

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7675970号
(P7675970)

(45)発行日 令和7年5月14日(2025.5.14)

(24)登録日 令和7年5月2日(2025.5.2)

(51)国際特許分類	F I
G 0 6 T 7/00 (2017.01)	G 0 6 T 7/00 C
G 0 6 F 3/01 (2006.01)	G 0 6 F 3/01 5 1 0
G 0 6 F 3/0346(2013.01)	G 0 6 F 3/01 5 7 0
G 0 6 F 3/038(2013.01)	G 0 6 F 3/0346 4 2 1
G 0 6 F 3/04815(2022.01)	G 0 6 F 3/0346 4 2 2
請求項の数 9 (全23頁) 最終頁に続く	

(21)出願番号	特願2020-557283(P2020-557283)	(73)特許権者	518102470 トヨタ リサーチ インスティテュート, インコーポレイティド アメリカ合衆国, カリフォルニア 9 4 0 2 2, ロスアルトス, エル カミノ リ アル 4 4 4 0
(86)(22)出願日	平成31年4月11日(2019.4.11)	(74)代理人	100099759 弁理士 青木 篤
(65)公表番号	特表2021-522564(P2021-522564 A)	(74)代理人	100123582 弁理士 三橋 真二
(43)公表日	令和3年8月30日(2021.8.30)	(74)代理人	100092624 弁理士 鶴田 準一
(86)国際出願番号	PCT/US2019/026999	(74)代理人	100114018 弁理士 南山 知広
(87)国際公開番号	WO2019/204118	(74)代理人	100117019
(87)国際公開日	令和1年10月24日(2019.10.24)		
審査請求日	令和4年4月11日(2022.4.11)		
審査番号	不服2024-1000(P2024-1000/J1)		
審査請求日	令和6年1月19日(2024.1.19)		
(31)優先権主張番号	15/955,333		
(32)優先日	平成30年4月17日(2018.4.17)		
(33)優先権主張国・地域又は機関	米国(US)		
最終頁に続く			

(54)【発明の名称】 非制約環境において人間の視線及びジェスチャを検出するシステムと方法

(57)【特許請求の範囲】

【請求項 1】

物体を位置特定し識別するためのシステムであって、
個人と関心対象物を含む非制約環境の360度画像をキャプチャするように構成された単眼カメラと、
前記非制約環境の三次元奥行マップを生成するように構成された赤外線奥行センサと、
知識ベースと、
プロセッサであって、
前記360度画像及び前記奥行マップから顕著性マップを生成し、
前記個人からの注目方向を判断し、
前記注目方向を表す三次元方向性ベクトルを生成し、
前記顕著性マップ上に前記個人からの前記三次元方向性ベクトルを投影し、
前記関心対象物と前記三次元方向性ベクトルとの交差を検出し、
前記関心対象物と前記三次元方向性ベクトルとの交差で検出された前記関心対象物のアイデンティティを前記知識ベースで検索するように構成されたプロセッサと、
を備える、システム。

【請求項 2】

前記単眼カメラは、パノラマカメラである、請求項1に記載のシステム。

【請求項 3】

前記360度画像は、つなぎ合わせた複数の画像を含む、請求項1に記載のシステム。

【請求項 4】

前記プロセッサは、前記個人のジェスチャ方向を推定するように更に構成され、前記注目方向は前記ジェスチャ方向から判断される、請求項 1 に記載のシステム。

【請求項 5】

前記プロセッサは、前記個人の視線及びジェスチャ方向を推定するように更に構成され、前記注目方向は前記視線及び前記ジェスチャ方向から判断される、請求項 1 に記載のシステム。

【請求項 6】

前記プロセッサは、前記関心対象物のアイデンティティが前記知識ベース中に発見されない時に学習プロセスを開始するように更に構成される、請求項 1 に記載のシステム。

【請求項 7】

前記学習プロセスは、ニューラルネットワークを備える、請求項 6 に記載のシステム。

【請求項 8】

前記プロセッサは、前記個人の視線を推定するように更に構成され、前記注目方向は前記視線から判断される、請求項 1 に記載のシステム。

【請求項 9】

前記視線は、赤外線センサを使用して推定される、請求項 8 に記載のシステム。

【発明の詳細な説明】**【技術分野】****【0001】**

本開示は概して人間の動作の検出に関するものであり、より具体的には、知能機械による、非制約環境における人間の視線又はジェスチャ動作に関するものである。

【背景技術】**【0002】**

人間の動きの機械検出は、インタラクティブなコンピュータや、機械学習や、人工知能の発展により普及してきた。組み込みや周辺機器としてのカメラを備えたデスクトップコンピュータ、スマートフォン、タブレット、ビデオゲームシステム、及びその他の装置は、人間の動きや、更にはユーザの視線（例えばユーザの目が見ている方向）をも検出できる能力を含むことがある。

【0003】

既知のジェスチャコントロールや視線追跡システムは狭視野カメラに依存し、ユーザの動きが何らかの方法で制約されている環境（例えば運転者又はコンピュータ画面の前にいる人の視線の推定）に向けて設計されている。これらのシステムはユーザの自由な動きを許さず、またほとんどの場合において、ユーザが観測しているのと同じ環境を観測することもできない。より一般的な環境では、ユーザは空間中の他の物体が、ある物体と係合（例えば、移動又は配送）するようロボットに教示してもよいが、これらのシステムは有益ではなく、注目対象物をカメラが捉えていないため、ユーザが見ているものや指さしているものを理解することができない。

【発明の概要】**【0004】**

本開示の実施形態は、非制約環境における視線及びジェスチャの検出システムを含む。システムは個人と関心対象物を含む、非制約環境の 360 度画像を生成するよう構成された画像キャプチャ装置を含んでもよい。システムは更に、非制約環境の三次元奥行マップと知識ベースを生成するよう構成された奥行センサを含んでもよい。360 度画像と奥行マップから環境マップを生成し、360 度画像中の個人の注目方向を判断するよう構成されたプロセッサを含んでもよい。プロセッサは更に、注目方向を表す方向性ベクトルを生成し、環境マップ上に個人からの方向性ベクトルを投影するよう構成されてもよい。プロセッサは方向性ベクトルと関心対象物の交点を検出し、関心対象物に関するデータを使用して、関心対象物のアイデンティティを知識ベースで検索するよう構成されてもよい。

10

20

30

40

50

【 0 0 0 5 】

システムの追加の機能は、画像キャプチャ装置として単眼パノラマカメラを含んでもよい。あるいは、360度画像は、複数の画像をつなぎ合わせたものであってもよい。プロセッサは、視線、個人のジェスチャ方向、又はそれらの組み合わせのうちの少なくとも一つから、注目方向を判断するよう構成されてもよい。奥行センサは赤外線センサ、奥行推定を訓練したニューラルネットワーク、又は自己位置推定(localization)とマッピングの同時実行センサのいずれか一つであってもよい。

【 0 0 0 6 】

本開示のさらなる実施形態では、物体の位置を特定し識別する方法を開示する。方法は個人と関心対象物を含む、非制約環境の360度画像をキャプチャし、環境マップを生成することを含んでもよい。キャプチャされた360度画像中の個人から注目方向が判断されてもよく、注目方向を示す方向性ベクトルが生成されてもよい。方向性ベクトルは環境マップ中の個人から投影され、方向性ベクトルと関心対象物の交差が検出されてもよい。

10

【 0 0 0 7 】

システム及び方法の追加の機能は、関心対象物のアイデンティティを知識ベースで検索し、関心対象物のアイデンティティが知識ベース中に見つからない時は学習プロセスを開始することを含んでもよい。学習プロセスはニューラルネットワークであってもよい。環境マップは、環境の奥行情報を取得し、奥行情報を360度画像と組み合わせることにより生成してもよい。奥行情報は赤外線センサ、自己位置推定とマッピングの同時実行プロセス、又はニューラルネットワークから取得してもよい。環境マップは顕著性マップ(saliency map)であってよく、そして/又は環境のセマンティックパーシング(semantic parsing)を含んでもよい。システムと方法は更に、視線及び/又は個人のジェスチャ方向を推定してもよく、注目方向は、視線及び/又はジェスチャから判断されてもよい。視線は赤外線センサから取得してもよい。

20

【 0 0 0 8 】

本開示の別の実施形態では、物体を位置特定し(locating)識別するためのシステムを含んでもよい。システムは個人と関心対象物を含む、非制約環境の360度画像をキャプチャするよう構成された単眼カメラを含んでもよい。赤外線奥行センサは、非制約環境の3D奥行マップを生成するよう構成してもよい。システムは更に知識ベースとプロセッサを含んでもよい。プロセッサは360度画像と奥行マップとから顕著性マップを生成し、個人の注目方向を判断するよう構成してもよい。プロセッサは更に、注目方向を表す三次元方向性ベクトルを生成し、環境マップ上に個人からの三次元方向性ベクトルを投影するよう構成してもよい。プロセッサは更に三次元方向性ベクトルと関心対象物の交点を検出し、キャプチャした関心対象物に関する画像データを使用して、関心対象物のアイデンティティを知識ベースで検索するよう構成されてもよい。

30

【 図面の簡単な説明 】

【 0 0 0 9 】

装置、システム、及び方法の実施形態が添付図面中の図画に示されるが、これらは例示的及び非限定的であることを意図し、同様の参照は同様の又は対応する部分を参照するよう意図される。

40

【 0 0 1 0 】

【 図 1 】 本開示の実施形態による知的機械によって人間による視線及び/又はジェスチャ動作を検出するシステム及び方法を実施する状況を示す、非制約環境の概念図である。

【 図 2 A 】 本開示の実施形態による、個人のジェスチャを観測する画像キャプチャ装置の概念図である。

【 図 2 B 】 本開示の実施形態による、個人のジェスチャを観測する画像キャプチャ装置の概念図である。

【 図 3 A 】 本開示の実施形態による、360度パノラマ画像、及び2D画像及び奥行センシングデータからシステムにより結果的に生成された顕著性マップの概念図である。

【 図 3 B 】 本開示の実施形態による、ヒートマップの概念図である。

50

【図4】本開示の実施形態による、人間による視線、及び/又はジェスチャ動作を検出するシステムを示す。

【図5】本開示の実施形態による、視線及びジェスチャ検出を使用して物体を位置特定し識別する方法のフローチャートである。

【図6】本開示の実施形態による、視線判断の精緻化を示す、さらなるフローチャートである。

【発明を実施するための形態】

【0011】

本開示は、非制約環境において物体を予測し、位置特定し、識別するために、人間による視線及びジェスチャを推定し追跡するシステム及び方法を記載する。システムは、人間のジェスチャ、視線、認識音声、及びそれらの任意の組み合わせの解釈に基づいて、物体とその位置を識別するために、高度な撮像技術、機械学習、及び人工知能に依存してもよい。本開示の実施形態は、以下に、添付図面を参照してより詳細に説明される。しかしながら、上記は、多くの異なる形態で実施されてもよく、本明細書内で示す実施形態に限定するものと解釈されるべきではない。

10

【0012】

本開示は、物体を位置特定し識別するために、人間による視線又はジェスチャに加えて、個人の周囲の環境をキャプチャするシステム及び方法を提供する。一実施形態によれば、360度単眼カメラは、非制約環境の2次元(「2D」)360度画像を提供してもよい。非制約環境の個人や物体の奥行推定パラメータを提供するために、キャプチャされたシーンを強化するための奥行マップを作成するために、一つ以上の赤外線奥行センサが使用されてもよい。

20

【0013】

システムは、個人の指さし方向等の、視線及び/又はジェスチャを示す三次元(「3D」)ベクトルを判断するために2D画像と奥行マップを処理してもよい。システムは、シーンの中で注目された可能性の高いエリアを定義するための顕著性マップを生成してもよい。顕著性マップは、個人が見ている又はジェスチャで示している物体又は位置をより高精細に予測することにおいて、技術的な利点を提供し得る。3Dベクトルが個人の周囲の環境中に識別された任意の物体及び位置と交差するかを判断するために、観測された環境画像上に3Dベクトルが投影されてもよい。システムが一度十分な確からしさで物体を位置特定し識別したなら、システムは更に物体と相互作用してもよい。もしも物体が位置特定されたがシステムにとって未知であれば、システムを更に教育し知識ベースを拡張するための機械学習プロセスが開始されてもよい。

30

【0014】

図1は本開示の実施形態による、非制約環境100、又は知的機械によって人間の視線及び/又はジェスチャ動作を検出するシステム及び方法を実施する状況を示す概念図である。視線及びジェスチャ検出システムは、画像キャプチャ装置110を備えたロボット105のような計算機内に実施されてもよい。環境100は一人以上の個人112、113、114、及び複数の物体115を含んでもよい。一人の方向指示を行っている個人114は、特に、複数の物体115のうち一つの方を見て、又はジェスチャしていてもよい。画像キャプチャ装置110は、個人114の視線方向121をキャプチャしてもよい。画像キャプチャ装置はまた、方向指示を行っている個人114によるジェスチャ122を、キャプチャ、又は検出してもよい。ジェスチャ検出情報は、指をさしている個人114の注目の方向を補う、又は確認するために使用されてもよい。例えば、図1に示すように、非制約環境100中の複数の物体115は、第1の窓117、第2の窓119、及び、二つの窓117と119の間の絵画118を含んでもよい。絵画118は、非制約環境100中の壁に掛けられても、又はスタンド又はイーゼル上に設置されてもよい。

40

【0015】

本明細書中で使用されるように、「非制約環境」の用語は、個人の動作又は行動が、カメラ又は画像キャプチャ装置110の空間的視野に制約されない任意の空間を含んでもよ

50

い。自動車の座席に拘束された運転者、又はコンピュータのカメラの視野のすぐ前に拘束されたコンピュータのユーザとは対照的に、例示する非制約環境 100 中の個人は、領域中を自由に移動してもよく、それでいて画像キャプチャ装置 110 の視野の内部にいることができる。

【0016】

計算機、又はロボット 105 は、画像キャプチャ装置 110 として、又はその構成要素として、360 度単眼カメラ 111 を含んでもよい。360 度単眼カメラ 111 を使用することにより、非制約環境 100 の、実質的に完全な視野が提供される。したがって、個人 112、113、114、又は複数の物体 115 の位置に関わらず、個人 112、113、114 又は複数の物体 115 はカメラ 111 によりキャプチャされてもよい。したがって、個人 112、113、114 及びロボット 105 は環境 100 内を移動してもよく、その時なおカメラ 111 は、非制約環境 100 内の個人 112、113、114 及び複数の物体 115 をキャプチャし得る。

10

【0017】

システムは更に、非制約環境 100 の奥行情報を収集する、追加の環境データ収集能力を含んでもよい。システムは、複数の物体 115、個人 112、113、114、及び、壁、柱、及びその他の構造のような環境の輪郭に関連する環境情報をキャプチャすることにより、奥行マップを生成してもよい。ロボット 105 及び画像キャプチャ装置 110 は、環境 100 をスキャン又は撮像する一つ以上の赤外線（「IR」）センサ 120 を含んでもよい。更に、ロボット 105 及び画像キャプチャ装置 110 は、IR 信号の変位を使用して、システムによって画像キャプチャ装置 110 からの環境中の物体又は個人の奥行の推定に使用され得る奥行画像を生成してもよい。

20

【0018】

ロボット 105、又は他の計算機が移動ロボットであれば、環境の奥行情報は自己位置推定とマッピングの同時実行（「SLAM」）技術を使用して取得してもよい。SLAM 技術は、移動ロボットにより、未知の環境のマップを、環境中を移動しながら構築し得るプロセスを提供する。ある構成では、非制約環境中におけるロボットの位置を推定するために SLAM 技術は撮像装置から抽出された特徴を使用してもよい。上記は、ロボット 105 が環境をマップし、そのためにロボット 105 が環境中の物体の 3D 相対位置を認識している、という状況で行われてもよい。上記は、ロボット 105 が過去に観測したが現在は他の物体やシーンの構造により塞がれている物体を含んでもよい。

30

【0019】

本開示の一実施形態によれば、環境の奥行情報は、図 1 の非制約環境のような非制約環境中における奥行推定を訓練した人工ニューラルネットワークを使用して生成されてもよい。人工ニューラルネットワークは環境の奥行情報を、前述した方法とシステムから、又は上記にて議論した方法とシステムに加えて、生成してもよい。ある構成では、人工ニューラルネットワークは物体情報を使用して訓練を行ってもよい。物体やその環境に関する先験がないシステムに基づいた時、人工ニューラルネットワークは、環境中の各物体又は個人に関する独自の定義的な特徴量を個別に獲得していてもよい。

【0020】

人工ニューラルネットワークは一つ以上の特定のタスクに関して訓練されてもよい。タスクは、人の検出、人体姿勢の推定、顔検出、物体のインスタンス分割、認識と姿勢推定、シーンの奥行の計量推定、及び、材料特性及び形容的説明のようなシーンのさらなる属性への分割を、限定せず含んでもよい。システムは、システムが自信をもって認識又は分割できない新しい物体又は構造を観測してもよい。このような場合は、システムは、抽出した属性と奥行を使用して、各物体に関する独自の定義的な特徴量を個別に獲得可能であってもよい。

40

【0021】

輪郭、及び非制約環境 100 中の輪郭、物体、及び個人をマッピングすることに加えて、システムは、方向指示を行っている個人 114 の視線情報をキャプチャ、又は検出する

50

能力を更に含んでもよい。視線情報は I R センサ、R G B カメラ、及び / 又はロボット 1 0 5 と通信中の一つ以上の光学追跡装置を使用して取得してもよい。例えば、I R 光線が I R センサ 1 2 0 により伝送され、ジェスチャを行っている個人 1 1 4 の目から反射される。反射された I R 光線はカメラ又は光学センサにより検出されてもよい。後述するように、更なる精緻化においては、視野中の視線方向又は関心対象物を更に獲得するために、複数の画像やスキャン又は同種のものを使用して視線データを一定期間観測してもよい。

【 0 0 2 2 】

次に、画像情報は、反射の変化から目の回転を抽出するためにプロセッサにより解析される。プロセッサは角膜反射及び瞳孔の中心を、経時的な特徴として使用してもよい。あるいは、角膜前部及びレンズ後部からの反射を、特徴として追跡してもよい。奥行マップの一部として、又は独立して撮像し処理されたものとして、システムは、ジェスチャを行っている個人 1 1 4 の目の I R 画像を使用して個人の視線方向を判断してもよい。後述するように、システムは、投影された視線及び関心対象物の交差点を探す中で、環境画像マップに渡る個人からの視線を外挿法により推定してもよい。

10

【 0 0 2 3 】

個人の位置と目の位置を観測しキャプチャすることに加えて、システムは更に、方向指示を行っている個人 1 1 4 のジェスチャ 1 2 2 (例えば指をさす姿勢)を検出してもよい。ジェスチャ 1 2 2 は、方向指示を行っている個人 1 1 4 の注目の方向性の識別、識別の補助、又は確認のために使用されてもよい。後述するように、システムは拡張された腕又は手を検出してもよい。方向指示を行っている個人 1 1 4 のジェスチャを行っている腕と体との間の角度に基づいて、システムは、非制約環境 1 0 0 中の方向性の線又はベクトルを、外挿法により推定してもよい。

20

【 0 0 2 4 】

図 2 A は方向指示を行っている個人 1 1 4 の視線 1 3 0 を観測する画像キャプチャ装置 1 1 0 の概念図 2 0 0 の例を示す。視線 1 3 0 を検出することに加えて、画像キャプチャ装置 1 1 0 は、方向指示を行っている個人 1 1 4 の位置又は関心対象物を三角推量により求めるために、方向指示を行っている個人 1 1 4 の指さし方向 1 2 4 又はその他の方向を示すジェスチャを識別してもよい。システムは I R センサ 1 2 0 又は 3 6 0 度単眼カメラ 1 1 1 からの画像マップを解析してもよい。解析に基づき、システムは、方向指示を行っている個人 1 1 4 及び指さし方向 1 2 4 を識別するためにパターン又は形状認識プロセスを適用してもよい。

30

【 0 0 2 5 】

ジェスチャを行っている個人 1 1 4 からの注目方向 1 2 6 を外挿法により推定するために、方向指示を行っている個人 1 1 4、視線 1 3 0 及びジェスチャ 1 2 2 を使用してもよい。システムは、方向指示を行っている個人 1 1 4 の視線 1 3 0 及びジェスチャ 1 2 2 の両方を検出してもよい。システムは方向指示を行っている個人 1 1 4 の体 1 2 8 のような大きなかたまりから拡張された、拡張された腕又は付属物を識別してもよい。視線方向及びジェスチャ方向からのデータは組み合わせられてもよい。システムは、方向性ベクトル 1 3 2 を判断するために、画像キャプチャ装置 1 1 0 からの、方向指示を行っている個人 1 1 4 及び指さし方向 1 2 4 の端部への距離 1 3 6 及び 1 3 7 を使用してもよい。図 2 A の例に示すように、方向性ベクトル 1 3 2 は、個人 1 1 4 の目からジェスチャ 1 2 2 の方向へ、外挿法により推定されてもよい。システムは、画像マップを使用し、方向性ベクトル 1 3 2 が絵画 1 1 8 のような注目対象物と交差することを判断してもよい。注目対象物は方向性ベクトル 1 3 2 と注目対象物の交差点に基づいて識別されてもよい。

40

【 0 0 2 6 】

図 2 B は、システムにより解釈され得る図 2 A を示す。図 2 B に示すように、システムは I R センサ 1 2 0 ' 又は 3 6 0 度単眼カメラ 1 1 1 ' からの画像マップを解析してもよい。方向性ベクトル 1 3 2 ' は、例えば、方向指示を行っている個人 1 1 4 ' の体 1 2 8 '、視線 1 3 0 '、及びジェスチャ 1 2 2 ' を三角測量することにより推定してもよい。システムは画像キャプチャ装置 1 1 0 ' からの、ジェスチャを行っている個人 1 1 4 ' の目への距離

50

136'と、視線130'とジェスチャ122'との交差点への距離137'を使用して、方向性ベクトル132'を推定してもよい。方向性ベクトル132'は視線130'に沿って、関心対象物118'(例えば絵画118)との交差点へ延長してもよい。図2A及び図2Bに示された例は視線及びジェスチャ認識の双方を使用し得るが、システムは視線のみを使用して関心対象物を位置特定及び識別してもよく、又は、視線検出の正確性を補うためにジェスチャ情報を使用してもよい。ジェスチャ情報のみで使用されてもよい。例えば、ジェスチャ122'のみを使用して、また個人114'からのジェスチャ方向122'のみにけるジェスチャベクトル134'を外挿法により推定してもなお、ジェスチャベクトル134'は関心対象物118'と交差する。

【0027】

システムは、シーンの中で注目された可能性の高いエリアを定義するための顕著性マップを生成してもよい。システムは、個人による視線及び/又はジェスチャ方向を示す3Dベクトルを顕著性マップに生成し投影するために、画像と奥行マップを解析してもよい。図3Aは、360度パノラマ画像300、及び2D画像及び奥行センシングデータに基づいてシステムにより生成された、顕著性マップ300'の概念図である。パノラマ画像300は2D画像であってもよい。この例では、パノラマ画像300は、画像キャプチャ装置が環境をスキャンするように、個人112、113、114、窓117、119及び絵画118を直線的に描写している。

【0028】

顕著性マップ300'は360度パノラマ画像300及びIR奥行マップに基づいていてもよい。前述のように、IR奥行マップは画像キャプチャ装置及びIRセンサにより取得されてもよい。顕著性マップ300'は、非制約環境の360度画像のような、関連する視覚的なシーンの顕著性を示す、地形的に配置されたマップであってもよい。画像データ及び奥行センシングデータは顕著性マップ300'を形成するために組み合わせられてもよい。顕著性マップ300'は環境中の関心対象エリアを識別してもよい。システムは個人112'、113'、114'、物体117'、119'(例えば窓117と119)及び関心対象物118'(例えば絵画118)を関心対象エリアとして識別してもよい。説明目的により、顕著性マップ300'中の関心対象物又はエリアはパノラマ画像300中にて反対のコントラストで描かれている。システムはこれらの物体及び個人を、周囲のエリアと非関心対象の既知のエリアとの差分から、関心対象と判断してもよい。

【0029】

例えば、システムは、壁、床、及び/又はその他の、画像データ中に連続した大きなエリアとして見える場所に比較して、環境中の識別可能な物体により関心があってもよい。壁及び/又は床のような低関心のエリアは環境中の大きなエリアに広がる、連続した、そして均質なピクセルデータであることから識別されてもよい。個人及び物体は、周囲のエリアのピクセルデータと比較して、異なっていてコントラストのあるピクセルデータであるため、関心対象物として識別されてもよい。システムは、個人の注目対象の物体又は位置を識別するために、顕著性マップ300'を視線及び/又はジェスチャ情報と共に解析してもよい。

【0030】

3D方向性ベクトル138'はシステムにより生成され顕著性マップ300'上に投影されてもよい。方向性ベクトル138'は、任意の不確実性境界を含む、観測されたシーンの残りの部分へ投影されてもよい。データが3Dモデルへ可視化された時、ノイズ及びその他の要因による不確実性がデータ中に存在し得る。不確実性のデータは、不確実性のレベル(例えば忠実度)を示すために、他のデータと共に提供されてもよい。不確実性境界は、視線検出システムにより、視野又は人の相対的姿勢から推定されてもよい。例えば、人がカメラの方向を見ていない時、カメラと個人の目の間の直接的な視線が欠如することにより、不確実性境界は大きくなり得る。一方、もしも人がカメラに近く、正面に平行に位置していれば、不確実性境界はより小さくなり得る。不確実性境界は推定3Dベクトルを、眼球を頂点とした円錐へと、効果的に変更し得る。

10

20

30

40

50

【 0 0 3 1 】

ジェスチャ検出及び関連情報は不確実性境界を狭めるために使用されてもよい。例えば、方向指示を行っている個人がカメラの方を向いていなく、そのため個人の目がシステムにより検出できない時、ジェスチャ検出及び情報を使用して、関心対象物を位置特定し識別する解析を補ってもよい。システムは、カメラの方向を向いていない個人の頭の方向に基づいて、広い注目の円錐を生成可能であってもよい。ジェスチャが検出された場合、システムは、検出されたジェスチャの方向性に基づいて、ジェスチャによる注目の円錐を生成してもよい。そしてシステムは視線による注目の円錐とジェスチャによる注目の円錐の交差又は重複を判断してもよい。二つの注目の円錐の交差はより狭い3Dベクトルの円錐を提供してもよく、システムはその中で注目対象物の可能性がある物体の位置特定と識別を行える。

10

【 0 0 3 2 】

前述のように、システムは、奥行マップデータを使用して、画像キャプチャ装置から方向指示を行っている個人114'への距離を判断してもよい。判断した距離に基づいて、システムは、関心対象物118'との交差を識別するために、ジェスチャを行っている個人114'の場所からの、検出された視線又はジェスチャの方向の方向性ベクトル138'を投影してもよい。図3の例では、システムは、方向性ベクトル138'の経路上に関心対象物118'を識別してもよい。したがって、システムは方向性ベクトル138'との交差に基づいて、関心対象物118'を、ジェスチャを行っている個人114'の注目対象物として識別してもよい。

20

【 0 0 3 3 】

前述のように、ジェスチャ情報はシステムによって観測されてもよく、システムが関心対象物118'を位置特定し識別するために使用されてもよい。ジェスチャベクトル140'は、顕著性マップ300'からの情報を使用して生成されてもよい。ジェスチャベクトル140'は、個人114'の体に対するジェスチャを行っている腕の角度に基づいてもよく、顕著性マップ300'中の方向に外挿法により推定されてもよい。図3に示すように、ジェスチャベクトル140'はなお同じ関心対象物118'と交差してもよい。正確性と確実性を向上するために、システムは視線とジェスチャベクトルを使用して、三角推量するか、もしくは、関心対象物との交差を検出してもよい。

【 0 0 3 4 】

不確実性を減少させるための更なる取り組みとして、システムは、方向指示を行っている個人の視線の、一定期間の離散的な視線のスナップショットを組み合わせたヒートマップ(heat map)を生成することにより、視線検出を精緻化してもよい。人間の目は常に動いているだけでなく、方向指示を行っている個人の注目は変化し得る。例えば、方向指示を行っている個人は瞬間的にロボットや、別の個人や、又は別の関心対象物の一部へと戻って注視することがあり得る。方向指示を行っている個人の目の一枚の画像キャプチャでは、正しい注目対象物を識別するには充分ではない、又は充分に正確ではないことがあり得る。

30

【 0 0 3 5 】

図3Bは、本開示の一実施形態による、ヒートマップ301の概念図である。方向指示を行っている個人の、一定時間の視線の一つ以上の画像又はスキャンを組み込んだヒートマップ301を使用することにより、さらなるレベルの精度を提供することができる。ヒートマップ301は、方向指示を行っている個人が最も高い頻度で見ているエリアを識別するために、図3Bに示されるように、期間 t に渡って取得された一連の視線画像又はスキャンからのデータを組み合わせて視線の範囲142''として形成してもよい。図3Bに示されるエリア112''、113''、117''、118''、119''は方向指示を行っている個人がその特定のエリアを見ていると判断され得る画像スキャンの密度及び/又は頻度を示す。システムは、スキャン期間 t において、方向指示を行っている個人114''は一つのエリア118''を、環境中の他のエリアよりも高い頻度で注視したと推定してもよい。例示的なヒートマップ301によれば、スキャンの密度に基づき、最も多い数スキャンさ

40

50

れた視線は、関心対象エリア 1 1 8 ' ' を示してもよい。

【 0 0 3 6 】

より頻度の少ない関心対象エリアが、隣接したエリア 1 1 7 ' '、1 1 9 ' ' (それぞれ窓 1 1 7 及び 1 1 9 を表す) に検出されてもよい。エリア 1 1 2 ' '、1 1 3 ' ' は環境中の他の個人 1 1 2、1 1 3 を表し、より低い頻度スキャンされたデータとして示されてもよい。ヒートマップ中に特定された、より頻度の低いエリアは、方向指示を行っている個人 1 1 4 ' ' が、期間 t の間に瞬間的に環境中の窓及び他の個人を見つめたことを示してもよい。図 3 B における方向指示を行っている個人 1 1 4 ' ' の表現は、追加の文脈及び参照の枠組みを提供するため、説明のみを目的として提供されることが理解されなければならない。方向指示を行っている個人が、システムが視線データを収集した期間 t の間に自分自身を注視することは起こりにくい。

10

【 0 0 3 7 】

ヒートマップ 3 0 1 は、ヒートマップ中の関心対象エリアと顕著性マップ中に識別された物体の交差を判断するために、顕著性マップと組み合わせられてもよい。システムは環境中の物体が物理的にどこに位置しているのか (例えば顕著性マップ) を識別し、また方向指示を行っている個人が注視している方向又はジェスチャしている方向 (例えばヒートマップ) を判断してもよい。二つのセットのデータを組み合わせることで、関心対象物の位置をあるレベルの確かさで示してもよい。

【 0 0 3 8 】

システムが前述の任意の方法を使用して関心対象物を識別したら、システムは識別した物体のアイデンティティを判断してもよい。システムは、識別された関心対象物 1 1 8 ' と知識ベース中の類似物体を関連付けるために、画像キャプチャ装置により取得した画像と奥行データを使用して知識ベース (例えばニューラルネットワークにより提供された特徴表現) を検索してもよい。ある程度以上の確からしさで一致がなされた場合は、システムは識別された関心対象物 1 1 8 ' を一致した物体と関連付けてもよい。

20

【 0 0 3 9 】

しかしながら、もしも関心対象物が一致しなければ、システムは、物体のアイデンティティを識別するため学習プロセスを開始してもよい。例えば、システムは識別された関心対象物が未知であることを示し、方向指示を行っている個人又は他のオペレーターへ情報を要求するよう促してもよい。システムの要求は、音声又は視覚的プロンプト、又は任意の他の入力要求を含んでもよく、これによりシステムは関心対象物のアイデンティティに関する情報を得る。システムは、物体に関する画像データと既知の物体のアイデンティティとを関連付けるために、機械学習技術及びユーザ入力を使用してよく、それにより関心対象物 (例えば絵画) 及び類似画像データを持つ観測された物体はシステムにとって既知となり得る。

30

【 0 0 4 0 】

システムはユーザ入力を使用して知識ベースへ識別情報を提供してもよい。例えば、システムは音声認識及び自然言語理解を使用して、ユーザ入力から物体の名前及び他の情報を取得してもよい。例えば、システムが知識ベース中の物体を識別できない場合は、システムは識別情報のための音声又は視覚的プロンプトを出力してもよい。プロンプトに回答して、個人は音声により物体の識別を行ってもよい。システムは個人による識別音声の音声信号を録音してもよく、自然言語プロセッサを使用して音声信号を処理して機械により理解可能なテキストへ解析及び転換してもよい。次にシステムは相互に関連がある識別情報を画像データと関連付けて、将来の呼び出しのために関連付けを知識ベースへ格納してもよい。更に、システムは、環境に対する物体の位置、他の識別可能な物体に対する物体の近接又は同種のものといった、物体に関する追加情報を記録してもよい。収集された追加データは、将来の問い合わせ又は環境中の物体の処理において、システムを補助してもよい。

40

【 0 0 4 1 】

システムはまた、コマンド、物体及び各物体のアフォーダンス (affordances) のライブ

50

ラリを学習していても、又はプログラムされていてもよい。アフォーダンスは個人又は機械、及び他の環境間に起こり得る、様々な相互作用を含んでいてもよい。システムは、物体のアイデンティティを学習する際に、物体を、システムが与えられた物体とどのように位置特定し、識別し、相互作用するかの追加の状況を提供するアフォーダンスのセットと関連付けてもよい。例えば、システムがテーブルを認識するよう訓練されていた場合、将来の相互作用のために、特定のアフォーダンスがその物体と関連付けられていてもよい。テーブルという物体のアイデンティティが与えられれば、システムはその物体と特定の特徴を関連付けてもよい。属性の例としては、アイテムを置くことができる平らな面、床よりも高いところにある表面又は同種のもを限定せず含んでもよい。このように、方向指示を行っている個人との将来の相互作用により、システムへ、「テーブルの上に」物体を位置特定する指示が行われてもよい。システムはテーブルのアイデンティティを認識し、関連属性を抽出し、テーブルが地面よりも高いところにある表面を含んでいることを知覚してもよい。

10

【0042】

このようなアフォーダンス及び物体に関連した属性は、システムがまずテーブルを識別し、次にテーブル（とその表面）の周辺の狭い環境中に、要求された物体を位置特定するように、環境情報を更にフィルタするために使用されてもよい。物体と属性及び特徴の関連付けは、本明細書中で説明する訓練中、セマンティックシーンパーシング中、又は他の機械学習ステージ中に成し遂げられてもよい。

【0043】

図4は本開示の実施形態によるシステム400を示す。システム400は専用の計算機405を含んでもよい。前に詳述したように、計算機405は移動ロボットとして実装されてもよく、又は関心対象物を人間の視線とジェスチャ検出を利用して位置特定し識別するために設計され実装された特別目的のコンピュータのような、別個の、統合又は分散計算機であってもよい。計算機405は本明細書中に説明されるシステムのために特別に設計されたデータソース、サーバ、及び/又はクライアント装置を含んでもよい。

20

【0044】

計算機405は人間と相互作用可能な相互作用ロボットを含んでいてもよい。計算機405はまた、ネットワーク402を介して他の装置又はセンサと相互作用できるが、本明細書中に前述したように、非制約環境に配置されている、ラップトップコンピュータ、デスクトップコンピュータ、パーソナルデジタルアシスタント、タブレット、携帯電話、テレビ、セットトップボックス、エアラブルコンピュータ及び同種のもののような装置であってもよい。特定の態様において、計算機405はハードウェア又はハードウェアとソフトウェアの組み合わせを使用して実装されていてもよい。計算機405はスタンドアロン装置、他の実体又は装置に統合された装置、複数の実体に分散されたプラットフォーム、又は仮想環境を実行する仮想装置であってもよい。

30

【0045】

ネットワーク402は、システム400の構成要素間のデータ及び制御情報通信を行うのに適した、（一つ以上の）データネットワーク又は（一つ以上の）インターネットワークを含んでもよい。これはインターネット、プライベートネットワーク、公衆交換電話網又は第三世代携帯電話技術（例えば3G又はIMT-2000）を使用した携帯電話ネットワーク、第四世代携帯電話技術（例えば4G、LTE、MT-Advanced、E-UTRA等）、WiMAX-Advanced（IEEE 802.16m）、及び/又はその他の技術のような通信ネットワークのような公衆ネットワークに加えて、コンピュータシステム400内にて、本明細書中に説明されるように、通信する構成要素間でデータを伝送するために使用され得る任意のスイッチ、ルータ、ハブ、ゲートウェイ及び同種のもを伴う任意の様々な企業区域、大都市区域、キャンパス又は他のローカルエリアネットワーク、又は企業ネットワークを含んでもよい。ネットワーク402はデータネットワークの組み合わせを含んでもよく、厳密にパブリック又はプライベートネットワークに限定される必要はない。システムは外部装置404を含んでもよい。外部装置404はネ

40

50

ットワーク 402 を介して 計算機 405 に接続される コンピュータ 又は他の外部リソース であってもよい。

【0046】

一般的に、 計算機 405 は プロセッサ 406、メモリ 408、ネットワークインターフェース 410、データ記憶部 412、及び一つ以上の入出力インターフェース 414 を含んでもよい。 計算機 405 は更に、入出力インターフェース 414 と接続し得る、リモートコントローラ、通信装置 又は同種のもののような有線 又は無線周辺機器 416 又は他の外部入出力装置を含むか、それらと通信中であってもよい。

【0047】

プロセッサ 406 は、 計算機 405 又はシステム 400 内にて実行する命令を処理可能なプロセッサ 又は処理回路 であってもよい。プロセッサ 406 は一つ以上のシングルスレッドプロセッサ、マルチスレッドプロセッサ、マルチコアプロセッサ 又は同種のものを含んでもよい。プロセッサ 406 はメモリ 408 又はデータ記憶部 412 に格納された命令を処理可能であり本明細書中に説明される機能を提供可能であってもよい。プロセッサ 406 は単一のプロセッサ であってもよく、又は、命令を並列処理するためにシステム 400 とその他のプロセッサと協働する多数のプロセッサを含んでもよい。

10

【0048】

メモリ 408 は 計算機 405 の内部に情報を格納してもよい。メモリ 408 は、ランダムアクセスメモリ (RAM)、フラッシュメモリ、リードオンリーメモリ (ROM)、プログラマブルリードオンリーメモリ (PROM)、消去可能プログラマブルリードオンリーメモリ (EPROM)、レジスタ、等を限定せず含む、揮発性 又は不揮発性メモリ、又はその他のコンピュータにより読み取り可能な媒体を含んでもよい。メモリ 408 は、 計算機 405 の動作を制御しユーザのための機能を実行するよう 計算機 405 を構成することに有用な、プログラム命令、プログラムデータ、実行ファイル、及びその他のソフトウェア及びデータを格納してもよい。メモリ 408 は、 計算機 405 の動作の異なる態様のために、多数の異なる段階及び型のメモリを含んでもよい。例えば、プロセッサは特定のデータ 又は命令へ速くアクセスするためのオンボードメモリ及び/又はキャッシュを含んでいてもよく、また必要に応じてメモリ容量を拡張するために別個のメインメモリ 又は同種のことを必要に応じて含んでいてもよい。これら全ての型のメモリは、本明細書で企図するように、メモリ 408 の一部であってもよい。

20

30

【0049】

メモリ 408 は、一般的に、 計算機 405 により実行された時、当コンピュータプログラムの実行環境 (例えばプロセッサファームウェア、プロトコルスタック、データベース管理システム、オペレーティングシステム、又はそれらの組み合わせを構成するコード) を作成し、フローチャートに説明 又は示されたステップの一部 又は全体、及び/又は本明細書中で示された他のアルゴリズムを実行するコンピュータコードを含んだ、不揮発性のコンピュータにより読み取り可能な媒体を含んでもよい。

【0050】

単一のメモリ 408 が描かれているが、 計算機 405 には多数のメモリが有用に組み込まれてもよいと理解されるであろう。例えば、第一のメモリは、 計算機 405 の電源が切られた時もファイル及びコードを永続的 又は長期的に記憶するディスクドライブのような不揮発性記憶装置を提供しても良い。ランダムアクセスメモリのような第二のメモリは、実行プロセスのための命令及びデータを格納するための揮発性 (しかしより高速な) メモリを提供してもよい。第三のメモリは、レジスタ、キャッシュ等の用途でプロセッサ 406 に物理的に隣接した高速メモリを提供することにより、性能を向上させるために使用されてもよい。

40

【0051】

ネットワークインターフェース 410 はネットワーク 402 を介して他のリソースと通信関係になるよう 計算機 405 を接続するハードウェア及び/又はソフトウェアを含んでもよい。これはインターネットを介してアクセス可能な遠隔リソースや、例えば物理接続

50

(例えばイーサネット(登録商標))、無線周波数通信(例えばWi-Fi)、光通信(例えば光ファイバー、赤外線又は同種のもの)、超音波通信、又は、それら又は計算機405及び他の装置間のデータ運送に使用され得る他のメディアの組み合わせを使用して近距離通信プロトコルを使用して利用可能なローカルリソースを含んでもよい。ネットワークインターフェース410は、例えば、ルータ、モデム、ネットワークカード、赤外線送受信機、無線周波数(RF)送受信機、近接通信インターフェース、無線周波数識別(RFID)タグリーダー、又は任意の他のデータ読み取り又は書き込みリソース又は同種のもの、を含んでもよい。

【0052】

より一般的には、ネットワークインターフェース410は、計算機405の構成要素を他の計算又は通信リソースへ接続することに適した、ハードウェアとソフトウェアの組み合わせを含んでもよい。これは、非限定例として、IEEE 802.11規格(又はその変種)に準拠して動作する有線又は無線イーサネット(登録商標)接続のための電子装置又は他の近距離又は遠距離ワイヤレスネットワークングコンポーネント又は同種のものを含んでもよい。これは他のローカルデバイスへの接続、又はインターネットのようなデータネットワーク402へ代わりに接続される、ローカルエリアネットワーク又は同種のものへの接続に使用され得るBluetooth(登録商標)又は赤外線送受信機のような近距離通信用のハードウェアを含んでもよい。これはWiMAX接続又はセルラーネットワーク接続(例えばCDMA、GSM(登録商標)、LTE、又は他の適合したプロトコル又はプロトコルの組み合わせを使用した)のためのハードウェア/ソフトウェアを更に含んでもよい。ネットワークインターフェース410は入出力インターフェース414の一部として含まれていてもよく、又はその逆でもよい。

【0053】

データ記憶部412は、ディスクドライブ、光学ドライブ、磁気ドライブ、フラッシュドライブ、又は、計算機405に大容量ストレージを提供可能である他の装置等の、コンピュータにより読み取り可能な媒体を提供する内蔵メモリ記憶装置であってもよい。データ記憶部412は、計算機405又はシステム400のための、コンピュータにより読み取り可能な命令、データ構造、プログラムモジュール、及びその他のデータを、不揮発性の形態で、比較的長期で永続的に記憶しその後引き出して使用するために格納してもよい。例えば、データ記憶部412はオペレーティングシステム、アプリケーションプログラム、プログラムデータ、データベース、ファイル、及び他のモジュール又は他のソフトウェアオブジェクト及び同種のことを格納してもよい。

【0054】

入出力インターフェース414は計算機405に接続され得る他の装置への入力と出力を支援してもよい。これは、例えば、シリアルポート(例えばRS-232ポート)、ユニバーサルシリアルバス(USB)ポート、光学ポート、イーサネット(登録商標)ポート、電話ポート、オーディオジャック、コンポーネントオーディオ/ビデオ入力、HDMI(登録商標)ポート等、を含み、どれもが他のローカルデバイスへ有線接続する際に使用されてもよい。他のローカルデバイスと通信関係になるよう無線接続する赤外線インターフェース、RFインターフェース、磁気カードリーダー、又は他の入出力システムも含んでもよい。ネットワーク通信のためのネットワークインターフェース410はローカルデバイス通信のための入出力インターフェース414と別個に説明されたが、これら二つのインターフェースは、同じであってもよく、又は、USBポートがWi-Fiアクセサリへの接続に使用されたり、イーサネット(登録商標)接続がローカルネットワークに接続された記憶装置へ使用されたりというように、機能を共有していてもよいことが理解されよう。

【0055】

周辺機器(peripheral)416は計算機405へ情報を提供するために使用される装置、又は計算機405から情報を受信する装置を含んでもよい。これは計算機405への入力を提供するために使用され得るリモートコントローラ、キーボード、マウス、マウスパッ

10

20

30

40

50

ド、トラックボール、ジョイスティック、マイク、フットペダル、タッチスクリーン、スキャナ、又はその他の装置のような、人間による入出力（I/O）装置を含んでもよい。これは、情報を提示するための、ディスプレイ、スピーカ、プリンタ、プロジェクタ、ヘッドセット、又はその他の任意の視聴覚装置を含んでもよい。周辺機器 4 1 6 は、他の装置又は構成要素の制御又は通信を支援する、デジタル信号処理装置、アクチュエーター、又はその他の装置を含んでもよい。

【 0 0 5 6 】

周辺機器 4 1 6 として使用するのに適した他の入出力装置は、触覚装置、三次元レンダリングシステム、拡張現実ディスプレイ等を含む。ある態様では、周辺機器 4 1 6 は短距離（例えば Bluetooth（登録商標）、Wi-Fi、赤外線、RF 又は同種のもの）又は長距離（携帯電話データ又は WiMAX）通信プロトコルを介しての通信を提供するように構成された USB 装置等のネットワークインターフェース 4 1 0 として機能してもよい。別の態様では、周辺機器 4 1 6 は、地球全即位システム（GPS）装置、セキュリティドングル、又は他の装置等の、追加の機能又は機能により、計算機 4 0 5 の動作を増強してもよい。別の態様では、周辺機器 4 1 6 は、フラッシュカード、USB ドライブ、又は他のソリッドステート装置、又は光学ドライブ、磁気ドライブ、ディスクドライブ、又は他の大容量記憶に適した装置又は装置の組み合わせ等の、記憶装置を含んでもよい。より一般的には、計算機 4 0 5 と共に使用されるのに適した装置又は装置の組み合わせが、本明細書にて企図されるように、周辺機器 4 1 6 として使用されてもよい。

【 0 0 5 7 】

制御部 4 1 8 は、計算機 4 0 5 の動作又は動作制御の制御の役割を務めてもよい。例えば、計算機 4 0 5 は移動ロボットとして実装されていてもよく、その中で制御部 4 1 8 は計算機 4 0 5 の動き及び動作の機能の全てをモニタし制御することを担当してもよい。制御部 4 1 8 は、計算機の稼働構成要素の動きや動作に関連するモータ、車輪、ギア、アーム又は関節付きの手足又は指の制御を任務としてもよい。制御部は入出力インターフェース 4 1 4 及び周辺機器 4 1 6 に連結されてもよく、ここで周辺機器 4 1 6 は計算機の動作を指示することが可能なりモートコントローラである。

【 0 0 5 8 】

画像キャプチャ装置 4 2 0 は、計算機 4 0 5 の環境に関連する画像データをキャプチャし記録する能力と共に使用する、カメラ、センサ、又は他の画像データ収集装置を含んでもよい。例えば、画像キャプチャ装置 4 2 0 は、非制約環境の 2 D 画像をキャプチャすることに適した 3 6 0 度パノラマカメラ 4 2 2 を含んでもよい。画像キャプチャ装置 4 2 0 は更にステレオカメラ、構造化光式カメラ、飛行時間型カメラ又は同種のもののような奥行センシング装置 4 2 4 を含んでもよく、それらは計算機 4 0 5 に環境の奥行マップを生成する能力を与え、2 D パノラマ画像と共に、環境内の個人、ジェスチャ、及び関心対象エリアを詳しく描く顕著性マップを生成する。画像キャプチャ装置 4 2 0 は更に、光学観測装置又は同種のもののような、追加の視線検出装置を含んでもよい。あるいは、奥行センシング装置 4 2 4 は、前述のように例えば IR カメラ及びセンサを使用して視線を検出可能であってもよい。

【 0 0 5 9 】

知識ベース 4 2 6 は前述のように計算機 4 0 5 の学習の中心として機能してもよい。知識ベース 4 2 6 は人工知能システム、ディープニューラルネットワーク、セマンティックシーンパーシングシステム、自己位置推定とマッピングの同時実行システム、コンピュータビジョンシステム又は同種のものゝ形であってもよい。知識ベースは、未知の物体の識別と関連付けに関する追加の情報のために、ネットワーク 4 0 2 を介して外部のデータソースと通信するために、入出力インターフェース 4 1 4 に又はネットワークインターフェース 4 1 0 に連結されてもよい。知識ベース 4 2 6 はシステムの近くに位置する関心対象物を、環境の画像データを使用して、識別及び関連付けを行う基礎として機能してもよい。知識ベース 4 2 6 は、既知及び未知の識別された物体のデータ及びそれらの物体に関連する属性を格納し呼び出すために、データ記憶部 4 1 2 又はメモリ 4 0 8 に連結されても

10

20

30

40

50

よい。

【 0 0 6 0 】

コプロセッサ、デジタル信号処理システム、数値演算コプロセッサ、グラフィックスエンジン、ビデオドライバ、マイクロフォン、又はスピーカのような、その他のハードウェア 4 2 8 が計算機 4 0 5 に組み込まれてもよい。上記その他のハードウェア 4 2 8 は拡張入出力ポート、追加メモリ、又は追加ドライブ（例えばディスクドライブ又は他のアクセサリ）を含んでもよい。

【 0 0 6 1 】

バス 4 3 0 又はバスの組み合わせは、プロセッサ 4 0 6、メモリ 4 0 8、ネットワークインターフェース 4 1 0、データ記憶部 4 1 2、入出力インターフェース 4 1 4、制御部 4 1 8、画像キャプチャ装置 4 2 0、知識ベース 4 2 6、及びその他のハードウェア 4 2 8 等の、計算機 4 0 5 の構成要素を相互に接続する電氣的又は電気機械的バックボーンとして機能してもよい。図に示されるように、計算機 4 0 5 の各構成要素は、制御、コマンド、データ、又は電源を共有する通信関係となるように、システムバス 4 3 0 を使用して相互に接続されてもよい。

10

【 0 0 6 2 】

図 5 は、本開示の実施形態による、人間の視線とジェスチャの検出を使用して物体を位置特定し識別する方法 5 0 0 を示すフローチャートである。前述のように、本明細書にて説明するシステムは、個人が興味を持っている物体を位置特定し識別するために、環境画像データ、人間の視線情報、及びジェスチャ情報を使用してもよい。ブロック 5 0 5 に示すように、システムは物体を識別する要求を受け取ってもよい。例えば、個人は、システムと通信しているリモートコントローラ又は他のインターフェース上で、音声コマンド又は触覚コマンドを使用して、位置特定と検出の要求を開始してもよい。

20

【 0 0 6 3 】

システムは、音声又はそれ以外の要求を受け取ると、自然言語処理、学習済みの行動又は同種のものを使用して要求を解析し、位置特定と識別処理を更に精緻化してもよい。前述のように、システムは、顕著性マップ及び物体の位置特定と検出を精緻化する使用可能な、コマンド、オブジェクト、及びアフォーダンスのライブラリを訓練又はプログラムされていてもよい。要求の言語はシステムが検索エリア又は環境を狭めることができる追加の状況の情報を、システムに提供してもよい。例えば、方向指示を行っている個人が「テーブルの上の物体を位置特定せよ」という要求を行い、テーブル上の箱を注視したならば、システムは要求を解析し、システムが物体を探し得る、より狭められた環境としてテーブルを識別してもよい。システムは、テーブルが、物体が置かれ得る、高いところにある平らな平面を持つことを知っていてもよい。システムはテーブルの既知の属性を共有しない壁、床、及び他の物体を環境から排除することができる。既知及び学習されたコマンド、オブジェクト、及びアフォーダンスのライブラリは、システムが位置特定と識別のプロセスを開始できる基礎的な知識ベースを提供してもよい。

30

【 0 0 6 4 】

ブロック 5 1 0 に示すように、システムは非制約環境に関する画像と奥行情報をキャプチャしてもよい。システムは非制約環境の 2 D 画像をキャプチャしてもよい。2 D 環境画像は 3 6 0 度パノラマカメラによりキャプチャされてもよく、又は一つ以上のカメラにより撮影された一連の画像をつなぎ合わせて環境の 3 6 0 度の表現を形成してもよい。システムは更に奥行画像をキャプチャしてもよい。奥行画像は I R カメラ、センサ又は同種のものを使用して取得してもよい。環境の奥行情報はまた S L A M（例えば移動ロボットシステム）、又は屋内のシーン等の非制約環境の奥行推定を訓練したディープニューラルネットワークを使用して取得してもよい。奥行情報は、システム、個人、物体、及び/又は他の定義可能な環境の輪郭をシステムが検出したら、それらの間の距離を測定するために使用されてもよい。

40

【 0 0 6 5 】

ブロック 5 1 5 に示すように、システムは環境マップを構築、又は引き出してもよい。

50

環境マップはシステムが取得した2Dパノラマ画像及び奥行センシングデータに基づいていてもよい。あるいは、環境マップは、環境が以前マッピングされた、又はシステムにアップロードされたなら、システムのメモリから引き出されてもよい。環境マップは、システムが解釈するためにより意味があり簡単な環境の表現として、画像及び奥行データを詳しく示す顕著性マップとして生成されてもよい。あるいは、環境マップは、シーンのセマンティックパーシングにより、物体とそれらのアイデンティティ（既知であれば）として生成されてもよい。システムは、区別される画像データ（例えば、普通の均質なピクセルエリアと比較して、区別されるピクセル又は画像データのエリアを識別する）に基づいて、関心対象エリアを定義してもよい。

【0066】

ブロック520に示すように、一度環境マップが構築又は引き出されれば、システムは、物体とは対照的な個人、及び個人の視線の位置特及び識別を行ってもよい。識別された個人の頭と目の周囲にバウンディングボックスが生成されてもよい。バウンディングボックス内では、個人の目の動きを位置特定し追従する目的で推定視線ベクトルを形成するために、IRセンサ又は光学/頭部追跡装置を使用して、個人の視線が検出されてもよい。

【0067】

システムは、ブロック525に示すように、個人による視線を検出してよい。環境マップを解析し、及び個人を識別した結果、方向指示を行っている個人の視線の方向が検出されてもよい。動きと個人の目の視線方向を判断するために、RGBカメラ、IRセンサ又は同種のものを使用して視線を検出してよい。

【0068】

ブロック530に示すように、システムは次に方向性ベクトルを生成してもよい。個人視線とジェスチャの解析により、システムは、環境のシーン上に、個人から環境マップ上に渡る個人による視線とジェスチャの推定方向に、方向性ベクトルを投影することができるようになってもよい。方向性ベクトルは、個人による環境中の視線及びジェスチャから、外挿法により推定されてもよい。

【0069】

ブロック535に示すように、システムは個人による視線とジェスチャに基づいて注目対象物を検索し位置特定してもよい。システムは、ベクトルが顕著性マップ中に識別された任意の物体又は関心対象エリアと交差するかを判断するために、外挿法により推定された方向性ベクトルを解析してもよい。ベクトルが顕著性マップ中の関心対象エリアと交差するならば、システムは関心対象エリアを個人の注目対象物であると判断してもよい。

【0070】

ブロック540に示すように、システムは物体が既知であるか判断してもよい。関心対象物が位置特定されたなら、システムはニューラルネットワーク、画像データベース、又は他のデータソースのような知識ベースに問い合わせ、物体の画像データと同様に関連付けられた画像データを有する既知の物体と比較してもよい。

【0071】

物体が既知であるならば、ブロック550に示されるように、システムは個人からの更なる指示を求めてもよい。例えば、システムは、物体が位置特定され識別されたこと、及び物体のアイデンティティを示す、音声による又は視覚的な指示を出してもよい。システムは個人へ、物体に関する、又は他の物体の位置特定と識別に関する、さらなる指示を要求してもよい。ブロック545に示すように物体がシステムにとって未知である場合は、システムは、個人と相互作用を行って知識ベースに物体の画像データ及び関連するアイデンティティを記録することでシステムが訓練されるような、学習プロセスを開始してもよい。システムが次回類似した画像データを持つ物体を識別するよう求められた時に、画像データ及び関連するアイデンティティが呼び出されてもよい。

【0072】

図6には、本開示の実施形態により視線判定を精緻化するためのさらなるフローチャートが示されている。ブロック605に示すように、また前述のように、システムは方向指

10

20

30

40

50

示を行っている個人から、関心対象物を位置特定し識別するよう、要求を受け取ってもよい。要求は音声によるものでもよく、又は、システム上又は遠隔に位置するユーザインターフェイス又は他のコントローラにより開始されてもよい。システムは、プログラムにより、又は高度な機械学習技術により、関心対象物の探索と位置特定を開始するための、複雑で状況に基づいた要求を解析してもよい。要求は、関心対象物が識別された後に実行する追加の指示を更に含んでもよい。システムは、追加の指示を解析し実行するために、プログラム又は高度な知識ベースに再び依存してもよい。

【 0 0 7 3 】

ブロック 6 1 0 に示すように、また前述のように、システムは非制約環境に関するセンサ情報をキャプチャしてもよい。環境をキャプチャするために、センサデータを RGB カメラ、IR センサ又は同種のものにより取得してもよい。ブロック 6 1 5 に示すように、システムは環境マップを構築又は引き出すために、キャプチャしたセンサ情報を使用してもよい。システムは、システムがメモリに格納し得ている他の環境マップと、新規に取得したセンサデータの一部又は全部を比較してもよい。環境がシステムにとって既知である場合は、以前格納した環境マップが引き出されてもよい。しかしながら環境がシステムにとって未知である場合は、システムは、その後の動作で使用するために、環境マップをメモリに登録してもよい。

【 0 0 7 4 】

ブロック 6 2 0 に示すように、また本明細書中に説明するように、システムはキャプチャしたセンサデータに基づいて環境中の個人も位置特定し識別してもよい。環境内の個人の存在を位置特定し識別するためにパターン認識、顔認識又は同種のものが実装されてもよい。環境中に個人が一人以上存在すれば、システムは、方向指示を行っている個人がどの個人であるのかを判断する追加の処理を実装してもよい。方向指示を行っている個人の注目対象物を位置特定し、その個人がどこを見つめているかを特定するために、システムは個人に関するデータを解析して、個人のうち一人がシステムを直接見ているか、ジェスチャを行っているか、又はシステムが位置特定のためにプログラムされ得る特定の特徴を持っているかを判断してもよい。例えば、方向指示を行っている個人が、システムが識別できるリモート又は他のコントローラを持っていてもよく、このことが、持っている人が方向指示を行っている個人であることを示してもよい。システムが方向指示を行っている個人を識別できない時は、システムは、更なる指示を求めるプロンプトを発行してもよい。

【 0 0 7 5 】

ブロック 6 2 5 に示すように、システムは方向指示を行っている個人の活動、視線及び/又はジェスチャを推定してもよい。前述のように、システムは、目の動きをスキャンし、方向指示を行っている個人がどこを見得ているのかを処理することにより、方向指示を行っている個人の視線を検出してもよい。視線の検出及び追跡は、その他の観測可能な情報により補完され、システムに解析を継続するための追加の状況の情報を与えてもよい。例えば、システムは、個人が特定の活動、姿勢、又は位置に関与しているかどうかを判断するために観測を行い、システムが注目対象物を探索している環境を精緻化してもよい。加えて、前述のように、システムは、方向指示を行っている個人のジェスチャを検出することで追加の情報を提供してもよい。これらの観測により、システムは、関心対象物を位置特定するために、より狭い興味範囲へと環境情報をフィルタするためにシステムが使用するパラメータのセットを形成してもよい。

【 0 0 7 6 】

ブロック 6 3 0 に示すように、画像及び奥行情報を含む環境マップは、顕著性マップを生成するために使用されてもよい。顕著性マップはセマンティックシーンパーシング、又は他の画像処理技術を含む、前述の技術を使用して生成してもよい。システムが取得した追加情報は顕著性マップの生成を補助してもよい。ブロック 6 2 5 に関連する個人の観測は、顕著性マップを更に発展又は制限するために使用されてもよい。方向指示を行っている個人が環境を更に精緻化するように観測された場合は、システムはその制限されたエリアのみに関心を持ってもよい。例えば、方向指示を行っている個人が環境の一部に向かっ

10

20

30

40

50

でジェスチャを行っている場合、システムは顕著性マップを、環境のそのエリアへと限定してもよい。

【0077】

加えて、システムは、環境を更に制限するために、状況の情報及びブロック605にて元の要求から抽出されたアフォーダンスを使用可能であってもよい。例えば、要求が角のような環境のあるエリアの識別を含んでいた場合、又は「窓を開ける」という文脈上の命令を含んでいた場合、システムは、環境及び顕著性マップを、壁の上の物体のみへと制限してもよい。このようにシステムは、より限定された、可能性のある物体による領域又は関心対象エリアを持つ、焦点の顕著性マップを生成することができる。

【0078】

ステップ635に示すように、システムは、観測された方向指示を行っている個人の活動と視線に基づいて、注目のヒートマップを生成してもよい。注目のヒートマップは、システムにより、ある期間における、方向指示を行っている個人の多数のスキャンを使用して生成されてもよい。前述のように、目の自然な動き及び要求を行う時に方向指示を行っている個人が複数のエリアを見つめるという性質から、システムは多数のスキャンを行い、方向指示を行っている個人の視線の位置の頻度と密度を詳しく説明するヒートマップを生成してもよい。方向指示を行っている個人による視線が最も集中しているエリアは、方向指示を行っている個人が注目しているエリア又は方向を意味してもよい。いくつかの離散的な視線方向が検出されるかもしれないが、いくつかの視線スキャンの融合により、注目対象物が位置するエリアへ向けられた有意に多数のスキャンが提供され得る。

【0079】

ブロック640に示すように、注目のヒートマップは、関心対象物がどこにあるかを判断するために、顕著性マップと組み合わせたり、重ね合わせたり、又は重複したりしてもよい。システムは、顕著性マップ上の関心対象エリアと検出された視線の集中領域の交差を判断することにより、関心対象物の位置を判断してもよい。システムがそのような交差を判断できるなら、システムは関心対象物を位置特定したかもしれず、その後、物体の識別へと移ってもよい。前述のように、システムが過程で得た任意の情報を使用して関心対象物が識別できる場合、システムはそのように示してもよい。物体がシステムにとって未知の場合は、学習プロセスを開始してもよい。

【0080】

本明細書にて説明される実施形態は人間の視線及びジェスチャを検出する移動ロボットを含むが、当業者はシステムがこのようなロボットのみ限定されず、分散型であろうと集中型であろうと、他の型の計算機にて本開示の主題から逸脱することなく実装されることを認めるであろう。

【0081】

更に、本開示において説明する例示的なシステム及び方法は、非制約環境の2D画像をキャプチャするために360度パノラマカメラを実装するが、当業者はその他の画像キャプチャ装置が使用されてもよいことを認めるであろう。例えば、システムのプロセッサがつながり合わせて環境の360度表現を形成してもよい一連の環境画像をキャプチャするために、RGBカメラ、回転カメラ、又は重複した視野を持つ複数のカメラが使用されてもよい。

【0082】

本開示の実施形態は、奥行マップ情報を生成するためにIRカメラ、センサ及び同種のものの使用を説明するが、当業者は、SLAM又は奥行推定を訓練したディープニューラルネットワークを限定せず含むその他の奥行センシング技術が、本開示の主題から逸脱することなく使用されてもよいことを認めるであろう。

【0083】

単数形の項目への言及は、別段の明示的な記載がない限り、又は本文から明らかでない限り、複数形の項目を含むものと理解されるべきであり、またその逆も同様である。文法的な接続詞は、別段の明示的な記載がない限り、又は文脈から明らかでない限り、節、文

10

20

30

40

50

、単語及び同種のもの任意のそして全ての離接及び接続を表現することを意図している。したがって、「又は」という用語は一般的に「及び/又は」等を表すと理解されるべきである。

【0084】

本明細書では、本明細書中に別段の記載がない限り、値の範囲への言及は、限定を意図せず、代わりにその範囲の任意の及び全ての値を個別に言及しており、そのような範囲内の各個別の値はそれが個別に本明細書で言及されているかのように本明細書に組み込まれている。数値又は方向を伴う「約」、「おおよそ」、「実質的に」又は同種ものは、意図する目的のために満足に動作させるよう行う逸脱を示すと当業者により評価されるものとして、解釈されるべきである。値の範囲及び/又は数値は例としてのみ本明細書中に提供されており、記載された実施形態の範囲の制限を構成するものではない。本明細書中に提供される、任意の、及び全ての例、又は例示的な言葉（「例えば」、「のような」又は同種のもの）は、単に実施形態をより明確にすることを意図しており、実施形態の範囲に制限を設けるものではない。本明細書のいかなる言語も、未請求の要素が実施形態の実現に不可欠であることを示すと解釈されるべきではない。

10

【0085】

明細書及び以下の請求項にて、「第一の」、「第二の」、「第三の」、「前」、「後」及び同種の用語は、便宜上のものであり、特に明記されない限り、限定的な用語として解釈されるべきではない。

【0086】

20

前述の方法及びシステムは例示を意図するものであり限定的なものではないことが理解されるべきである。数多くの変形、追加、省略、及びその他の修正は当業者にとって明らかであろう。加えて、前述の説明中の方法のステップ及び図の順序は、特定の順序が明示的に要求されるか、又は文脈から明らかでない限り、言及されたステップを順序通りに実施することを要求するものではない。したがって、特定の実施形態を示し説明したが、本開示の精神及び範囲から逸脱することなく形態及び詳細の様々な変更及び修正が行われ得ることが当業者にとって明らかであり、これは、以下の請求項によって定義する、法律によって許容される最も広い意味で解釈されるべき本開示の一部を形成するよう意図される。本明細書に開示される発明は以下の態様を含む。

〔態様1〕

30

物体を位置特定し識別するためのシステムであって、
個人と関心対象物とを含む非制約環境の360度画像を生成するように構成された画像キャプチャ装置と、
非制約環境の奥行マップを生成するように構成された奥行センサと、
知識ベースと、
プロセッサであって、
前記360度画像及び前記奥行マップから環境マップを生成し、
前記360度画像中の前記個人からの注目方向を判断し、
前記注目方向を表す方向性ベクトルを生成し、
前記環境マップ上に前記個人からの前記方向性ベクトルを投影し、
前記方向性ベクトルと前記関心対象物との交差を検出し、
前記関心対象物に関するデータを使用して前記関心対象物のアイデンティティを前記知識ベースで検索するように構成されたプロセッサと、
を備える、システム。

40

〔態様2〕

前記画像キャプチャ装置はパノラマカメラである、態様1に記載のシステム。

〔態様3〕

前記プロセッサは更に、視線、前記個人のジェスチャ方向、又はそれらの組み合わせのうち少なくとも一つから、前記注目方向を判断するように構成された、態様1に記載のシステム。

50

〔態様 4〕

前記奥行センサは赤外線センサ、奥行推定を訓練したニューラルネットワーク、又は自己位置推定とマッピングの同時実行センサの一つである、態様 1 に記載のシステム。

〔態様 5〕

物体を位置特定し識別する方法であって、

個人と関心対象物を含む非制約環境の 360 度画像をキャプチャすることと、

環境マップを生成することと、

前記キャプチャした 360 度画像中の前記個人からの注目方向を判断することと、

前記注目方向を表す方向性ベクトルを生成することと、

前記環境マップ上へ前記個人からの前記方向性ベクトルを投影することと、

前記方向性ベクトルと前記関心対象物との交差を検出することと、

を含む、方法。

10

〔態様 6〕

前記関心対象物のアイデンティティを知識ベースで検索することと、

前記関心対象物の前記アイデンティティが前記知識ベース中に発見されない時に学習プロセスを開始することと、を更に含む、態様 5 に記載の方法。

〔態様 7〕

前記学習プロセスはニューラルネットワークを備える、態様 6 に記載の方法。

〔態様 8〕

前記 360 度画像は単眼パノラマカメラによりキャプチャされる、態様 5 に記載の方法。

20

〔態様 9〕

複数の画像をキャプチャすることを更に含み、前記 360 度画像は前記複数の画像をつなぎ合わせたものを含む、態様 5 に記載の方法。

〔態様 10〕

前記環境マップを生成することは

環境の奥行情報を取得することと、

前記環境の奥行情報を前記 360 度画像と組み合わせることと、

を含む、態様 5 に記載の方法。

〔態様 11〕

前記環境の奥行情報は赤外線センサにより取得される、態様 10 に記載の方法。

30

〔態様 12〕

前記環境の奥行情報は自己位置推定とマッピングの同時実行プロセスから取得される、態様 10 に記載の方法。

〔態様 13〕

前記環境の奥行情報はニューラルネットワークから取得される、態様 10 に記載の方法。

〔態様 14〕

前記環境マップは顕著性マップを含む、態様 5 に記載の方法。

〔態様 15〕

前記環境マップは前記環境のセマンティックパーシングを備える、態様 5 に記載の方法。

〔態様 16〕

前記個人の視線を推定することを更に含み、前記注目方向は前記視線から判断される、態様 5 に記載の方法。

40

〔態様 17〕

前記視線は赤外線センサを使用して推定される、態様 16 に記載の方法。

〔態様 18〕

前記個人のジェスチャ方向を推定することを更に含み、前記注目方向は前記ジェスチャ方向から判断される、態様 5 に記載の方法。

〔態様 19〕

前記個人の視線及びジェスチャ方向を推定することを更に含み、前記注目方向は前記視線及び前記ジェスチャ方向から判断される、態様 5 に記載の方法。

50

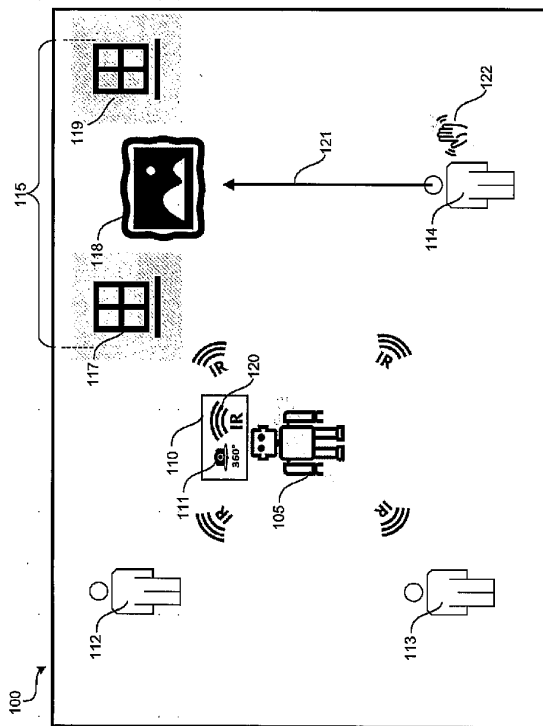
〔態様 20〕

物体を位置特定し識別するためのシステムであって、
 個人と関心対象物を含む非制約環境の360度画像をキャプチャするように構成された単
 眼カメラと、
 前記非制約環境の三次元奥行マップを生成するように構成された赤外線奥行センサと、
 知識ベースと、
 プロセッサであって、
 前記360度画像及び前記奥行マップから顕著性マップを生成し、
 前記個人からの注目方向を判断し、
 前記注目方向を表す三次元方向性ベクトルを生成し、
 前記顕著性マップ上に前記個人からの前記三次元方向性ベクトルを投影し、
 前記三次元方向性ベクトルと前記関心対象物との交差を検出し、
 前記関心対象物に関するキャプチャされた画像データを使用して、前記関心対象物のアイ
 デンティティを前記知識ベースで検索するように構成されたプロセッサと、
 を備える、システム。

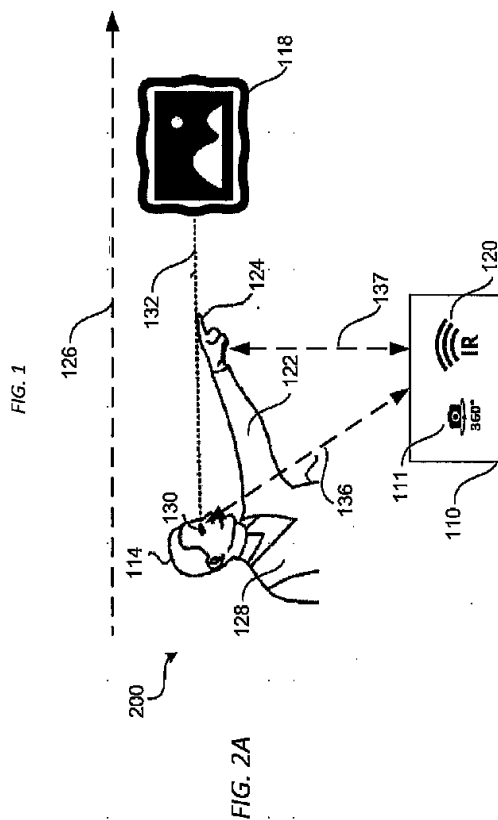
10

【図面】

【図 1】



【図 2 A】



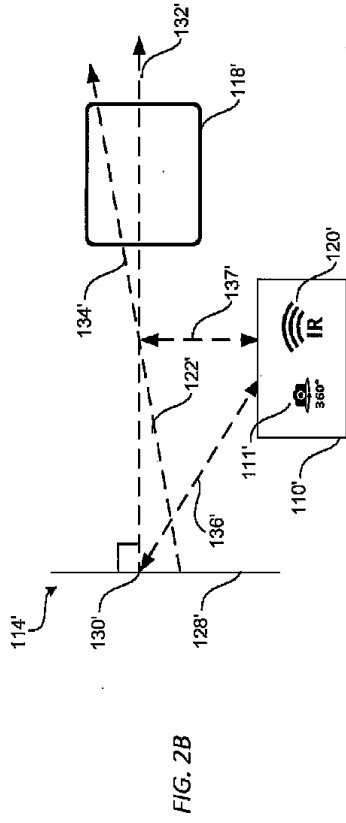
20

30

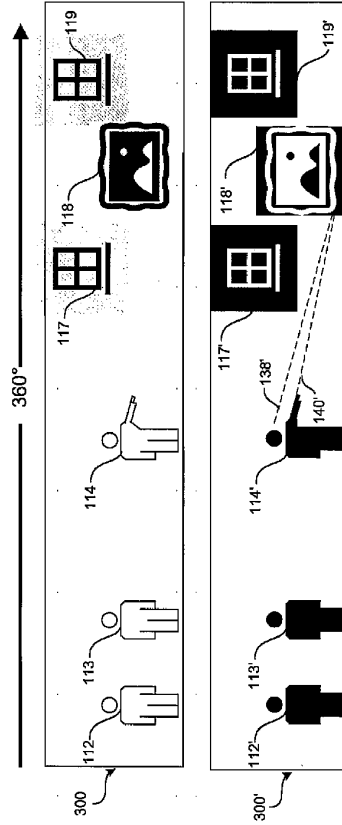
40

50

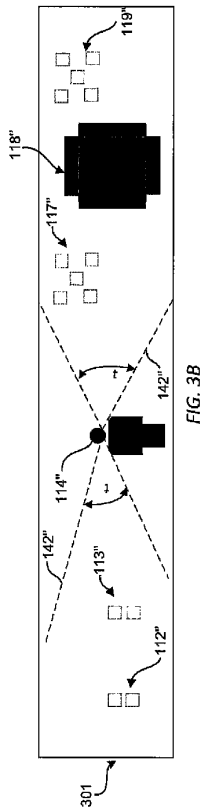
【図 2 B】



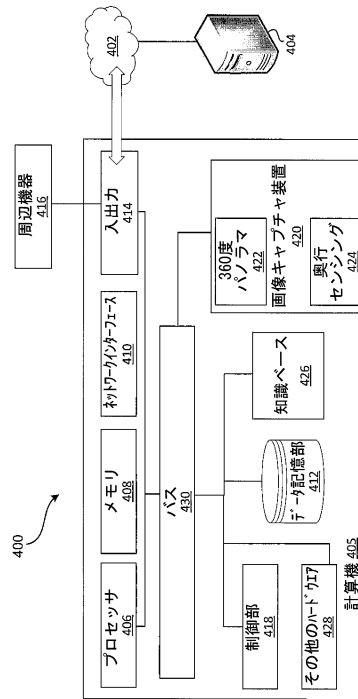
【図 3 A】



【図 3 B】



【図 4】



10

20

30

40

50

【 図 5 】

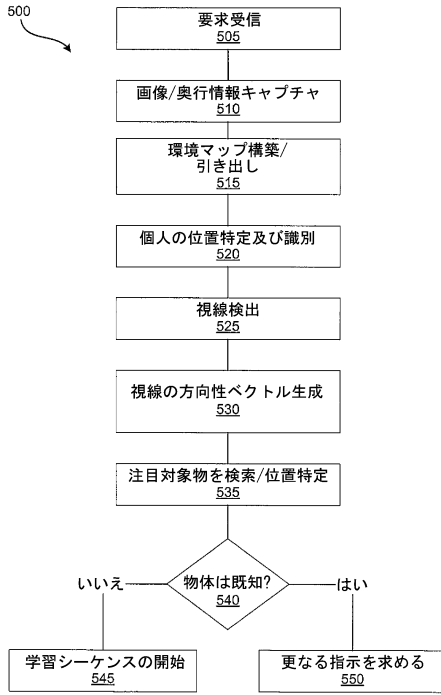


FIG. 5

【 図 6 】

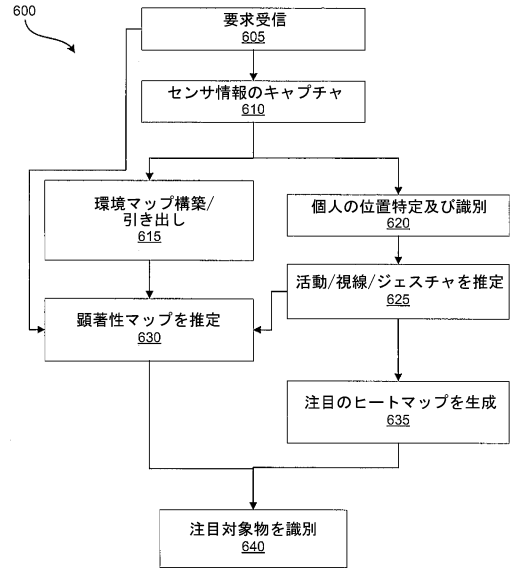


FIG. 6

10

20

30

40

50

フロントページの続き

(51)国際特許分類

F I
 G 0 6 F 3/0346 4 2 3
 G 0 6 F 3/038 3 1 0 A
 G 0 6 F 3/04815

弁理士 渡辺 陽一

(74)代理人 100108903

弁理士 中村 和広

(72)発明者 サイモン エー . アイ . ステント

アメリカ合衆国, カリフォルニア 9 4 0 2 2 , ロス アルトス, エル カミーノ レアル 4 4 4 0

合議体

審判長 廣川 浩

審判官 河合 弘明

審判官 中木 努

(56)参考文献

特開 2 0 1 7 - 0 4 1 2 4 2 (J P , A)

特開 2 0 0 0 - 0 5 7 3 4 9 (J P , A)

特開 2 0 1 0 - 2 6 8 1 5 8 (J P , A)

特開 2 0 1 5 - 1 1 6 3 1 9 (J P , A)

特開 2 0 1 7 - 2 2 4 1 8 4 (J P , A)

Hengcan Shi et al., "Hierarchical Parsing Net: Semantic Scene Parsing From Global Scene to Objects", IEEE Transactions on Multimedia, 米国, IEEE, 2018年03月05日, 発行日, Vol.20, No.10, pp.2670-2682

Simone Milani et al., "Joint denoising and interpolation of depth maps for MS Kinect sensors", 2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 米国, IEEE, 2012年03月25日, 発行日, pp.797-800

Sen Wang et al., "Towards Robotic Semantic Segmentation of Supporting Surfaces", 2015 IEEE International Conference on Computational Intelligence & Communication Technology, 米国, IEEE, 2015年02月13日, 発行日, pp.775-779

早川 俊介、外 3 名, " 共同注意実現を目指した視線及び頭部動作を認識するロボットの開発 ", 電子情報通信学会技術研究報告, 日本, 社団法人電子情報通信学会, 2003年11月14日, 発行日, Vol.103, No.455, pp.31-36

(58)調査した分野 (Int.Cl., D B 名)

G06T 1/00,7/00-7/90,19/00

G06V 10/00-10/98

G06F 3/01,3/038-3/04815

C S D B