

(19) 日本国特許庁(JP)

(12) 公開特許公報(A)

(11) 特許出願公開番号

特開2017-157201
(P2017-157201A)

(43) 公開日 平成29年9月7日(2017.9.7)

(51) Int.Cl. F I テーマコード (参考)
G06T 7/00 (2017.01) G06T 7/00 200Z 5L096

審査請求 有 請求項の数 18 O L (全 24 頁)

(21) 出願番号 特願2017-15619 (P2017-15619)
(22) 出願日 平成29年1月31日 (2017.1.31)
(31) 優先権主張番号 15/057,032
(32) 優先日 平成28年2月29日 (2016.2.29)
(33) 優先権主張国 米国 (US)

(特許庁注：以下のものは登録商標)

1. ETHERNET

(71) 出願人 000003207
トヨタ自動車株式会社
愛知県豊田市トヨタ町1番地
(74) 代理人 100100549
弁理士 川口 嘉之
(74) 代理人 100085006
弁理士 世良 和信
(74) 代理人 100113608
弁理士 平川 明
(74) 代理人 100123319
弁理士 関根 武彦
(74) 代理人 100123098
弁理士 今堀 克彦
(74) 代理人 100143797
弁理士 宮下 文徳

最終頁に続く

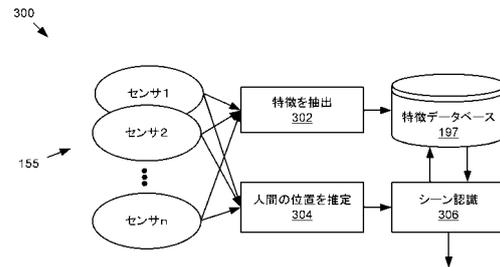
(54) 【発明の名称】 人間を中心とした場所認識方法

(57) 【要約】

【課題】対象者が存在する場所を認識する際の精度を向上させる。

【解決手段】コンピュータによって実行される方法であって、一つ以上のセンサを用いて、環境をセンシングしたセンサデータを取得する取得ステップと、前記センサデータに基づいて、前記環境内にいる対象者を検出する検出ステップと、前記対象者の位置を判定する判定ステップと、前記環境から事前に抽出された、特徴に関するデータが蓄積された特徴データベースに対して、前記対象者の位置を用いて問い合わせを行い、前記対象者の近傍にある特徴を抽出するクエリステップと、前記特徴が特定のシーンタイプに含まれる確率を表す関連付け尤度値に基づいて、複数のシーンタイプの中からシーンタイプを選択する選択ステップと、を含む。

【選択図】 図3



【特許請求の範囲】

【請求項 1】

コンピュータによって実行される方法であって、
一つ以上のセンサを用いて、環境をセンシングしたセンサデータを取得する取得ステップと、

前記センサデータに基づいて、前記環境内にいる対象者を検出する検出ステップと、
前記対象者の位置を判定する判定ステップと、

前記環境から事前に抽出された、特徴に関するデータが蓄積された特徴データベースに対して、前記対象者の位置を用いて問い合わせを行い、前記対象者の近傍にある特徴を抽出するクエリステップと、

前記特徴が特定のシーンタイプに含まれる確率を表す関連付け尤度値に基づいて、複数のシーンタイプの中からシーンタイプを選択する選択ステップと、

を含む、方法。

【請求項 2】

選択されたシーンタイプに基づいて、一つ以上の自律ルーチンを実行する実行ステップをさらに含む、

請求項 1 に記載の方法。

【請求項 3】

前記関連付け尤度値は、時刻に応じて異なる値を取りうる、

請求項 1 に記載の方法。

【請求項 4】

前記関連付け尤度値は、前記特徴が前記シーンタイプに属する確率、前記特徴を使用して正確にシーンタイプを分類する事前確率、任意の基準点に対する N 次元における位置およびサイズ、をそれぞれ記述するスコアである、

請求項 1 に記載の方法。

【請求項 5】

前記関連付け尤度値は、下記の数式を使用して計算される確率である、

$$P(ff | s) = \eta \prod_i P(s | ff_i) P(ff_i)$$

請求項 1 に記載の方法。

【請求項 6】

前記一つ以上のセンサを使用して前記環境を事前に走査し、前記一つ以上のセンサから得られたセンサデータを使用して、前記環境から特徴を抽出し、前記抽出した特徴を前記特徴データベースに記憶させる登録ステップをさらに含む、

請求項 1 に記載の方法。

【請求項 7】

前記選択ステップでは、

前記対象者の近傍のエリアについて特徴の偏りを検出し、前記偏りに基づいて判定したシーンの指向性にさらに基づいてシーンタイプを選択する、

請求項 1 に記載の方法。

【請求項 8】

前記一つ以上のセンサのうち少なくとも一つは、距離画像を取得するカメラである、

請求項 1 に記載の方法。

【請求項 9】

前記判定ステップでは、前記対象者の移動先を推定し、当該移動先を前記対象者の位置とする、

請求項 1 に記載の方法。

【請求項 10】

一つ以上のセンサを用いて、環境をセンシングしたセンサデータを取得する取得手段と

、

10

20

30

40

50

前記センサデータに基づいて、前記環境内にいる対象者を検出する検出手段と、
 前記対象者の位置を判定する判定手段と、
 前記環境から事前に抽出された、特徴に関するデータが蓄積された特徴データベースに
 対して、前記対象者の位置を用いて問い合わせを行い、前記対象者の近傍にある特徴を抽
 出するクエリ手段と、
 前記特徴が特定のシーンタイプに含まれる確率を表す関連付け尤度値に基づいて、複数
 のシーンタイプの中からシーンタイプを選択する選択手段と、
 を含む、システム。

【請求項 11】

選択されたシーンタイプに基づいて、一つ以上の自律ルーチンを実行する実行手段をさ
 らに含む、
 請求項 10 に記載のシステム。

10

【請求項 12】

前記関連付け尤度値は、時刻に応じて異なる値を取りうる、
 請求項 10 に記載のシステム。

【請求項 13】

前記関連付け尤度値は、前記特徴が前記シーンタイプに属する確率、前記特徴を使用し
 て正確にシーンタイプを分類する事前確率、任意の基準点に対する N 次元における位置お
 よびサイズ、をそれぞれ記述するスコアである、
 請求項 10 に記載のシステム。

20

【請求項 14】

前記関連付け尤度値は、下記の数式を使用して計算される確率である、

$$P(ff|s) = \eta \prod_i P(s|ff_i) P(ff_i)$$

請求項 10 に記載のシステム。

【請求項 15】

前記一つ以上のセンサを使用して前記環境を事前に走査し、前記一つ以上のセンサから
 得られたセンサデータを使用して、前記環境から特徴を抽出し、前記抽出した特徴を前記
 特徴データベースに記憶させる登録手段をさらに含む、
 請求項 10 に記載のシステム。

30

【請求項 16】

前記選択手段は、
 前記対象者の近傍のエリアについて特徴の偏りを検出し、前記偏りに基づいて判定した
 シーンの指向性にさらに基づいてシーンタイプを選択する、
 請求項 10 に記載のシステム。

【請求項 17】

前記一つ以上のセンサのうちの少なくとも一つは、距離画像を取得するカメラである、
 請求項 10 に記載のシステム。

【請求項 18】

前記判定手段は、前記対象者の移動先を推定し、当該移動先を前記対象者の位置とする
 、
 請求項 10 に記載のシステム。

40

【発明の詳細な説明】

【技術分野】

【0001】

本開示は、人間を中心とした場所認識方法に関する。

【背景技術】

【0002】

(関連出願への相互参照)

本出願は、"AUTONOMOUS HUMAN-CENTRIC PLACE RECOGNITION"と題し、2016年2月2

50

9日に出願された米国特許出願第15/057,032号の優先権を主張する。当該出願は、その全体が本明細書中に参照として組み込まれる。

【0003】

今日、多くの自律コンピュータシステムは、様々な異なるアプリケーション向けの検出技法および認識技法に依存している。場所および環境の分類において、複雑な聴覚、視覚、または多様な学習を使用して、システムは迅速に向上している。しかしながら、課題は、良いデータの分類ではなく、むしろ知識が必要な際の不十分なセンサのポジショニングを克服することにある。例えば、長椅子の上の人物と対話するロボットは、人物の後方にある大きい壁を捉える場合があるが、その壁は環境を正確に分類するための十分な複雑さを持っていない場合がある。あるいは、比較的視界が開けている環境であっても、いくつかの部屋は多目的であり、単純な分類ストラテジーが通用しない。さらに、ロボットがある部屋から隣の部屋へ移動するとき、ロボットはしばしば、その移動を識別し、正確な状況を判断することが困難である。ロボットなどの自律エージェントが分類結果に基づいて判断を行う必要があるとき、これらの「境界条件」は、モバイルセンサの配置に対する重大な障害になる。

10

【0004】

場所の認識またはラベル付けは新しい分野ではない。それは通常、シーン(シーン)認識や場所分類とも呼ばれる。現時点では、現在観察されているシーンのタイプを識別するために使用できる様々な手法およびセンサが存在する。非特許文献1に記載されている、環境内で物体のタイプを分類し、次いでそれらの物体に関連する意味のある場所ラベルを学ぶ既存の方法などが存在する。

20

【0005】

例えば、非特許文献2に記載されているような、現在オンラインで利用可能な、単一視点分類用の大規模画像データベースを利用する、深層学習における新しい研究である直線画像ベースの分類を行う既存の方法も存在する。

【0006】

しかしながら、上記の単一観察による分類方法では、カメラまたはセンサのポジショニングが十分でない場合において、自身で場所ラベルを識別することができない。例えば、二つ以上の環境を含んでいる場所をカテゴリ分けする際に、どちらか片方の環境のみを回答したり、あるいはどちらも回答しないとといった失敗をすることがある。

30

【0007】

センサのポジショニングにおけるこれらの問題を修正するために、ロボット工学コミュニティは、物理空間上でセンサデータを融合させることに注目した。議論された一つの可能性は、一般に入手可能な位置センサ、例えばGPSを使用して、画像の位置を既にラベル付けされたマップと比較することである。そして、場所を推定するために、GPSが予測した位置と分類された場所ラベルの組合せが使用される。「Scene Classification for Place Recognition」と題し、Boris Babenko, Hartwig Adam, John FlynnおよびHartmut Nevenによって、特許文献1において説明されているこの手法は、不十分にポジショニングされたセンサによる問題を改善する。しかし、当該発明は、都市または観光地のような大きい場所カテゴリ用に設計されている。また、当該発明は、小さい屋内環境、遷移領域、または多目的空間によるラベル付けの課題を解決しない。

40

【0008】

別の形態のセンサ融合は、環境のトポロジカルなマップを構築することである。非特許文献3に記載されている着想は、単一画像の手法ではなく、それらの前に来た画像と著しく異なる画像を探ることで、モバイルセンサからのビデオストリームを活用する。

環境をマッピングする際、ロボットは、同様の種類の領域をクラスタ化し、ある部屋と隣の部屋との間の変化点(change points)を自己識別する。得られたマップはトポロジカルグラフである。この方法は、不十分なセンサのポジショニングに対する改善を示すが、各部屋が同一の目的を有し、遷移(transitions)が明確に定義されていること(実際の環境では該当しないことがある)を仮定している。実際のマップを生成しないで分類ア

50

ルゴリズムを改善するためのさらなる同様の手法は、「Detecting and Labeling Places using Runtime Change-point Detection」と題し、Ananth Ranganathanによって特許文献2において説明されている。

【0009】

代替的な融合方法は、占有グリッドである。Ananth RanganathanおよびJongwoo Limは、非特許文献4において、場所認識アルゴリズムによる各測定結果を使用し、ロボットが空間内を移動するに従って占有グリッドを更新することを開示している。重要なことに、各測定結果の更新は、カメラによって観察された視界領域を反映し、占有グリッド内の障害物と空きスペースの両方についての分類を学習するよう試みる。トポロジカルマップと同様に、このセンサ融合戦略は、特にカメラからの基本的な指向性問題を克服する助けになるが、さらなる問題ももたらす。

第一に、マップは場所認識問題に直接回答しない。マップを仮定すると、ロボットはそのアプリケーション内で使用するための場所ラベルをどのように識別するのか？このグループは、いずれのアプリケーションにもマップを適用せず、そのため、人間-ロボット間対話、または任意の他の領域において得られた融合表現を最も良く利用する方法に対処しない。第二の問題は、このマップが点群データに焦点が当てられた静的表現であることである。マップ表現に対してリアルタイムで変更を行うか、または非点群データを組み込むことは、いずれも多目的環境における曖昧さを低減することに役立つ場合があるが、困難さを伴う。

【先行技術文献】

【特許文献】

【0010】

【特許文献1】米国特許第8798378号公報

【特許文献2】米国特許第8565538号公報

【非特許文献】

【0011】

【非特許文献1】Shrihari Vasudevan, Stefan Gachter, Marc Berger & Roland Siegwart, "Cognitive Maps for Mobile Robots - An Object based Approach", Intelligent Robots and Systems(IROS), サンディエゴ, 米国, 2007年

【非特許文献2】Bolei Zhou, Agata Lapedriza, Jianxiong Xiao, Antonio Torralba, Aude Oliva, "Learning Deep Features for Scene Recognition using Places Database", NIPS, 2014年

【非特許文献3】Aravindhan K Krishnan, K Madhava Krishna, "A Visual Exploration Algorithm using Semantic Cues that Constructs Image based Hybrid Maps", Intelligent Robots and Systems(IROS), 台北, 台湾, 2010年

【非特許文献4】Ananth Ranganathan, Jongwoo Lim, "Visual Place Categorization in Maps", Intelligent Robots and Systems(IROS), サンフランシスコ, 米国, 2011年

【発明の概要】

【0012】

人間-ロボット間対話などの自律的な実装形態向けの場所認識のユーザビリティを改善するため、本明細書は、エビデンスが蓄積されたデータベースに問い合わせるために対話対象者(例えば、人間)の検出位置を使用する、発明的な人間中心の手法を開示する。

人間-ロボット間対話の状況において、この手法は、人物が位置する場所の機能に部分的に基づいて、ロボットが必要とする関連情報を識別することができる。この態様は、モバイルロボットによる場所認識の失敗原因となる境界条件を安定させるために使用することができる。

【0013】

本開示に記載される主題の一つの発明的態様によれば、システムは、一つ以上のセンサを用いて、環境をセンシングしたセンサデータを取得する取得手段と、前記センサデータに基づいて、前記環境内にいる対象者を検出する検出手段と、前記対

10

20

30

40

50

象者の位置を判定する判定手段と、前記環境から事前に抽出された、特徴に関するデータが蓄積された特徴データベースに対して、前記対象者の位置を用いて問い合わせを行い、前記対象者の近傍にある特徴を抽出するクエリ手段と、前記特徴が特定のシーンタイプに含まれる確率を表す関連付け尤度値に基づいて、複数のシーンタイプの中からシーンタイプを選択する選択手段と、を含む。

【0014】

一般に、本開示に記載される主題の別の発明的態様は、

コンピュータによって実行される方法であって、一つ以上のセンサを用いて、環境をセンシングしたセンサデータを取得する取得ステップと、前記センサデータに基づいて、前記環境内にいる対象者を検出する検出ステップと、前記対象者の位置を判定する判定ステップと、前記環境から事前に抽出された、特徴に関するデータが蓄積された特徴データベースに対して、前記対象者の位置を用いて問い合わせを行い、前記対象者の近傍にある特徴を抽出するクエリステップと、前記特徴が特定のシーンタイプに含まれる確率を表す関連付け尤度値に基づいて、複数のシーンタイプの中からシーンタイプを選択する選択ステップと、を含む方法において具現化することができる。

10

【0015】

他の態様は、上記その他の発明的態様のための対応する方法、システム、装置、およびコンピュータプログラム製品を含む。

【0016】

上記その他の実装形態は、各々、場合によっては、以下の特徴および/または動作のうちの一つまたは複数を含む場合がある。

20

例えば、特徴および/または動作は、選択されたシーンタイプに基づいて、一つ以上の自律ルーチンを実行してもよい。

また、前記関連付け尤度値は、時刻に応じて異なる値を取り得てもよい。

また、前記関連付け尤度値は、前記特徴が前記シーンタイプに属する確率、前記特徴を使用して正確にシーンタイプを分類する事前確率、任意の基準点に対するN次元における位置およびサイズ、をそれぞれ記述するスコアであってもよい。

また、前記関連付け尤度値は、下記の数式を使用して計算されてもよい。

$$P(ff|s) = \eta \prod_i P(s|ff_i) P(ff_i)$$

30

また、前記一つ以上のセンサを使用して前記環境を事前に走査し、前記一つ以上のセンサから得られたセンサデータを使用して、前記環境から特徴を抽出し、前記抽出した特徴を前記特徴データベースに記憶させてもよい。

また、前記対象者の近傍のエリアについて特徴の偏りを検出し、前記偏りに基づいて判定したシーンの指向性にさらに基づいてシーンタイプを選択してもよい。

また、前記一つ以上のセンサのうち少なくとも一つは、距離画像を取得するカメラであってもよい。

また、前記対象者の移動先を推定し、当該移動先を前記対象者の位置としてもよい。

【0017】

本開示において提示される新規の検出技術は、いくつかの点で特に有利である。例えば、場所認識方程式 (place recognition equation) に人間の位置を組み込むことで、人間-ロボット間対話に対する回答の関連性が最大化される。具体的には、人間の位置と、検索用に設計されたデータベースを用いることで、人間-ロボット間対話のための複雑なシーンの曖昧さを除去し、指向性センサ、多目的ルーム (例えば、スタジオアパート、台所/ダイニング、寝室/オフィスなど)、および遷移領域を伴うシナリオにおいて、場所の認識性能を向上させることが可能になる。

40

【0018】

加えて、通常、物体の移動や、環境が変化する様々な空間を使用することなどに起因して時間とともに不正確になる、ある環境内の物体の固定位置をマッピングする固定マップのみを利用することとは対照的に、本検出技法は、ロボットなどの自動化デバイスが人間

50

とのより多くの情報に基づく計算的対話に使用することができる、環境の動的、文脈的、かつ時間的に関係する描写を生成する。

より具体的には、検索可能なデータベース内に場所認識のエビデンスを生成することで、動的環境をより良く反映することができる、より動的な分類アルゴリズムが可能になる。例えば、予想されるインタラクションの時間によって検索することで、人間が直ちに認識するであろう部屋の、時間変化する様子を明らかにすることができる。また、認識された物体の人間への近さが場所認識スコアに組み込まれ、対話中のロボットの動きおよび/または注意をガイドするための、局所的な場所の遷移を推定することができる。上記の利点は例として提供されること、および本技法は、多数の他の利点および利益を有することができることを理解されたい。

10

【0019】

本開示は、添付図面の図において限定としてではなく例として示され、添付図面では、同様の構成要素を参照するために同様の参照番号が使用される。

【図面の簡単な説明】

【0020】

【図1】本発明に係る例示的なシステムのブロック図である。

【0021】

【図2A】例示的なコンピューティングデバイスのブロック図である。

【0022】

【図2B】例示的な特徴モジュールのブロック図である。

20

【0023】

【図3】抽出された特徴と人間の位置に基づいてシーンを認識するためのプロセス例のフローチャートである。

【0024】

【図4】特徴データベースに記憶されるデータの例である。

【0025】

【図5】シーンを認識するための方法例のフローチャートである。

【0026】

【図6】特定の環境から特徴を抽出し、特徴データベースに投入するための方法例のフローチャートである。

30

【0027】

【図7】スコアを計算するための方法例のフローチャートである。

【0028】

【図8】シーンを認識する自律ロボットを含む家屋の例である。

【0029】

【図9】本発明に係る場所認識技法の利点を表す表である。

【発明を実施するための形態】

【0030】

図1は、人間中心の場所認識を行う例示的なシステム100のブロック図である。図示されたように、システム100は、計算サーバ101と、(ライン118で描写されたように)ユーザ125によってアクセスされ、当該ユーザと対話する検出システム103を含む。

40

実装形態に応じて、計算は、二つ以上のコンピューティングシステム(例えば、計算サーバ101および検出システム103、二つ以上の検出システム103)にわたって分散して行われてもよい。また、それらの間で動的にシフトされてもよく、検出システム103によって別個に実行されてもよい。そのため、システム100は、計算サーバ101を含まない場合もある。

【0031】

計算サーバ101が含まれる実施形態においては、検出システム103および計算サーバ101は、それぞれ信号線106および108を介して、ネットワーク105に通信可

50

能に結合する。例えば、検出システム 103 および計算サーバ 101 は、センサデータ、特徴データ、認識データなどのデータを交換するために、ネットワーク 105 を介して互いに通信可能に結合する。図 1 の信号線 106 および 108 は、一つまたは複数の有線接続またはワイヤレス接続であり得る。さらなる例として、検出システム 103 は、処理用に計算サーバ 101 にセンサデータを送信することができ、計算サーバ 101 は、本明細書に記載されるようにデータを処理して物体を検出および認識し、認識された物体を記述するデータや結果を検出システム 103 に送ることができる。

計算サーバ 101 が含まれない実施形態では、検出システム 103 は、自律的に、または他の検出システム 103 (不図示) と連携して動作し、物体、シーン、人間の位置などを検出および認識することができる。例えば、検出システム 103 は、本明細書において説明される計算を実行するために、他の同様の検出システム 103 とコンピュータネットワークを介してネットワーク接続されてもよい。

10

【0032】

なお、図 1 は、単一の検出システム 103 および計算サーバ 101 を描写するが、様々な異なるシステム環境およびシステム構成が可能であり、考察され、本開示の範囲内であることを理解されたい。例えば、いくつかの実施形態は、追加のまたはより少ないコンピューティングデバイス、サービス、ネットワークを含む場合があり、他のコンピューティングデバイス上にローカルまたはリモートに様々な機能を実装する場合がある。さらに、様々なエンティティが、単一のコンピューティングデバイスもしくはシステムに統合されるか、または追加のコンピューティングデバイスもしくはシステムなどにわたって分散される場合がある。例えば、検出モジュール 135 は、コンピューティングデバイスやシステムの組合せにわたって、または一つのコンピューティングデバイスやシステムの中に記憶され、それらによって実行可能であり、分散される場合がある。

20

【0033】

ネットワーク 105 は、標準タイプのネットワーク、有線、またはワイヤレスであり得るし、スター形、トークンリング、または他の既知の構成などの、任意の数の構成を有する場合がある。ネットワーク 105 は、一つもしくは複数のローカルエリアネットワーク (LAN)、ワイドエリアネットワーク (WAN) (例えばインターネット)、仮想プライベートネットワーク (VPN)、ピアツーピアネットワーク、近距離無線通信 (例えば Bluetooth (登録商標))、セルラーネットワーク (例えば、3G、4G、他の世代)、または、それを介して複数のコンピューティングノードが通信することができる任意の他の相互接続されたデータパスを含む場合がある。データは、例えば、様々なインターネットレイヤ、トランスポートレイヤ、またはアプリケーションレイヤのプロトコルを含む、様々な異なる通信プロトコルを使用して、ネットワーク 105 のノード間で、暗号化または非暗号化された形態で送信される場合がある。例えば、データは、伝送制御プロトコル/インターネットプロトコル (TCP/IP)、ユーザデータグラムプロトコル (UDP)、伝送制御プロトコル (TCP)、ハイパーテキスト転送プロトコル (HTTP)、セキュアハイパーテキスト転送プロトコル (HTTPS)、動的適応ストリーミングオーバー HTTP (DASH)、リアルタイムストリーミングプロトコル (RTSP)、リアルタイム転送プロトコル (RTP) およびリアルタイム転送制御プロトコル (RTCP)、ボイスオーバーインターネットプロトコル (VOIP)、ファイル転送プロトコル (FTP)、ウェブソケット (WS)、ワイヤレスアクセスプロトコル (WAP)、様々なメッセージングプロトコル (SMS、MMS、XMS、IMAP、SMTP、POP、WebDAV など)、または他の既知のプロトコルを使用して、ネットワークを介して送信される場合がある。

30

40

【0034】

検出システム 103 は、環境内にある物体の有意性を知覚、認識、および解釈して、アクションを実行することが可能な自律コンピューティングシステムであるか、またはそれに含まれる場合がある。例えば、検出システム 103 は、車に乗っているドライバーまたは同乗者を認識する能力を有するインテリジェントカーであるか、またはその中に組み込

50

まれる場合がある。さらなる例では、検出システム 103 は、人間や他のロボットと連携して、様々なタスクを実行することができるソーシャルロボット、または人口密集環境内で動作する自律システムであるか、またはその中に組み込まれる場合がある。

いくつかの実施形態では、検出システム 103 は、物体を検出および認識するための構成要素として他のシステムの中に組み込まれる場合がある。例えば、検出システム 103 は、ゲーミングシステム、テレビジョン、スマートフォン、タブレット、ラップトップ、ワークステーション、サーバなどのクライアントデバイスの中に組み込まれる場合がある。例えば、検出システム 103 は、ある特定の一人または複数の人物が特定の位置に存在するかどうかを判定するためにマシンまたはコンピュータシステムの中に組み込まれる場合があり、マシンまたはコンピュータシステムは、その特定の一人または複数の人物が特定の位置に存在する場合、特定のプログラムをオン/オフまたは実行することができる。

10

【0035】

いくつかの実施形態では、検出システム 103 は、一つもしくは複数のセンサ 155、(一つもしくは複数のプロセッサを表す場合がある)プロセッサ 195 および検出モジュールのインスタンス 135 を含む計算ユニット 115、特徴データベース 197、インターフェース 175 を含む場合がある。描写されたように、センサ 155 は、信号線 122 を介して計算ユニット 115 に通信可能に結合される。特徴データベース 197 は、信号線 124 を介して計算ユニット 115 に通信可能に結合される。インターフェース 175 は、信号線 126 を介して計算ユニット 115 に通信可能に結合される。いくつかの実施形態では、検出モジュールのインスタンス 135、またはその様々な構成要素は、本明細書内の他の場所に記載されるように、計算サーバ 101 に記憶され、それによって実行可能であり得る。検出モジュールのインスタンス 135 はまた、本明細書では個別に呼ばれるか、または総称して検出モジュール 135 と呼ばれる。

20

【0036】

計算ユニット 115、センサ 155、特徴データベース 197、およびインターフェース 175 の各々が、単数形または複数形で本明細書において描写および参照されているが、検出システム 103 は、任意の数の計算ユニット 115、センサ 155、特徴データベース 197、インターフェース 175 を含むことができることを認識されたい。さらに、構成に応じて、検出システム 103 は、オペレーティングシステム、プログラム、様々な追加センサ、モーター、ムーブメントアセンブリ、スピーカ、ディスプレイデバイスなどの入力/出力デバイス、ネットワーク 105 に通信可能に結合された他のデバイス(例えば、計算サーバ 101、他の検出システム 103 (不図示)、任意の他の適切なシステム(不図示))とのワイヤレス通信のトランシーバユニットおよびアンテナなどの、図 1 に示されていない他の構成要素を含む場合があることを理解されたい。

30

【0037】

センサ 155 は、周辺環境から光および他の信号を取り込み、深度データなどのセンサデータを生成および処理するように構成された、一つまたは複数のセンサを含む。例えば、センサ 155 は、RGB-D カメラ、立体カメラ、構造光カメラ/スキャナ、飛行時間カメラ、干渉計、変調イメージャ、レーザー測距器、明視野カメラ、強化 CCD カメラなどのレンジカメラを含む場合があるが、これに限定されない。また、超音波センサ、カラーカメラ、赤外線カメラなどの他のタイプのセンサが使用される場合があることを理解されたい。いくつかの実施形態では、センサ 155 や検出システム 103 は、加速度計、ジャイロスコープ、温度計、気圧計、熱電対、マイクロフォン、または他の従来検知デバイスなどの、様々なタイプのセンサの組合せを含む場合がある。MESA Imaging による Swiss Ranger センサ、Microsoft による Kinect (登録商標) センサ、様々な立体視システムなどは、センサ 155 が含む場合があるカメラのさらなる非限定的な例である。センサ 155 は計算ユニット 115 の中に組み込まれる場合があるか、またはワイヤレス接続もしくは有線接続を介して計算ユニット 115 に結合される異種デバイスであり得る。

40

【0038】

50

様々な実施形態では、センサ 155 は、本明細書内の他の場所に記載されるように、センサデータ（例えば、センサ 155 によって取り込まれた物体に関連する距離情報を記述する深度データ）を生成し、処理用に計算ユニット 115 や計算サーバ 101 に送ることができる。

【0039】

計算ユニット 115 は、図 2 A に描写されるコンピューティングデバイス 200 などの、任意のプロセッサベースのコンピューティングデバイスを含む。一実施形態では、計算ユニット 115 は、本明細書内の他の場所に記載されるように、センサ 155 からセンサデータを受信し、センサデータを処理し、処理に基づいてインターフェース 175 を介して提示用の結果を生成および供給し、処理に基づいて様々なプログラムをトリガし、処理に基づいて検出システム 103 または関連システムの挙動や運動を制御し、計算サーバ 101 と連携してセンサデータを処理することができる。いくつかの実施形態では、計算ユニット 115 は、処理されたセンサデータや、そこから処理された結果を特徴データベース 197 に記憶することができる。プロセッサ 195 および検出モジュール 135 については、少なくとも図 2 A ~ 図 9 を参照してより詳細に記載される。

10

【0040】

インターフェース 175 は、ユーザ 125 と計算ユニット 115 との間の通信を処理するように構成されたデバイスである。例えば、インターフェース 175 には、ユーザ 125 に検出情報を表示するための画面、ユーザ 125 に音声情報を出力するためのスピーカ、音声やボイスコマンドを取り込むためのマイクロフォン、インジケータ（例えば、LED）、ならびに、ユーザ 125 との通信を容易にする任意の他の入力/出力構成要素のうちの一つまたは複数が含まれる。いくつかの実施形態では、インターフェース 175 は、計算ユニット 115 からユーザ 125 に出力を送信するように構成される。例えば、インターフェース 175 は、ユーザ 125 が近傍内にいることを計算ユニット 115 が検出した場合に、ユーザ 125 に音声で挨拶するためのオーディオシステムを含む。インターフェース 175 は、本明細書に記載される機能を提供するための他のタイプのデバイスを含む場合があることを理解されたい。

20

【0041】

ユーザ 125 は人間のユーザである。一実施形態では、ユーザ 125 は、道路上にある車両に座っているドライバーまたは同乗者である。別の実施形態では、ユーザ 125 は、ロボットと対話する、家屋内に位置する人間である。さらなる実施形態では、ユーザは、コンピューティングデバイスの従来のユーザである。ユーザ 125 は、計算ユニット 115 との間で様々なタイプのデータを送受信するインターフェース 175 と対話するか、または場合によっては、それに入力を供給し、出力を受信することができる。

30

【0042】

特徴データベース 197 は、一つもしくは複数のスタティックランダムアクセスメモリ（SRAM）デバイス、フラッシュメモリ、ハードディスクドライブ、ディスクベースのメモリデバイス（例えば、CD、DVD、Blue-ray（登録商標）、フロッピーディスク（登録商標）など）、または他の何らかの既知の揮発性もしくは不揮発性の物理的記憶デバイスなどの、非一時的記憶媒体に記憶される。特徴データベース 197 は、検出システム 103、または、検出システム 103 とは別個であるか、それに結合されるか、もしくはそれによってアクセス可能な別のコンピューティングデバイスやストレージシステムに含まれる場合がある。いくつかの実施形態では、特徴データベース 197 は、検出システム 103 や計算サーバ 101 によって動作可能なデータベース管理システム（DBMS）に関連するデータを記憶することができる。例えば、DBMS は、構造化問合せ言語（SQL）の DBMS、非 SQL の DBMS、フラットファイルシステム、従来型ファイルシステムなどを含んでもよい。場合によっては、DBMS は、行と列から構成される多次元テーブルにデータを記憶し、プログラムに基づいた動作を使用して、データの行を操作、すなわち、挿入、問合せ、更新、削除することができる。特徴データベース 197 のさらなる態様は下記で説明される。

40

50

【0043】

計算サーバ101は、検出システム103が物体を検出および認識することを容易にするための、プロセッサ(不図示)と、非一時的コンピュータ可読記憶媒体(例えば、メモリ)(不図示)とを有する、任意のコンピューティングデバイスである。いくつかの実施形態では、計算サーバ101は、検出モジュールのインスタンス135を含む。ネットワークベースの実施形態では、計算サーバ101は、検出システム103や他の情報源から、センサデータ、認識データ、位置データ、本明細書で説明された任意の他のデータを受信することができ、データを処理し、検出システム103に必要な処理結果を送る。

【0044】

図2Aは、図示された実施形態による、検出モジュール135と、プロセッサ195と、メモリ237と、通信ユニット245と、センサ155と、特徴データベース197とを含む、コンピューティングデバイス200のブロック図である。コンピューティングデバイス200の構成要素は、バス220によって通信可能に結合される。いくつかの実施形態では、コンピューティングデバイス200は、検出システム103や計算サーバ101のアーキテクチャの代表である。

【0045】

メモリ237は、データを記憶し、コンピューティングデバイス200の他の構成要素にデータへのアクセスを提供することができる。いくつかの実装形態では、メモリ237は、プロセッサ195によって実行され得る命令やデータを記憶することができる。例えば、メモリ237は、検出モジュール135やその構成要素を記憶することができる。メモリ237はまた、例えば、オペレーティングシステム、ハードウェアドライバ、他のソフトウェアアプリケーション、データベースなどを含む、他の命令およびデータを記憶することが可能である。メモリ237は、プロセッサ195およびコンピューティングデバイス200の他の構成要素との通信のために、バス220に結合することができる。

【0046】

メモリ237は、プロセッサ195によって、またはそれとともに処理するための、命令、データ、コンピュータプログラム、ソフトウェア、コード、ルーチンなどを、含有、記憶、通信、伝搬、または搬送することができる装置またはデバイスを含む、一つまたは複数の非一時的コンピュータ使用可能(例えば、読取り可能、書込み可能などの)媒体を含む。いくつかの実装形態では、メモリ237は、揮発性メモリおよび不揮発性メモリのうちの一つまたは複数を含む。例えば、メモリ237は、ダイナミックランダムアクセスメモリ(DRAM)デバイス、スタティックランダムアクセスメモリ(SRAM)デバイス、ディスクリットメモリデバイス(例えば、PROM、FPROM、ROM)、ハードディスクドライブ、光ディスクドライブ(CD、DVD、Blue-ray(登録商標)など)のうちの一つまたは複数を含む場合があるが、これらに限定されない。メモリ237は単一のデバイスであり得るか、または複数のタイプのデバイスおよび構成を含む場合があることを理解されたい。

【0047】

通信ユニット245は、ワイヤレスまたは有線接続を使用して通信可能に結合された他のコンピューティングデバイスにデータを送信し、それらからデータを受信する。通信ユニット245は、データを送受信するための一つまたは複数の有線インターフェースやワイヤレストランシーバを含む場合がある。通信ユニット245は、ネットワーク105に結合し、(構成に応じて)検出システム103や計算サーバ101などの他のコンピューティングノードと通信することができる。通信ユニット245は、ネットワーク105に関して上記で説明された通信方法などの標準の通信方法を使用して、他のコンピューティングノードとデータを交換することができる。

【0048】

検出モジュール135は、センサデータを受信するためにセンサ155に結合することができる。いくつかの実施形態では、センサ155から受信されるセンサデータは、深度画像を表す深度データを含む場合がある。深度画像は、一つまたは複数の物体を含むシー

10

20

30

40

50

ンを描写する画像であり得る。物体は、生物または非生物、生命体または非生命体などであり得る。例示的な物体には、人間、動物、家具、備品、車、器具などが含まれるが、これらに限定されない。

【0049】

例示的な検出モジュール135を描写する図2Bにおいて示されたように、検出モジュール135は、画像プロセッサ202と、特徴抽出器204と、分類モジュール206と、対象者位置推定器208と、シーン認識モジュール210とを含む場合があるが、検出モジュール135は、追加の構成要素を含む場合があり、また、様々な構成要素は単一のモジュールに組み合わせられるか、もしくはさらなるモジュールに分割される場合があることを理解されたい。

10

【0050】

画像プロセッサ202、特徴抽出器204、分類モジュール206、対象者位置推定器208、シーン認識モジュール210は、ソフトウェア、ハードウェア、または上記の組合せとして実装することができる。いくつかの実装形態では、画像プロセッサ202、特徴抽出器204、分類モジュール206、対象者位置推定器208、シーン認識モジュール210は、バス220やプロセッサ195により、互いに、またはコンピューティングデバイス200の他の構成要素に通信可能に結合することができる。いくつかの実装形態では、構成要素135、202、204、206、208、210のうちの一つまたは複数は、それらの機能を実現するためにプロセッサ195によって実行可能な命令のセットである。さらなる実装形態では、構成要素135、202、204、206、208、210のうちの一つまたは複数は、メモリ237に記憶され、それらの機能を実現するためにプロセッサ195によってアクセス可能かつ実行可能である。上記の実装形態のいずれにおいても、これらの構成要素135、202、204、206、208、210は、プロセッサ195およびコンピューティングデバイス200の他の構成要素との連携および通信に適合することができる。

20

【0051】

画像プロセッサ202、特徴抽出器204、分類モジュール206、対象者位置推定器208、シーン認識モジュール210のさらなる構造、動作、機能は、少なくとも図3～図9を参照して下記で説明される。

【0052】

図3は、抽出された特徴および推定された人間の位置に基づいてシーンを認識するための例示的なプロセスのフローチャートである。図示されたように、ロボットなどの検出システムは、人間などの対象者と対話する際に、検出された対象者の位置を利用して、対象環境についてのデータが蓄積されたデータベースに問い合わせを行う。検出システム103は、対象者検出/追跡システムを含むか、それと連携することができる。場合によっては、これは、対象者位置推定器208によって具現化される。対象者の位置は、本明細書で言及されるように、検出された対象者の現在の位置であってもよく、対象者の現在の速度に基づいて算出された、将来のある時点で予想される位置であってもよい。対象者位置推定器208は、データベースを検索するために、必要に応じてn次元で対象者を検出するように構成される。これらのn次元は、特徴データベース197に記憶された他の項目である特徴と同じ素性(origin)に関して参照されるか、またはそれと調整可能である。

30

40

【0053】

対象者の位置は、いくつかの異なる方法で検出することができる。いくつかの実施形態では、対象者位置推定器208は、本明細書内の他の場所で説明されるように、フレーム内に対象者を含むセンサ155によって取り込まれた画像データを処理して、既知の基準点に対する対象者の位置を特定することができる。さらなる実施形態では、対象者は、位置特定センサ(例えば、GPSセンサ)を備える携帯型電子デバイス(例えば、ウェアラブル、携帯電話、タブレット、ラップトップなど)などの、位置情報を特定できるデバイス、あるいは、そのデバイスのIPアドレス、またはサードパーティによりデバイス上で実行されたマルチラレーションもしくは三角測量に基づいてサードパーティによって特

50

定された位置情報を受信および供給することが可能なデバイスを携帯している場合がある。これらの実施形態では、対象者のデバイスは、ネットワーク 105（例えば、PAN, LAN, WAN など）を介して検出システム 103 または計算サーバ 101 に位置データを送信するようにプログラムすることができ、対象者位置推定器 208 は、通信ユニット 245 を介して位置データを受信し、それを処理して対象者の位置を推定することができる。

【0054】

ブロック 304 において対象者の位置が推定されると、検出システム 103 は、ブロック 306 において、本明細書によって開示されるシーン認識方法を使用して、対象者と対話する方法を決定する。場所（より具体的には、対象者に直接関係する可能性が最も高いシーン）を特定することで、検出システム 103 によって遂行されるべきインタラクションの適切な方法についての貴重なデータ（例えば、どの固有情報が対象者に伝達されるべきか、どの情報が対象者から要求されるか、困難な手法を実行する方法、対象者が検出システム 103 に現在要求しているタスクのタイプなど）が提供される。

10

【0055】

人物の近傍で収集されたセンサデータが多くない場合、以下の理由のうちの少なくとも一つまたは複数に起因して、正確な場所認識ができない場合がある。

- ・対象者が、シーン分類のための情報が十分に無い環境にいる場合（例えば、カメラが壁に向いている場合や、マイクが静かな部屋に設置されている場合など）
- ・対象者が、二つの部屋の間にある境界領域（boundary region）に立っている場合（センサが指向性である場合、センサの方向によって、得られる分類結果が変わってしまう）
- ・対象者が位置する部屋が、多目的な部屋である場合（部屋の異なる部分が、時刻によって異なる目的に変わる場合など）

20

【0056】

これらの場合の各々について考えると、少なくとも、正しい分類結果は、センサが見ることができるものではなく、対話対象者が環境内で位置する場所に依存する。すなわち、背景技術で記載した解決策は不十分なものである。

図 3 において描写され、本明細書を通して記載される手法は、対象者の位置を検出および追跡することによってこの問題を解決し、当該位置を使用して、環境内で取り込まれたセンサデータから抽出された特徴を蓄積したデータベース 197 に問い合わせる。

30

詳細には、一つまたは複数のセンサ 155 が、センサデータの様々なセットを取り込む。ブロック 302 において、特徴抽出器 204 は、センサデータから環境に応じた特徴を抽出するように構成される。

なお、本発明における特徴とは、環境中に存在するものであれば、どのようなものであってもよい。例えば、特徴は、人間の周辺に存在しうる物体（家具、食器、書籍、文具類、電化製品、小物類、生活用品等）であってもよい。

分類モジュール 206 は、確率スコアを使用して特徴をシーンタイプに分類するように構成される。抽出された特徴を表す特徴データと、確率スコアを反映する分類データは、互いに環境に関連して、特徴データベース 197 に記憶される。

40

【0057】

図 4 は、特徴データベース 197 に記憶されるデータを表した図である。図示されたように、特徴データベース 197 は、特徴データ 402、分類データ 404、シーンデータ 406、環境データ 408 を記憶する。

【0058】

特徴データベース 197 には、多様な方法でデータを投入することができる。例えば、検出システム 103（例えば、ロボット）は、検出システム 103 がセンサ 155 を利用して、環境を自律的に自己探索するようにプログラムすることができ、環境の様々な態様を表すセンサデータを取り込むことができる。また、特徴抽出器 204 は、センサデータを処理して特徴を抽出し、抽出された特徴を特徴データベース 197 に投入することができる。

50

別の例では、環境内に設置された固定センサ、一つ以上の検出システム103のセンサ155、環境内の他のデバイスに含まれるセンサ（例えば、家電、ビルコントロール、セキュリティシステム、ポータブルデバイスなどのユーザエレクトロニクス製品）などの複数の異なるセンサが、センサデータを取得し、ネットワーク105を介して、特徴抽出302による処理のために検出システム103や計算サーバ101に供給し、供給されたこれらのセンサデータのセットや、これらから抽出された特徴を調整（例えば、同一の基準点から取得、スケールマッチング、同じ素性について格納する等）してもよい。

【0059】

特徴データ402は、特定の環境内で検出され、特定の環境に関係する多様な特徴を含む。例えば、特徴は点群に含まれる点である。例えば、M個の点群がRGB-Dセンサによって取得されると、点群または個別センサのいずれかを使用して特徴やシーンが分類され、各3Dポイントに対応するM個の特徴が、得られた分類スコアとともに記録される。

10

【0060】

別の例では、特徴は、一点の家具または検出されたカップもしくは皿のような環境内の対象物を反映した情報である。各物体は三次元の物理的位置、関連付けられたサイズ、および関連する検出時刻（detection time）を有する。さらに、各物体は様々なタイプのシーンの各々に含まれる尤度を有する。例えば、皿は、ほとんどの場合、台所または食事エリアで見つかり、その存在はそれらの環境（台所または食事エリア）のうちの一つを暗示するはずである。

【0061】

いくつかの実施形態では、特徴は分類スコアと関連付けられる。分類スコアは、分類器によって認識された特定のシーンタイプに特徴が属する確率、その特徴を使用して正確にシーンを分類する事前確率、任意の基準点に対するN次元における位置およびサイズを表したデータである。

20

【0062】

図示されたように、各特徴は、追跡および検索の目的で、特徴IDなどの一意の識別子を持つ場合がある。特徴の説明、当該特徴が環境内にあるはずの時間または時刻（毎日でない場合、当該特徴が環境内にあるはずの任意の日付）、特徴のサイズ、および環境内の特徴の位置などの、各特徴の様々な態様を記憶することができる。

【0063】

分類データ404は、一つまたは複数の分類された環境内の位置、位置に関連する特徴、および特徴がそれらの位置内に存在する尤度を反映する尤度スコアをインデックス付ける。いくつかの実施形態では、環境内のいくつかの位置は、環境についての追加情報を取得するにつれて時間とともに更新されるシードデータ（初期尤度）をプリフェッチし、投入することができる。所定の環境についての追加の位置や特徴などを表すデータが取り込まれるにつれて、分類データは、既存の特徴および位置についてのより関係する尤度スコア、新しい特徴および位置についての新しい尤度スコアなどの、より正確な情報で更新される。

30

【0064】

描写された例では、分類データは、各特徴の特徴IDを、各位置（2D、3Dなどの座標または他の適切な一意の識別情報を使用して表現することができる）と、当該特徴IDに関連付いた特徴が存在する尤度を表す尤度スコアと関連付ける。加えて、データの完全性の目的で、二つ以上の環境が所定の検出システム103によって処理されるため、特徴IDおよび位置は、特徴が抽出された特定の環境の環境IDに合わせることができる。

40

【0065】

シーンデータ406は、シーンの種別（以下、シーンタイプ）ごとの一意の識別子、タイトルおよび説明を含むデータである。また、環境データ408は、検出システム103によって監視されている任意の環境についての一意の識別子、各環境の寸法、および各環境の説明を含むデータである。なお、シーンタイプは、対象者が自分自身を見つけることができる任意の特定のシーンを反映するようにあらかじめ決められることができ、（例えば、

50

ロボット、ロボットのグループ、一つまたは複数のロボットと通信するサーバシステムなどによって)機械学習することができる。例示的なシーンタイプとして、自宅またはオフィスなどの住居の様々な部屋、バス停留所、中庭、歩道、横断歩道、建物の階段といった公共の屋外ロケーションが含まれ得る。

【0066】

図5は、シーンを認識するための例示的な方法500のフローチャートである。

ブロック502において、センサ155が、特定の環境を表すセンサデータを取り込む。

センサデータは特徴抽出器204に供給され、ブロック504において、対象者位置推定器が当該センサデータを処理し、環境内にいる一人または複数の対話対象者、および環境内の当該対象者の位置を検出する。追加または代替として、対象者位置推定器208は、本明細書内の他の場所で説明されるサードパーティソースなどの他のソースから位置データを受信して、環境内にいる対象者の位置を特定してもよい。

10

【0067】

ブロック506において、シーン認識モジュール210が、環境に関係し、関連付けられ、または関連付け可能である、多様な特徴が投入された特徴データベース197に、位置(location)および探索領域の寸法(search region dimension)を用いて問い合わせる。いくつかの実施形態では、尤度値は事前に生成され、(例えば、分類データとして)特徴データベース197に記憶される。さらなる実施形態では、尤度値は、シーンタイプごとに生成され、実行時に更新される場合がある。

20

【0068】

ブロック508において、シーン認識モジュール210が、探索領域の寸法によって定義された探索領域内に特徴が存在する尤度値に基づいて、様々なシーンタイプの中からシーンタイプを選択する。

【0069】

ブロック510において、検出モジュール135は、選択されたシーンタイプに基づいて一つまたは複数の自律ルーチンを実行し、それにより、検出システム103が対象者と正確かつ文脈的に対話することが可能になる。

例えば、シーンタイプを特定した後で、検出モジュール135は、シーンタイプに基づいて、シーンタイプに関連する情報の取出し、対象者や一つ以上の特徴と対話する(例えば、ユーザの名前を使用してユーザに挨拶する)ための一つまたは複数の出力デバイス(例えば、ディスプレイ、スピーカ、センサ、モチベータなど)の制御、物体(例えば、特定の人物/ユーザ)に関連するアカウント情報の取得などの、シーンタイプに基づく動作を実行するプログラムの動作をトリガすることができる。

30

【0070】

図6は、特定の環境の特徴を抽出して分類し、分類された特徴を特徴データベースに投入するための例示的な方法600のフローチャートである。

ブロック602において、画像プロセッサ202が、センサ155を使用して特定の環境をスキャンする。センサからセンサデータを受信すると、画像プロセッサ202は物体についてのセンサデータを処理する。

40

【0071】

画像プロセッサ202は、センサデータを受信および処理するためにセンサ155に通信可能に結合され、本明細書内の他の場所で説明されるさらなる処理のために、特徴抽出器204および対象者位置推定器208に処理されたセンサデータを供給することができる。いくつかの実施形態では、センサデータは、基準点に対する物体の位置を記述する深度画像データを含む場合がある。例えば、センサ155は、センサ155によって取り込まれた物体を含む深度画像を記述する多次元(例えば3D)データを生成する多次元深度センサを含む。深度画像データは、画像内の物体を形成するピクセルについてのRGB値を含む場合がある。場合によっては、深度画像データは、トリプレットまたは空間座標の

50

アレイの形態の多次元（例えば3D）深度点群などの、物体と関連付けられた位置情報を含む場合がある。場合によっては、深度画像データは、そのX座標とY座標を表す各ピクセルの列番号と行番号、ならびにそのZ座標を表すピクセルの値を含む場合がある。

【0072】

画像プロセッサ202は、センサ155によって取り込まれた深度画像を記述する深度画像データを使用して、深度画像に含まれる個別の物体を特定することができる。深度撮像を使用すると、物体のセグメント化を簡略化することを含む様々な利点を得ることができる。深度画像内で、物体は、しばしば、それらの相対深度により画像内で互いから分離することができる。例えば、（センサ155の位置などの所定の基準点から測定された）同じ相対距離を有する二つの隣接するピクセルは、同じ物体に属する可能性があるが、基準点に対してかなり異なる距離を有する二つのピクセルは、画像内の異なる物体に属する可能性がある。これは、独立して立っている物体を互いからより容易に区別することに役立つ。

10

【0073】

ブロック604において、特徴抽出器204が、センサ155によって供給され、画像プロセッサ202によって処理されたセンサデータを使用して、環境の特徴を抽出する。例えば、特徴抽出器204は、画像プロセッサ202からセンサデータ内の物体を記述する物体データを受信することができ、物体データおよびセンサデータをさらに処理して、基準点に対する特徴の位置、特徴のサイズ、特徴の位置特定時間（location time of the feature）、特徴の位置特定頻度（location frequency of the feature）、特徴につい

20

【0074】

ブロック606において、分類モジュール206が、各シーンタイプに対する特徴ごとの尤度スコアを生成する。尤度スコアは、その特徴が当該シーンタイプ内で現れる確率を表す。尤度スコアを生成するための例示的な方法が、図7を参照して下記に記載される。

【0075】

尤度スコアが生成された後、ブロック608において、分類モジュール206が、本明細書内の他の場所でさらに説明されるように、検出された物体ごとに抽出された特徴、およびシーンタイプに対する特徴についての分類スコアを、環境用の特徴データベース197に投入する。

30

【0076】

いくつかの実施形態では、各特徴は2次元または三次元の特徴仕様（feature specification）を有し、追加の属性（検索基準）として、時間を有する場合もある。これは、結果の解像度を高めることができるため有利である。例えば、多目的空間（例えば部屋）の用途は、時間とともに変化する場合がある。テーブルを有する空間は、1日のほとんどの間便利な会議空間であり、物体ベースまたはリアルタイム画像/音声分類は、それを正確に分類することができる。しかしながら、昼食時には、それは食事スペースとなり、食べ物および皿、ならびに、食事に関する他の物体（例えば図8参照）の存在が、割り当てられた分類スコアに影響を及ぼす。

40

対象者の位置（または対象者の位置ベースの探索エリア）と時間の両方によってデータベースを検索することで、元来記述することが困難である多目的環境の、より微細であり、正確な分類が可能になる。

【0077】

いくつかの実施形態では、特徴データベース197は、検出された対象者の近傍における場所認識の傾斜（偏り）を生成するために使用することができる。ある特定の場所タイプ、例えば台所を示す特徴が対象者の周囲にランダムに分散されている場合、シーン認識モジュール210によって生成された傾斜は、「台所らしさ」の指向性が最小であることを示す。

しかし、台所に強く関連する特徴や、台所についてのシーン結果が、人物の片側におい

50

て支配的である、すなわちランダムに分散されていない場合、傾斜は、支配的な特徴への強い指向性を示す。すなわち、特定のシーンが特定の方向に強く現れていることを示す。

シーン認識モジュール 210 は、傾斜を使用して遷移領域（すなわち、台所と廊下の間といったように場所やシーンが変化する領域）を識別し、それにより、対象者と対話するように検出システム 103 の挙動を制御する。

【0078】

例えば、シーン認識モジュール 210 によって特定された指向性は、ロボットが対話のための状況を探ることを支援することができる。ある特定のシーンに強く関連付けられた領域の近くに位置する人物は、それらのエリアを利用するために、しばしばその近傍にいる。指向性センサ（例えばカメラ）との局所的な傾斜または注意機構の焦点（focus of an attention mechanism）に従うことによって、例えば、図 8 に関して本明細書内の他の場所で説明されるように、ロボットは、人物との対話のための環境状況を識別するか、または人間の活動を認識するその能力を向上させることができる。

10

【0079】

図 7 は、スコアを計算するための例示的な方法 700 のフローチャートである。図示されたように、ブロック 702 において、対象者位置推定器 208 は、本明細書内の他の場所で説明されるように、 n 次元で表された対象者の位置を特定し、ブロック 704 において、シーン認識モジュール 210 は、探索エリアの寸法を特定する。いくつかの実施形態では、探索エリアの寸法は、アプリケーションに最も適するように、実行時またはコンパイル時に指定される場合がある。

20

いくつかのさらなる実施形態では、シーン認識モジュール 210 は、適応的に探索エリアの寸法を計算する。例えば、初期寸法から始めて、位置が当該寸法を満たす特徴についてデータベースへの問い合わせを行い、結果が曖昧な場合（すなわち、信頼度が所定の閾値よりも低い場合や、対象者周辺における顕著な特徴が所定の閾値よりも小さい場合）に、所定の閾値（最大、最小等）が満たされるまで寸法を大きく（または小さく）していてもよい。

探索エリアの寸法の一例は探索半径であり得るが、探索エリアは他の形態（多角形、楕円、不連続など）を取る場合があり、寸法は、その境界を特定するのに十分なエリアの複数の態様を記述することができることを理解されたい。

【0080】

次に、ブロック 706 において、方法 700 は、探索エリアの寸法によって定義された探索エリア内の特徴のサブセット f_f を特定する。いくつかの実施形態では、シーン認識モジュール 210 は、探索エリアの寸法によって定義されたエリア内に含まれている特徴について特徴データベース 197 に問い合わせる。特徴は、ある特定の基準点に対する環境内の位置およびサイズによってインデックス付けすることができ、問合せは、探索エリアの寸法に基づいて対象のエリアを指定することができる。

30

【0081】

図 8 は、探索エリアおよびその中に含まれている特徴のさらなる例を示す。具体的には、図 8 は、シーンを認識する自律ロボット 826 を含む家屋 800 を表したものである。家屋 800 は住宅である。対象者 820 は、居間 802 と台所 804 との間に位置し、住宅は、寝室 806、浴室 808、およびトイレ、リアエントランス 810 などの、いくつかの他の部屋を含む。これらの部屋の各々についてのシーンタイプは、あらかじめ定義され、自律ロボット 826（この場合、検出システム 103 を反映する）によるアクセスのために記憶される。

40

【0082】

居間 802 は、ソファ 818、サイドテーブル 822、絨毯 816 などの、居間の中に収容される典型的な物体を含む。同様に、台所 804 は、この例ではおおよそ食事時なので、食器類がセットされているダイニングテーブル 814 などの、典型的な台所にある様々な物体を含む。

【0083】

50

ロボット 826 は、対話対象者 820 の近くに位置する。ロボット 826 は、居間 802 と台所 804 との間を移動する対象者 820 の位置を追跡している。点線 828 は、居間 802 と台所 804 の間の移動を表すものである。

ロボット 826 は対象者 820 が移動していると判断したため、ロボット 826 のシーン認識モジュール 210 は、対象者 820 が移動している先の近傍に位置する、半径 r によって定義された探索エリア内の特徴（この場合、食器がセットされたダイニングテーブル）について特徴データベース 197 に問い合わせる。

特徴データベース 197 において、85% の確率でダイニングテーブル 814 が台所 804 内に位置し、95% の確率で食器類が台所内に位置するものとする。

対照的に、45% の確率でダイニングテーブル 814 が居間 802 内に位置し、25% の確率で食器類が居間 802 内に位置するものとする（対象者は、時々ソファで食事を取るため）。

これらの値および/または他の変数（例えば、時刻、以前の履歴など）を利用して、シーン認識モジュール 210 は、ユーザが台所に向かっていると判断する。また、当該判断に回答して、検出モジュール 135 が、ユーザが台所に向かっていると判断に基づいて、ロボット 826 の挙動を制御する。例えば、ロボット 826 は、冷蔵庫または流し台に向かう誘導が必要とされるかどうか、食事を準備するための情報が必要かどうか、または任意の他の人々をテーブルに呼ぶ必要があるかどうかを、対象者 820 に尋ねることができる。

【0084】

図 7 に戻る。ブロック 708 において、人の位置を中心とする、指定された半径内の特徴のサブセット ff が識別される。また、方法 700 は、下記の式を使用して、個々の特徴（ ff における i ）が与えられた各シーンタイプの確率 $P(s | ff_i)$ 、および ff_i を観測する確率から、複合シーン確率 $P(ff | s)$ を、シーンタイプ s ごとに計算する。

【数 1】

$$P(ff | s) = \eta \prod_i P(s | ff_i) P(ff_i)$$

【0085】

背景技術において説明された手法などの他の手法に比べて、本明細書に記載された、人間中心の場所認識技法のさらなる利点は、図 9 の表 900 にて要約されている。

【0086】

上記の説明では、説明の目的で、本開示を完全に理解するために、多数の具体的な詳細が明記されている。しかしながら、本明細書に記載された技術は、これらの具体的な詳細なしに実践され得ることを理解されたい。さらに、説明を不明瞭にすることを回避するために、様々なシステム、デバイス、および構造がブロック図の形で示される。例えば、特定のハードウェア、ソフトウェア、およびユーザインターフェースを有するように、様々な実装形態が記載される。しかしながら、本開示は、データおよびコマンドを受信することができる任意のタイプのコンピューティングデバイス、ならびにサービスを提供する任意の周辺デバイスに適用される。

【0087】

場合によっては、様々な実装形態は、コンピュータメモリ内のデータビット上の演算のアルゴリズムおよびシンボル表現の観点から、本明細書に提示される場合がある。本明細書において、かつ一般的に、アルゴリズムは、望ましい結果に導く自己矛盾のない 1 組の演算であると考えられる。演算は、物理量の物理的操作を必要とする演算である。必ずしもそうであるとは限らないが、通常、これらの量は、記憶、転送、合成、比較、および他の方法で操作することが可能な電気信号または磁気信号の形態をとる。主に共通使用の理由で、これらの信号をビット、値、素子、シンボル、文字、用語、番号などとして参照することは、時には好都合であると証明されている。

【0088】

10

20

30

40

50

しかしながら、これらおよび同様の用語のすべてが、適切な物理量と関連付けられ、これらの量に適用される便利なラベルにすぎないことは留意されるべきである。以下の説明から明白であると別段に明記されていない限り、本開示全体にわたって、「処理」、「計算」、「算出」、「決定」、「表示」などを含む用語を利用する説明は、コンピュータシステムのレジスタおよびメモリ内の物理（電子）量として表されるデータを操作し、コンピュータシステムのメモリもしくはレジスタ、または他のそのような情報を記憶、送信、もしくは表示するデバイス内の物理量と同様に表される他のデータに変換する、コンピュータシステムまたは同様の電子コンピューティングデバイスのアクションおよびプロセスを参照する。

【0089】

本明細書に記載された様々な実装形態は、本明細書において演算を実行するための装置に関係する場合がある。この装置は、必要な目的のために特別に構築される場合があるか、またはコンピュータに記憶されたコンピュータプログラムによって、選択的にアクティブ化または再構成される汎用コンピュータを備える場合がある。そのようなコンピュータプログラムは、限定はしないが、各々がコンピュータシステムバスに結合される、フロッピーディスク、光ディスク、CD ROM、および磁気ディスクを含む任意のタイプのディスク、読取り専用メモリ（ROM）、ランダムアクセスメモリ（RAM）、EPROM、EEPROM、磁気カードもしくは光カード、不揮発性メモリを有するUSBキーを含むフラッシュメモリ、または電子命令を記憶するのに適した任意のタイプの媒体を含む、コンピュータ可読記憶媒体に記憶される場合がある。

【0090】

本明細書に記載された技術は、ハードウェアの実装形態、ソフトウェアの実装形態、またはハードウェアとソフトウェアの両方の構成要素を含んでいる実装形態の形態をとることができる。例えば、技術は、限定はしないが、ファームウェア、常駐ソフトウェア、マイクロコードなどを含む、ソフトウェアに実装することができる。さらに、技術は、コンピュータもしくは任意の命令実行システムによって使用されるか、またはそれとともに使用されるプログラムコードを提供するコンピュータ使用可能媒体またはコンピュータ可読媒体からアクセス可能な、コンピュータプログラム製品の形態をとることができる。この説明の目的で、コンピュータ使用可能媒体またはコンピュータ可読媒体は、命令実行システム、命令実行装置、または命令実行デバイスによって使用されるか、またはそれとともに使用されるプログラムを、内蔵、記憶、通信、伝搬、または転送することができる、任意の非一時的記憶装置であり得る。

【0091】

プログラムコードを記憶または実行するのに適したデータ処理システムは、システムバスを介してメモリ素子に直接的または間接的に結合された少なくとも一つのプロセッサを含む場合がある。メモリ素子は、プログラムコードの実際の実行中に利用されるローカルメモリ、バルクストレージ、および、実行中にバルクストレージからコードが取り出されねばならない回数を低減するために、少なくともいくつかのプログラムコードの一時的ストレージを提供するキャッシュメモリを含むことができる。（限定はしないが、キーボード、ディスプレイ、ポインティングデバイスなどを含む）入力/出力デバイスまたはI/Oデバイスは、直接または介在するI/Oコントローラを介して、システムに結合することができる。

【0092】

ネットワークアダプタはまた、データ処理システムが、介在する専用ネットワークおよび/または公共ネットワークを介して、他のデータ処理システム、ストレージデバイス、リモートプリンタなどに結合されるようになることを可能にするために、システムに結合することができる。ワイヤレス（例えば、Wi-Fi（登録商標））トランシーバ、Ethernetアダプタ、およびモデムは、ネットワークアダプタのほんの数例にすぎない。専用ネットワークおよび公共ネットワークは、任意の数の構成および/またはトポロジーを有する場合がある。データは、例えば、様々なインターネットレイヤ、トランスポート

10

20

30

40

50

トレイヤ、またはアプリケーションレイヤのプロトコルを含む、様々な異なる通信プロトコルを使用して、ネットワークを介してこれらのデバイス間で送信することができる。例えば、データは、伝送制御プロトコル/インターネットプロトコル(TCP/IP)、ユーザデータグラムプロトコル(UDP)、伝送制御プロトコル(TCP)、ハイパーテキスト転送プロトコル(HTTP)、セキュアハイパーテキスト転送プロトコル(HTTPS)、動的適応ストリーミングオーバーHTTP(DASH)、リアルタイムストリーミングプロトコル(RTSP)、リアルタイム転送プロトコル(RTP)およびリアルタイム転送制御プロトコル(RTCP)、ボイスオーバーインターネットプロトコル(VOIP)、ファイル転送プロトコル(FTP)、ウェブソケット(WS)、ワイヤレスアクセスプロトコル(WAP)、様々なメッセージングプロトコル(SMS、MMS、XMS、IMAP、SMTP、POP、WebDAVなど)、または他の既知のプロトコルを使用して、ネットワークを介して送信される場合がある。

10

【0093】

最後に、本明細書に提示された構造、アルゴリズム、および/またはインターフェースは、任意の特定のコンピュータまたは他の装置に本質的に関係しない。様々な汎用システムは、本明細書における教示によるプログラムとともに使用される場合があるか、または必要な方法ブロックを実施するためにより特化した装置を構築することが好都合であると証明する場合がある。様々なこれらのシステムに必要な構造は、上記の説明から明らかになる。加えて、本明細書は、いかなる特定のプログラミング言語も参照して記載されていない。本明細書に記載された明細書の教示を実装するために、様々なプログラミング言語を使用できることが諒解されよう。

20

【0094】

上記の発明を実施するための形態は、例示および説明の目的で提示されている。本明細書を開示されたそのままの形態に徹底または限定するものではない。上記の教示に照らして、多くの修正形態または変形形態が可能である。本開示の範囲は、この発明を実施するための形態によって限定されず、むしろ本出願の特許請求の範囲によって限定されるものである。当業者によって理解されるように、本明細書は、その趣旨または本質的な特性から逸脱することなく、他の固有の形態で具現化される場合がある。同様に、モジュール、ルーチン、特徴、属性、方法、および他の態様の特定の命名および分割は、必須または重要ではなく、本明細書またはその特徴を実装するメカニズムは、様々な名称、分割、および/またはフォーマットを有する場合がある。

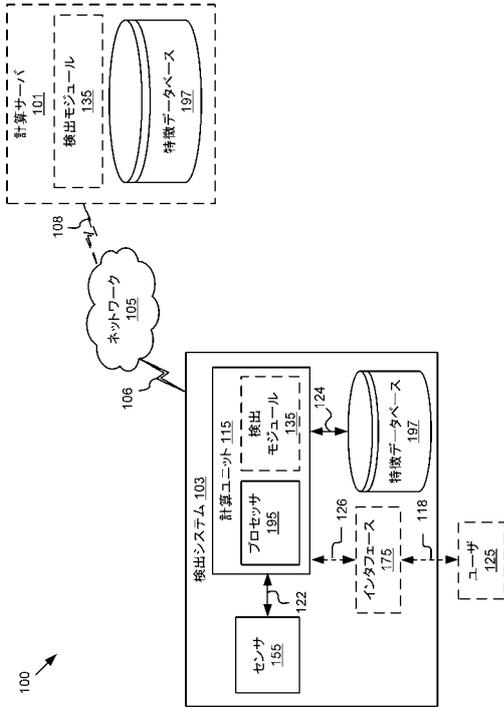
30

【0095】

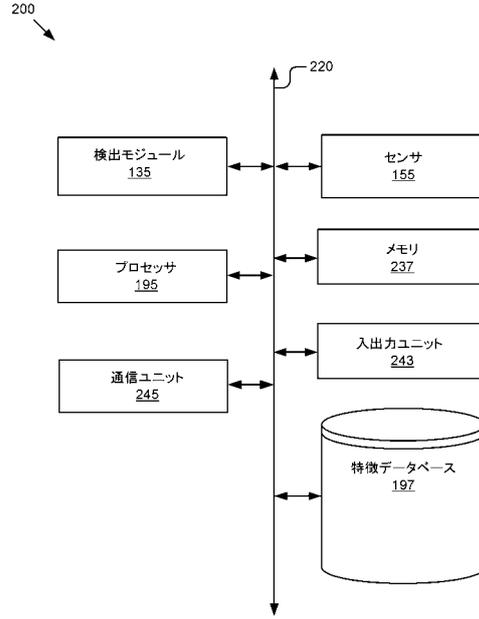
さらに、本開示のモジュール、ルーチン、特徴、属性、方法、および他の態様は、ソフトウェア、ハードウェア、ファームウェア、または上記の任意の組合せとして実装することができる。また、その例がモジュールである本明細書の構成要素がソフトウェアとして実装される場合はいつでも、構成要素は、スタンドアロンプログラムとして、より大きいプログラムの一部として、複数の別々のプログラムとして、静的もしくは動的にリンクされたライブラリとして、カーネルのロード可能なモジュールとして、デバイスドライバとして、ならびに/または、現在知られているか、もしくは将来知られるすべておよび任意の他の方法で、実装することができる。加えて、本開示は、任意の固有のプログラミング言語における、または任意の固有のオペレーティングシステムもしくは動作環境のための実装に少しも限定されない。

40

【図1】



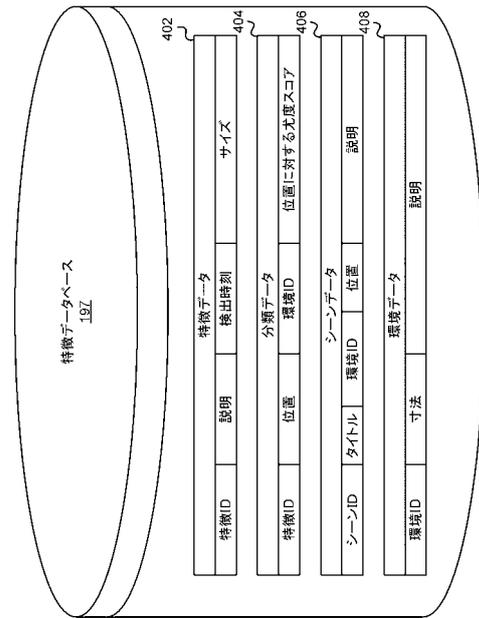
【図2A】



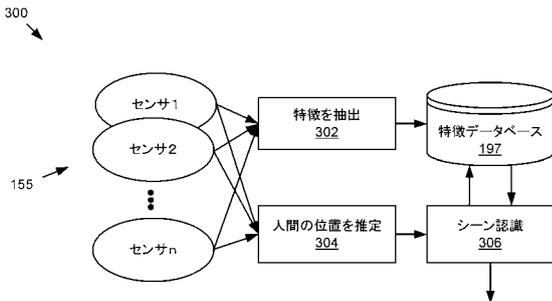
【図2B】



【図4】

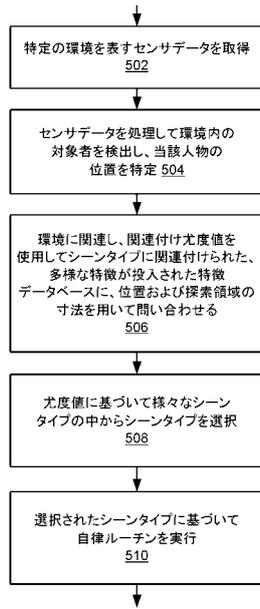


【図3】



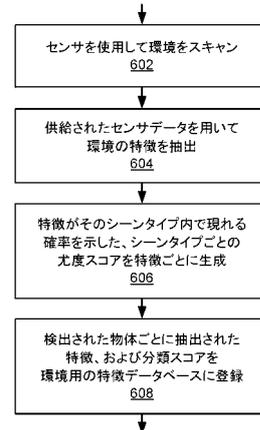
【 図 5 】

500



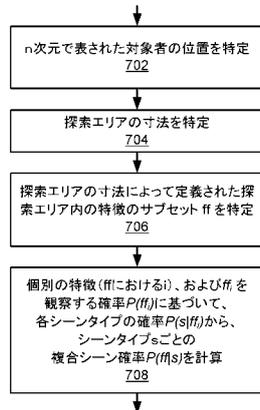
【 図 6 】

600



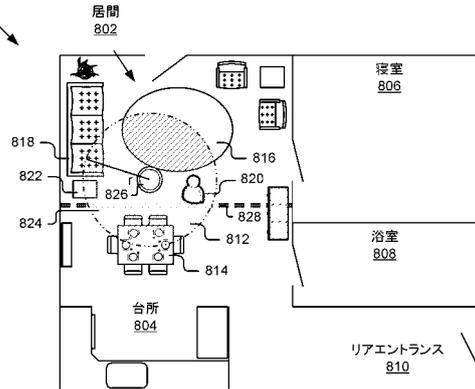
【 図 7 】

700



【 図 8 】

800



【 図 9 】

900 →

	物体ベースの推定	ディープラーニング によるソリューション	トポジカルマップ	占有グリッド	人間中心の 場所認識
シーンタイプの推定	True	True	True	True	True
センサ位置に対する 強固さ	False	False	True	True	True
人間ロボット間 対話についての 曖昧さを無くせるか	False	False	False	False	True
物体とポイントデータ の両方を含む	False	False	False	False	True
時間ベースの探索	False	False	False	False	True

フロントページの続き

(74)代理人 100138357

弁理士 矢澤 広伸

(74)代理人 100176201

弁理士 小久保 篤史

(72)発明者 マーティンソン, エリック

愛知県豊田市トヨタ町 1 番地 トヨタ自動車株式会社内

(72)発明者 キム, デヴィッド

愛知県豊田市トヨタ町 1 番地 トヨタ自動車株式会社内

(72)発明者 中野 雄介

愛知県豊田市トヨタ町 1 番地 トヨタ自動車株式会社内

Fターム(参考) 5L096 AA06 AA09 BA05 CA02 FA69 GA30 HA09 JA03 JA11