可能であり、画像の中の他のオブジェクト（例えば、道路、縁石）を理解することと合わせて、この情報とキーポイントの集約を回転に基づいて、意図判断モジュール３３２へ送り、ユーザの意図を判断することが可能である。

【００２４】

キーポイントを集約した後に、意図判断サービス３３０のプロセッサ２０２は、意図判断モジュール３３２を実行し、集約されたキーポイントを任意に、追加情報と共にフィードする。意図判断モジュール３３２は、集約されたキーポイントを潜在的な追加の情報と共に用いて、人間１１４の意図を判断する。いくつかの実施形態において、意図判断モジュール３３２は、集約されたキーポイントをポーズとして分類する。本明細書で用いられるポーズは、所与の時点での静止した人間、またはジェスチャーなどの人間の動きであってよい。したがって、分類されたポーズは、静止したポーズの集積、または上述のように、画像から画像へのキーポイント間の移動を説明するベクトルであってよい。

【００２５】

意図判断モジュール３３２は、ポーズテンプレートデータベース３３４にクエリを行い、人間１１４の意図を判断する。ポーズテンプレートデータベース３３４は、さまざまなキーポイントの集約のエントリと、集約間の運動ベクトルとを含む。これらのエントリは、対応するポーズへマッピングされる。例えば、図４に戻って参照すると、テンプレートデータベース３３４は、人間４５０に対するキーポイントの集約とマッチングするキーポイントの集約の記録を含んでよく、「電話中」に対応するポーズを指す場合がある。任意の数の対応するポーズは、起立している、座っている、自転車に乗っている、走っている、電話をしている、別の人間と手をつないでいるなど、キーポイントの集約、および運動ベクトルへマッピングすることができる。さらに、キーポイントの集約は、１つよりも多くのレコードとマッチングする場合がある。例えば、キーポイントの集約は、走っていることに対応するテンプレート、および電話をしていることに対応するテンプレートの両方とマッチングする場合がある。

【００２６】

ポーズテンプレートデータベース３３４は、キーポイントの集約を人間１１４の動き、または姿勢を凌ぐ人間１１４の特徴へマッピングし、また、心の状態を含むことができる。図６は、本開示のいくつかの実施形態によって、ユーザが注意深いかを検出すること、および注意散漫のタイプを検出することのためにキーポイントを描写する。ポーズテンプレートデータベース３３４は、例えば、人間１１４の顔に対するキーポイントの集約をユーザが注意散漫であるか、または注意を払っているかでマッピングし、人間が注意散漫である場合は、ポーズテンプレートデータベース３３４が、キーポイントの集約を注意散漫のため追加してマッピングすることができる。例えば、ポーズテンプレートデータベース３３４は、キーポイントの集約を注意深い人間１１４へマッピングするレコード６１０を含んでよい。また、ポーズテンプレート３３４は、レコード６２０と、レコード６３０と、レコード６４０とを含み、それぞれが注意散漫である人間へマッピングし、横を見ている、電話をしている、下方を向いていることから注意散漫である人間へそれぞれマッピングすることができる。クエリに応答してポーズテンプレート３３４からマッチングするテンプレートを受け取ると、意図判断モジュール３３２は、人間１１４のポーズが１つまたは複数のポーズ、および心の状態とマッチングすると判断することができる。

【００２７】

ポーズが作成された場所に応じて、同じポーズが２つの異なるものを意味し得る状況が存在する。例えば、人間１１４がヨーロッパにおいて彼の手を挙げた場合、通常の文化的応答は、これを「停止する」合図として理解する。このジェスチャーが通常「継続する」

ことを意味する国において、人間１１４が手を挙げた場合、意図判断モジュール３３２が、キーポイントとそのような意味との間の対応を示すポーズテンプレートデータベース３３４のレコードに基づいて、これは、停止するジェスチャーであると判断することは誤りであるということになる。したがって、いくつかの実施形態において、ポーズテンプレートデータベース３３４は、単にポーズが何であるかを詳細に示すだけであり、例えば、人間の手が振られていること、および別のデータベースである文化的ポーズ意図データベース３３６が、人間の地理的な場所に応じて、そのポーズを人間の願望、または意図に変換する。この変換は、図７について説明されることとなり、本開示のいくつかの実施形態によって、ジェスチャーの差ベースの地理の図を描写する。

【００２８】

上述のシステム、および方法を用いて、意図判断モジュール３３２は、ポーズテンプレートデータベース３３４におけるエントリに基づいて、人間７１０は、彼の手を挙げていると判断し、人間７２０は、彼の手を水平に保持していると判断する。この判断を行った後に、意図判断モジュール３３２は、文化的ポーズ意図データベース３３６のエントリに対して、判断された人間７１０、および人間７２０のポーズを参照する。文化的ポーズ意図データベース３３６は、ポーズが判断され、取得された画像からの場所に基づいて、ポーズを意図にマッピングする。例えば、人間７１０の画像は、ヨーロッパにおいて取得されたが、人間７２０の画像は、日本において取得された。文化的ポーズ意図データベース３３６のそれぞれのエントリは、ポーズを画像が取得された場所において対応する文化的意味へマッピングする。したがって、人間７１０のポーズの意味は、ヨーロッパにおける場所に対応するエントリに基づいて判断されることとなるのに対して、人間７２０のポーズの意味は、日本における場所に対応するエントリに基づいて判断されることとなる。いくつかの実施形態において、文化的ポーズ意図データベース３３６は、いくつかのデータベース間で分散され、それぞれのデータベースは、異なる場所に対応する。適切なデータベースは、画像が取得された地理的な場所に基づいて参照することができる。車両１１０内で意図判断サービス３３０が部分的、または完全にインストールされている状況において、プロセッサ２０２は、車両１１０が新しい地理的位置に入ったこと（例えば、車両１１０が境界を越えて、異なる都市、州、国、公国などへと入った場合）を検出することに応答して、文化的ポーズ意図データベース３３６から自動的にデータをダウンロードすることができる。これは、外部ネットワーク（例えば、ネットワーク１２０）を介して文化的ポーズ意図データを要求することにおいて待ち時間を回避することによって、システムの効率を改善する。

【００２９】

人間１１４のポーズを判断した後に、ポーズの文化的意味、および人間１１４が上述の方法において注意散漫であるかなどの他の情報を任意に判断し、意図判断モジュール３３２は、人間１１４の意図を予測することが可能である。図８は、本開示のいくつかの実施形態によって、いくつかの例示的なポーズを予測される意図へマッピングするためのフローダイヤグラムを描写する。実際面では、図８に示されるデータフローは、人間が人間検出モジュール３３３によって検出される毎に始める必要はない。例えば、人間が人間検出モジュール３３３によって検出されたが、人間が車両１１０から遠いか、または車両１１０が進行している経路に沿って遠いと判断された場合は、人間の意図は、車両１１０の進行には重要ではなく、結果として、判断する必要はない。したがって、いくつかの実施形態において、意図判断サービス３３０のプロセッサ２０２は、人間検出モジュール３３３が人間１１４を検出した場合に、人間１１４の活動は、人間１１４の安全性のために、車両１１０が経路に沿ってその進行を変更する必要が出てくることになるか、影響を与え得るかの初期予測を判断することができる。

【００３０】

意図判断サービス３３０が、人間１１４の活動が車両１１０の進行に重要であり得るという初期予測を判断する場合、上述のように、意図判断モジュール３３２は、人間１１４のポーズ、およびいくつかの実施形態においてその対応する意味を判断する。したがって

、データフロー８１０に示されるように、意図判断モジュール３３２は、人間１１４が左腕を挙げていると判断する。意図判断モジュール３３２は、そこから人間１１４の予測された意図を判断する。いくつかの実施形態において、ポーズテンプレートデータベース３３４は、上述のように、ポーズとマッチングする意図を示し、したがって、意図判断モジュール３３２は、そこから、データフロー８１０において示されるように、人間が横断することを待っているなどの人間１１４の意図を判断する。

【００３１】

ポーズテンプレートデータベース３３４におけるポーズ意図マッピングは、追加のソースからの情報によって補足されてよい。例えば、意図判断モジュール３３２は、ディープラーニングを用いて、ポーズ（および人間１１４から道路までの距離などの潜在的な追加の情報）を人間１１４の意図に変換することができる。例えば、ディープラーニングフレームワークをトレーニングする分類された歩行者の意図（横断している、および横断していないなど）のデータセットが与えられた場合、意図判断モジュール３３２は、人間１１４が横断するつもりか、または横断しないつもりでいるか、またどんな信頼を持っているかをリアルタイムで取得された一連の画像の次の画像からのデータを予測することが可能である。

【００３２】

さらに、意図判断モジュール３３２は、人間１１４の動き、ボディーランゲージ、社会基盤との相互作用、活動、および行動における統計分析を実行し、人間１１４の以前の行動、人間１１４の分類をタイプ別（例えば、1人の歩行者、グループ、サイクリスト、身体障害者、子供など）、歩行者の速度、相互作用の背景（例えば、横断歩道、通りの中央、都市、田舎など）などの異なる入力を与えることで、人間１１４の将来の行動を判断することができる。さらに、意図判断モジュール３３２は、心理的行動モデルを適用し、意図予測を改善することができる。心理的行動モデルは、歩行者の動き、ボディーランゲージ、および行動のコンピュータビジョンから、歩行者の意図の新規な行動に関する心理学的モデルを用いて、補完または拡張される、上述の実施形態などのボトムアップ、データ駆動のアプローチの能力に関連する。さらにまた、意図判断モジュール３３２は、人間１１４の会話を分析するマイクなどのカメラ１１２以外のセンサー（例えば、センサー１１３）を用いて、その予測を改善することができる。これらの改善を合わせることで、データフロー８２０、データフロー８３０、およびデータフロー８４０は、意図判断モジュール３３２によって判断されたとして、意図ポーズから意図への変換のさらなる例を示す。

【００３３】

車両１１０の危険ゾーンへ近いことは、人間１１４に対する意図の判断をトリガーするかにおける唯一の要因ではない。図９は、本開示のいくつかの実施形態によって、目印を参照することで、その目印を参照して人間の方向を判断することに用いられるキーポイントを描写する。画像９１０において、人間１１４は、上述のシステム、および方法を用いて判断されたように、縁石に背を向けている。これに基づいて、意図判断モジュール３３２は、人間１１４が道路に近づく可能性が低いと判断することができ、したがって、車両１１０は、人間１１４の意図に基づいてその経路を変更する必要がない。しかしながら、画像９２０において、人間１１４は縁石の方に向いており、画像９３０においては、人間１１４は縁石に対面している。これらの活動に基づいて、画像判断モジュール３３２は、人間１１４の意図が道路に入る場合は、車両１１０が人間１１４に対して危険を生じさせないように命令されることを確実とするよう、人間１１４の意図が判断されなければならないと判断させることができる。

【００３４】

上述の説明はマクロレベルでキーポイントを分析している一方で、キーポイント判断モジュール３３７は、人間１１４の意図を判断する場合に、意図判断モジュール３３２に対してよりロバストな情報を提供するようキーポイントを粒状に判断し、処理することができる。図１０は、本開示のいくつかの実施形態によって、人間の身体のさまざまな部分を特定するためにキーポイントクラスターを描写する。画像１０１０は、人間の顔の例示的

なキーポイントを図示している。より多くの数のキーポイントを使用することによって、意図判断モジュールは、意図判断モジュール３３２によって人間１１４の意図を判断することが可能となるより多くの情報を提供する、人間１１４の感情状態（例えば、しかめている、うれしいなど）を判断することが可能である。同様に、意図判断モジュール３３２は、画像１０２０に図示されるような腕、画像１０３０に図示されるような脚、および画像１０４０に図示されるような身体において、より多くのキーポイントを判断することによって、より正確なジェスチャー、および類似物を判断することが可能である。

【００３５】

図１１は、本開示のいくつかの実施形態によって、ビデオフィードから受信された画像を意図判断へ変換するためにフローダイヤグラムを描写する。プロセス１１００は、意図判断サービス３３０のプロセッサ２０２が、ビデオフィードから（例えば、上述の図１を参照して説明したように、ネットワーク１２０を経由してカメラ１１２から）複数の一連の画像を取得すること１１０２で始まる。画像は、所定の時間の長さ、または現在の時間に対して所定の数のフレームに対して取得することができる。人間が（例えば、人間検出モジュール３３３によって）検出されない画像は、破棄されてよい。次に、意図判断サービス３３０のプロセッサ２０２は、（例えば、図３を参照し、上述したように、キーポイント判断モジュール３３７を実行することによって）複数の一連の画像のそれぞれの画像における人間に対応するそれぞれのキーポイントを判断する１１０４。

【００３６】

次に、意図判断サービス３３０のプロセッサ２０２は、（例えば、図３について上述したように、キーポイント集約モジュール３３１を実行することによって）それぞれの画像に対するそれぞれのキーポイントを人間のポーズへと集約する１１０６。画像の中に複数の人間がいる場合は、それぞれの人間（または上述のように、それぞれの人間の誰かの意図が車両１１０の動作に影響を与える場合がある）に対するキーポイントが集約され、それぞれに対してポーズが判断される。次に、意図判断サービス３３０のプロセッサ２０２は、例えば、意図判断モジュール３３２を実行させ、クエリをポーズテンプレートデータベース３３４へ伝送し、ポーズを候補ポーズから意図へ変換する複数のテンプレートポーズと比較することによって、ポーズを意図テンプレートと比較する１１０８。次に、意図判断モジュール３３２は、マッチングするテンプレートが存在するかを判断する１１１０。例えば、意図判断モジュール３３２は、人間の意図、またはマッチングするテンプレートを特定することができないことのいずれかを示す応答メッセージをデータベースから受信する。応答メッセージが人間の意図を示す場合は、意図判断モジュール３３２は、マッチングするテンプレートがあったと判断し、判断された意図に対応するコマンドを出力（例えば、人間１１４の動きを回避するために旋回する、人間１１４が横断を可能とするよう減速するなど）する１１１２。

【００３７】

応答メッセージがマッチングするテンプレートを特定することができないことを示す場合は、意図判断サービス３３０の意図プロセッサ２０２は、車両１１０に安全モードに入るようコマンドする１１１４。例えば、意図判断サービス３３０は、安全モード動作のデータベース３３５を含むことができる。上述のように、意図判断サービス３３０は、一連の画像から、人間の位置、ならびに画像内の他の障害物および特徴を判断することができる。安全モード動作データベース３３５は、検出された障害物および特徴、さらにそれらの障害物および特徴、且つ車両１１０から人間の相対距離に基づいて、人間１１４の意図の知識がない場合に取るべき安全動作を示す。例えば、安全動作モードに入った場合に、車両に、旋回するか、加速するか、移動の停止をするか、オペレータへ制御の提供をするか、ホーンを鳴動させるか、または音を通じてメッセージの伝達するのか少なくとも１つを実行するようコマンドすることを含む。カメラ１１２の視野内で検出された物体に基づいて判断されるように、安全動作モードは、人間１１４の意図を判断するか、または人間１１４がカメラ１１２の視野を離れるか、または危険区域を離れるまで、入っていることができる。

【００３８】
（追加の構成を考慮）
本明細書を通じて、複数のインスタンスは、単一のインスタンスとして説明されるコンポーネント、オペレーション、または構造を実装することができる。１つまたは複数の方法の個々の動作が別個の動作として図示、および説明されているが、個々の動作の１つまたは複数を同時に実行することができ、動作を図示されている順序で実行する必要はない。同様に、単一のコンポーネントとして提示される構造、および機能は、別個のコンポーネントとして実装されてよい。これら、および他の変形例、修正、追加、および改善は、本明細書の主題の範囲内にある。
【００３９】
本明細書では、特定の実施形態は、論理もしくはいくつかのコンポーネント、またはモジュールもしくは機構を含むものとして説明される。モジュールは、ソフトウェアモジュール（例えば、機械可読媒体上、または伝送信号内で実施されるコード）、またはハードウェアモジュールのいずれかを構成することができる。ハードウェアモジュールは、特定の動作で実行することが可能な有形のユニットであり、特定の方法で構成、または配置することができる。例示的な実施形態において、１つまたは複数のコンピュータシステム（例えば、スタンドアローン、クライアント、またはサーバコンピュータシステム）、またはコンピュータシステムの１つまたは複数のハードウェアモジュール（例えば、プロセッサ、またはプロセッサのグループ）は、本明細書に記載される特定の動作を実行するよう動作するハードウェアモジュールのように、ソフトウェア（例えば、アプリケーション、またはアプリケーションの一部）によって構成されてよい。
【００４０】
さまざまな実施形態において、ハードウェアモジュールは、機械的、または電子的に実装されてよい。例えば、ハードウェアモジュールは、永続的に構成される専用の回路、または論理（例えば、フィールドプログラマブルゲートアレイ（FPGA）、または特定用途向け集積回路（ASIC）などの特定用途プロセッサとして、特定の動作を実行する）を備えてよい。また、ハードウェアモジュールは、ソフトウェアによって一時的に構成されるプログラム可能な論理、または回路（例えば、汎用プロセッサ、または他のプログラム可能なプロセッサ内に包含されるもの）を含むことができ、特定の動作を実行する。専用の永続的に構成される回路、または一時的に構成される回路（例えば、ソフトウェアによって構成される）において、ハードウェアモジュールを機械的に実装する判断は、コスト、および時間の考慮によって駆動されることができると認識されることとなる。
【００４１】
したがって、「ハードウェアモジュール」という用語は、有形のエンティティを包含することが理解されるべきであり、そのエンティティは、つまり物理的に構築され、永続的に構成される（例：ハードワイヤ化）エンティティ、または一時的に構成（例：プログラム化）され、特定の方法で動作する、または本明細書で説明する特定の動作を実行する。本明細書で用いられるように、「ハードウェア実装モジュール」は、ハードウェアモジュールを参照する。ハードウェアモジュールが一時的に構成（例えば、プログラム化）される実施形態を考慮すると、それぞれのハードウェアモジュールは、時間内に、任意の１つのインスタンスで構成、またはインスタンス化される必要はない。例えば、ハードウェアモジュールがソフトウェアを用いて構成される汎用プロセッサを含む場合、汎用プロセッサは、それぞれ異なるハードウェアモジュールとして、異なる時間で構成されてよい。したがって、ソフトウェアは、例えば、ある瞬間に特定のハードウェアモジュールを構成し、別の瞬間に異なるハードウェアモジュールを構成するようプロセッサを構成することができる。
【００４２】
ハードウェアモジュールは、情報を他のハードウェアモジュールへ提供したり、情報を他のハードウェアモジュールから受信したりすることが可能である。したがって、説明されるハードウェアモジュールは、通信可能に結合されているとみなすことができる。複数

のそのようなハードウェアモジュールが同時に存在する場合、ハードウェアモジュールを接続する信号伝送を通じて（例えば、適切な回路、およびバスを超えて）通信を実現することができる。複数のハードウェアモジュールが異なる時間に構成、またはインスタンス化される実施形態において、そのようなハードウェアモジュール間の通信は、例えば、複数のハードウェアモジュールがアクセスするメモリ構造の中の情報の格納、および検索を通じて達成することができる。例えば、１つのハードウェアモジュールは、動作を実行し、その動作の出力を、それが通信可能に結合されているメモリデバイスの中に格納することができる。次に、さらなるハードウェアモジュールが、その後で、メモリデバイスにアクセスし、格納された出力を検索、および処理することができる。また、ハードウェアモジュールは、入力デバイスとの、または出力デバイスとの通信を開始することができ、リソース（例えば、情報の収集）を操作することが可能である。

【００４３】

本明細書で説明される例示的な方法のさまざまな動作は、関連する動作を実行するよう一時的に（例えば、ソフトウェアによって）、または永続的に構成される１つまたは複数のプロセッサによって少なくとも部分的に実行することができる。一時的に、または永続的に構成されているかをそのようなプロセッサは、1つまたは複数の操作もしくは機能を実行するよう動作するプロセッサ実装モジュールを構成することができる。本明細書で参照されるモジュールは、いくつかの例示的な実施形態において、プロセッサ実装モジュールを含むことができる。

【００４４】

同様に、本明細書で説明される方法は、少なくとも部分的にプロセッサに実装されてよい。例えば、方法の動作の少なくともいくつかは、１つまたは複数のプロセッサもしくはプロセッサ実装ハードウェアモジュールによって実行することができる。特定の動作性能は、１つまたは複数のプロセッサの間で分散され、単一の機械内に存在するだけでなく、いくつかの機械に渡って配置されてよい。いくつかの例示的な実施形態において、単一のプロセッサ、または複数のプロセッサは、単一の場所（例えば、家庭環境内に、オフィス環境内に、またはサーバーファームとして）配置することができ、一方、他の実施形態おいて、プロセッサは、いくつかの場所を渡って分散することができる。

【００４５】

また、１つまたは複数のプロセッサは、「クラウドコンピューティング」環境において、または「サービスとしてのソフトウェア」（ＳａａＳ）として関連する動作性能をサポートするよう動作することができる。例えば、少なくともいくつかの動作は、（プロセッサを含む機械の例として）コンピュータのグループによって実行されてよく、これらの動作は、ネットワーク（例えば、インターネット）、および１つまたは複数の適切なインターフェース（例えば、アプリケーションプログラミングインターフェース（APIs））を介してアクセス可能である。

【００４６】

特定の動作性能は、１つまたは複数のプロセッサの間で分散され、単一の機械内に存在するだけでなく、いくつかの機械に渡って配置されてよい。いくつかの例示的な実施形態において、１つまたは複数のプロセッサ、またはプロセッサ実装モジュールは、単一の地理的場所（例えば、家庭環境内、オフィス環境内、またはサーバーファーム内）に配置することができる。他の例示的な実施形態において、１つまたは複数のプロセッサ、またはプロセッサ実装モジュールは、いくつかの地理的場所に渡って分散されることができる。

【００４７】

本明細書の一部は、機械メモリ（例えば、コンピュータメモリ）内のビット、またはバイナリデジタル信号として格納されたデータにおける動作のアルゴリズム、または記号表現の観点から提示される。これらのアルゴリズム、または記号表現は、データ処理分野における当業者が彼らの仕事の内容を他の当業者に伝えるよう用いる技術の例である。本明細書で用いられるように、「アルゴリズム」は、所望の結果へ導く首尾一貫した一連の動作、または同様の処理である。この文脈において、アルゴリズム、および動作は、物理量

の物理的操作を伴う。通常、必ずしもそういうわけではないが、そのような量は、機械によって、格納すること、アクセスすること、変換すること、結合すること、比較すること、またはさもなければ操作することが可能である電気的、磁気的、または光学的な信号の形態を取ることができる。主に共通使用の理由のために、「データ」、「コンテンツ」、「ビット」、「値」、「要素」、「記号」、「文字」、「用語」、「数」、または「数字」などの単語を用いてそのような信号を参照すると便利な場合がある。しかしながら、これらの単語は、単に便利なラベルに過ぎなく、且つ適切な物理量に関連付けられている。

【００４８】

特に明記しない限り、「処理」、「コンピューティング」、「計算」、「判断」、「提示」、または「表示」などの単語を用いる本明細書での説明は、1つまたは複数のメモリ内（例えば、揮発性メモリ、不揮発性メモリ、またはそれらの組み合わせ）、レジスタ内、または情報を受信、格納、伝送、または表示する他の機械コンポーネント内に物理（電子、磁気、または光学）量として表されるデータを操作、または変換する機械（例えば、コンピュータ）のアクション、またはプロセスを参照することができる。

【００４９】

本明細書に用いられるように、「一実施形態」、または「実施形態」への参照は、その実施形態に関連して記載される特定の要素、特徴、構造、または特性が少なくとも１つの実施形態において含まれることを意味する。本明細書におけるさまざまな場所において、「一実施形態において」という表現の記載は、必ずしもすべてが同じ実施形態を参照しているわけではない。

【００５０】

いくつかの実施形態は、「結合された」、および「接続された」という表現をそれらの派生語と共に用いて説明することができる。これらの用語は、互いに同義語として意図されていないことを理解されたい。例えば、いくつかの実施形態は、２つまたはそれより多くの要素が互いに直接物理的、または電気的に接触していることを示すよう「接続した」という用語を用いて説明することができる。別の例において、いくつかの実施形態は、２つまたはそれより多くの要素が直接物理的、または電気的接触にあることを示すよう「結合した」という用語を用いて説明することができる。しかしながら、また、「結合した」という用語は、２つまたはそれより多くの要素が互いに直接接触していないが、依然として互いに協働または相互作用することを意味してよい。実施形態は、本文脈において限定されない。

【００５１】

本明細書で用いられる場合、「備える」、「備えている」、「含む」、「含んでいる」、「有する」、「有している」、またはそれらの任意の他の変形例は、非排他的な包含を含むことが意図されている。例えば、要素のリストを備えるプロセス、方法、物品、または装置は、必ずしもそれらの要素のみに限定されず、明示的に記載されていない、またはそのようなプロセス、方法、物品、または装置に内在する他の要素を含んでよい。さらに、それとは反対に明示的に述べられていない限り、「または」は、「包括的なまたは」を参照するものであり、「排他的なまたは」を参照しない。例えば、条件ＡまたはＢは、以下のいずれか１つによって満たされる。Ａは真（または存在する）でＢは偽（または存在しない）であり、Ａは偽（または存在しない）で、Ｂは真（または存在する）であり、さらにＡおよびＢの両方が真（または存在する）である。

【００５２】

さらに、「ａ」または「ａｎ」の使用は、本明細書における実施形態の要素、および構成要素を説明するよう用いられる。これは、単に便宜上、および本発明の一般的な意味を付与するために行われているに過ぎない。本説明は、１つまたは少なくとも１つを含むものであり、また、そうでないことを意味することが明らかでない限り、単数形が複数形を含むように読まれるべきである。

【００５３】

本開示を読むことで、当業者は、本明細書に開示された原理を通じて、人間の意図を判

断するためのシステム、およびプロセスに対するさらなる追加的な代替の構造、および機能的なデザインを理解することになるであろう。したがって、特定の実施形態および用途が図示され、且つ説明されてきた一方で、開示された実施形態は、本明細書に開示された正確な構造、および構成要素に限定されないことを理解されたい。さまざまな修正例、変更例、および変形例は、当業者にとって明らかとなることとなり、添付の特許請求の範囲において定義される趣旨および範囲から逸脱することなく、明細書に開示された方法、および装置の配置、動作、および詳細を作ることができる。

【図面】

【図１】

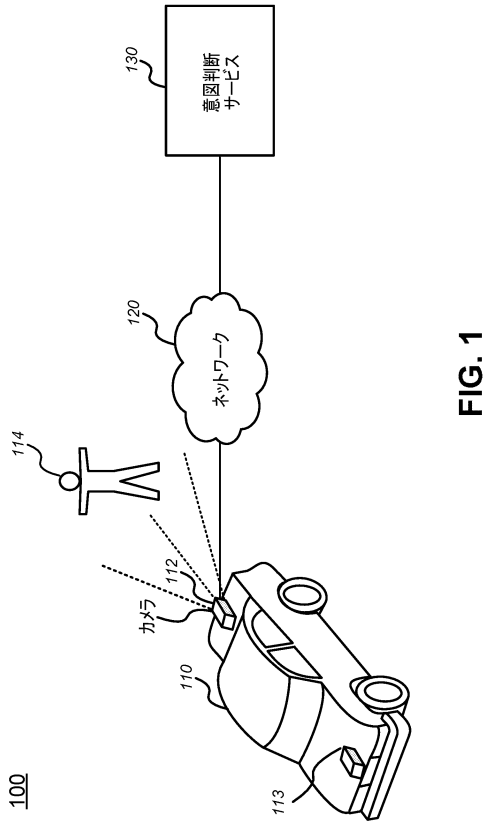

FIG. 1

【図２】

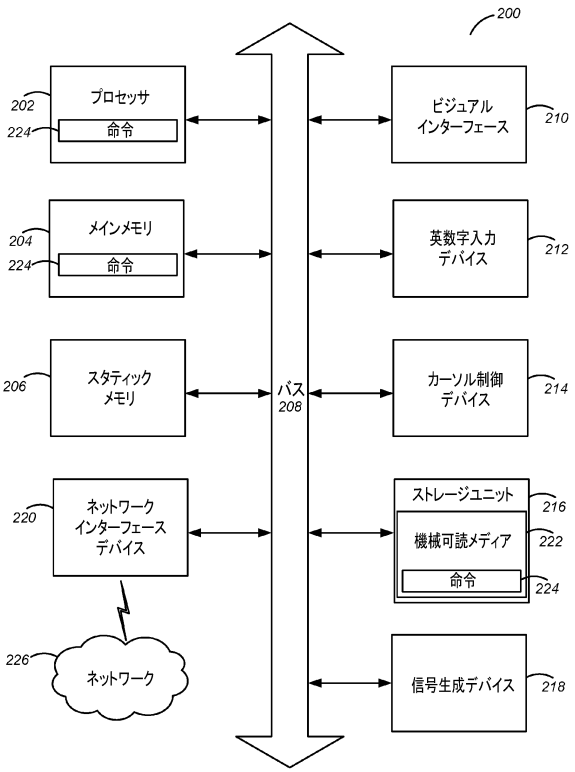

FIG. 2

【図３】

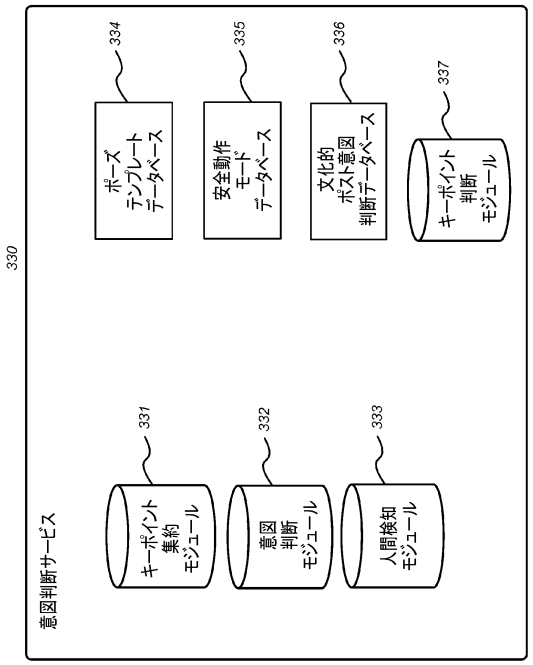

FIG. 3

【図４】

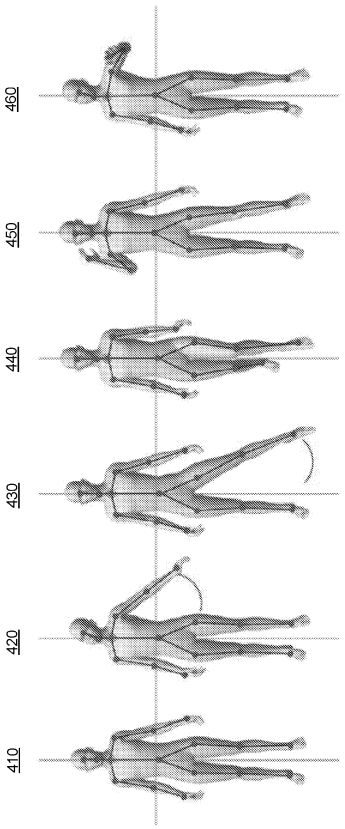

FIG. 4

【図５】

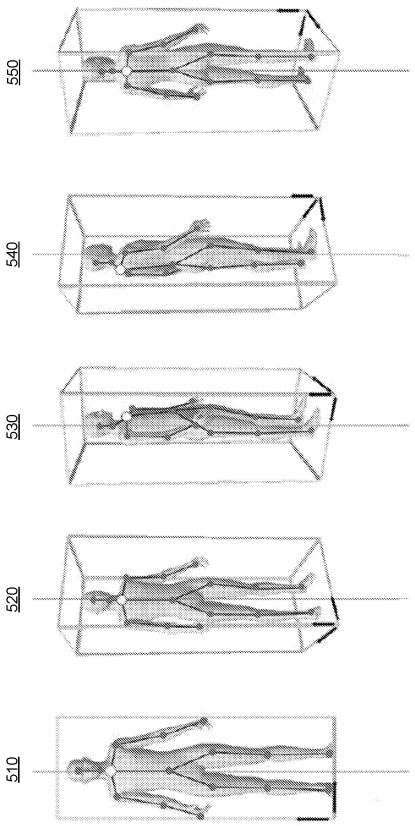

FIG. 5

【図６】

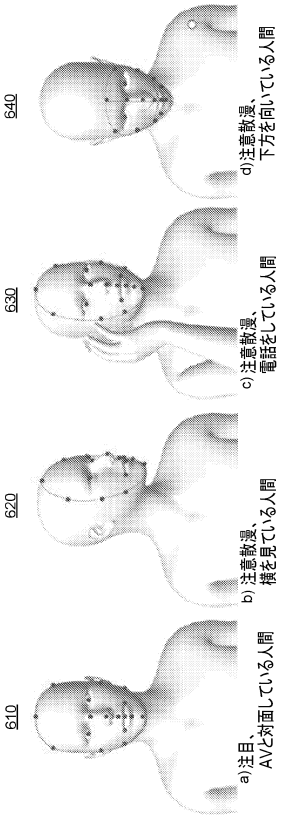

FIG. 6

【図７】

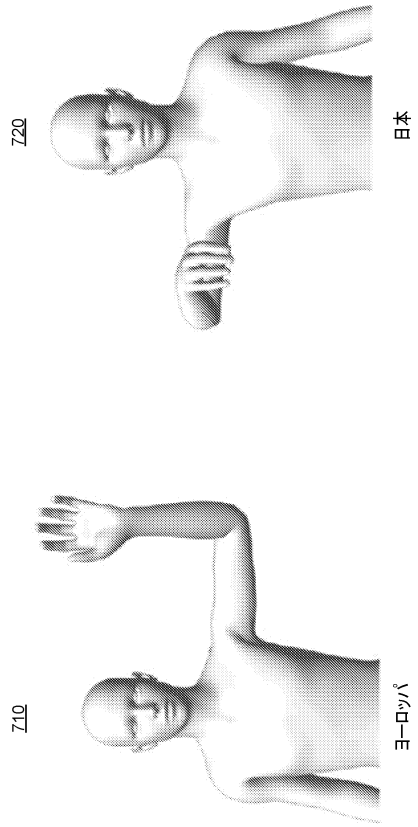

FIG. 7

【図８】

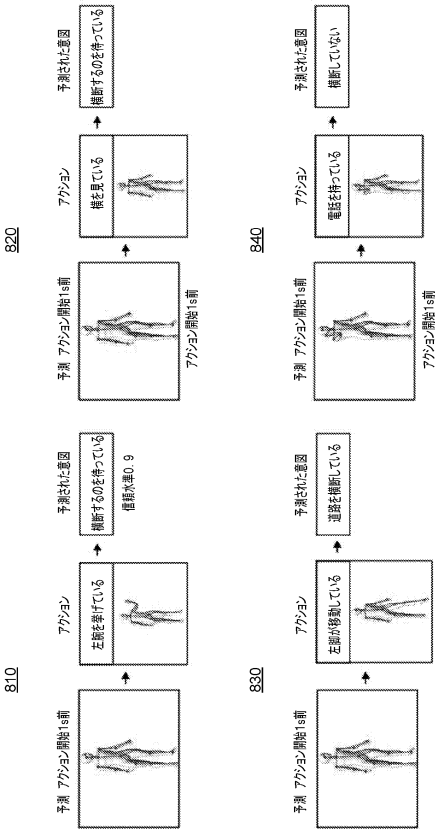

FIG. 8

【図９】

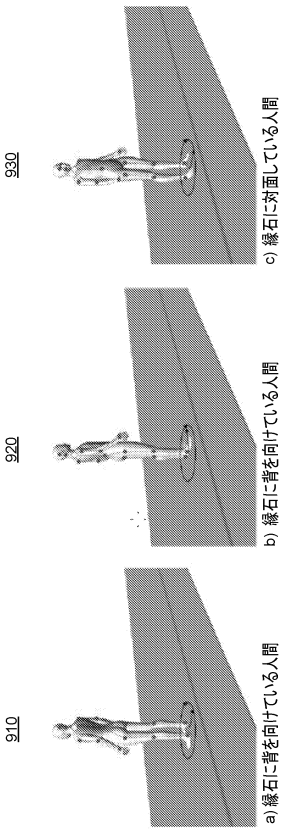

FIG. 9

【図１０】

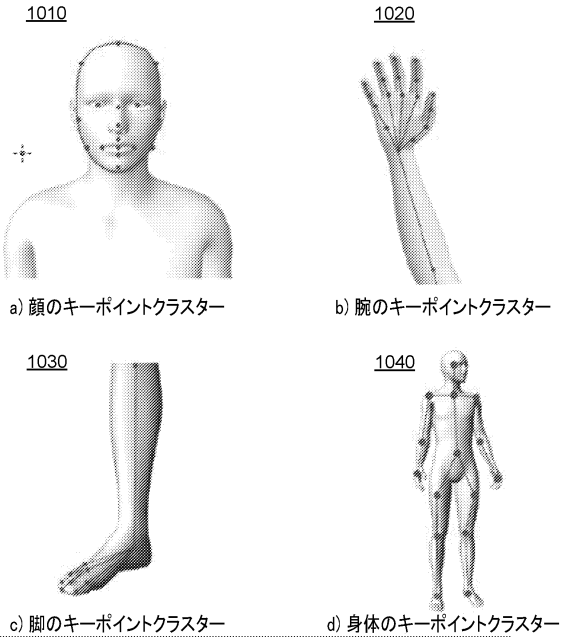

FIG. 10

【図１１】

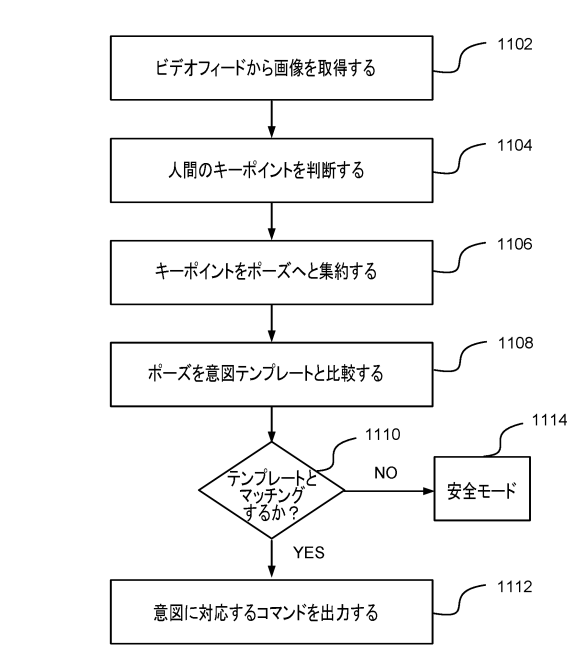
1100
ビデオフィードから画像を取得する
1102
人間のキーポイントを判断する
1104
キーポイントをポーズへと集約する
1106
ポーズを意図テンプレートと比較する
1108
テンプレートとマッチングするか？
1110
NO
安全モード
1114
YES
意図に対応するコマンドを出力する
1112

FIG. 11

フロントページの続き

米国(US)
エー ロンドン ストランド サマセット ハウス（番地なし） ヒューマニシング オートノミー リミテッド内
(72)発明者 レスリー セース ノーテボーム
イギリス ダブリュシー２アール １エルエー ロンドン ストランド サマセット ハウス（番地なし） ヒューマニシング オートノミー リミテッド内
(72)発明者 アダム ジョシュア バーンスタイン
イギリス ダブリュシー２アール １エルエー ロンドン ストランド サマセット ハウス（番地なし） ヒューマニシング オートノミー リミテッド内

合議体
審判長 畑中 高行
審判官 樫本 剛
審判官 高橋 宣博

(56)参考文献
特開２０１３－２０５７８（ＪＰ，Ａ）
特開２０１２－６６０２６（ＪＰ，Ａ）
国際公開第２０１５／１８６４３６（WO，Ａ１）
国際公開第２０１６／１６７３３１（WO，Ａ１）
特開２０１１－１８６５７６（ＪＰ，Ａ）
古橋 竜祐、山田 啓一、“複数フレームの見かけの姿勢に基づく歩行者の横断意図の推定”、電気学会論文誌Ｃ、Vol.133、No.5、2013年5月1日、pp.1069-1075
指田 岳彦、外２名、“３次元人物姿勢情報を用いた睡眠時行動認識への取り組み”、第２０回画像センシングシンポジウム 講演論文集、2014年6月11日、IS2-30、pp.1-5
八尾 智幸、外４名、“テレビ視聴行動からの個人的選好獲得のための行動認識”、電子情報通信学会技術研究報告、Vol.103、No.585、2004年1月16日、PRMU2003-209、pp.65-70
野田 悠介、外３名、“劇場型プレゼンテーション支援システム～Ｊｏｂｓ～”、情報処理学会第７５回全国大会 講演論文集（４）、2013年3月6日、5ZB-3、pp.203-204

(58)調査した分野 (Int.Cl.，ＤＢ名)
G06T 1/00
G06T 7/00 - 7/90
G06V 10/00 - 10/98
G08G 1/16
G06F 3/01
H04N 7/18