

(12) 特許協力条約に基づいて公開された国際出願

(19) 世界知的所有権機関
国際事務局

(43) 国際公開日
2021年3月18日(18.03.2021)



(10) 国際公開番号

WO 2021/048985 A1

(51) 国際特許分類:

G06T 7/70 (2017.01)

(21) 国際出願番号: PCT/JP2019/035972

(22) 国際出願日: 2019年9月12日(12.09.2019)

(25) 国際出願の言語: 日本語

(26) 国際公開の言語: 日本語

(71) 出願人: 株式会社ソニー・インタラクティブエンタテインメント (SONY INTERACTIVE ENTERTAINMENT INC.) [JP/JP]; 〒1080075 東京都港区港南1丁目7番1号 Tokyo (JP).

(72) 発明者: 堀川 勉 (HORIKAWA, Tsutomu); 〒1080075 東京都港区港南1丁目7番1号 株式

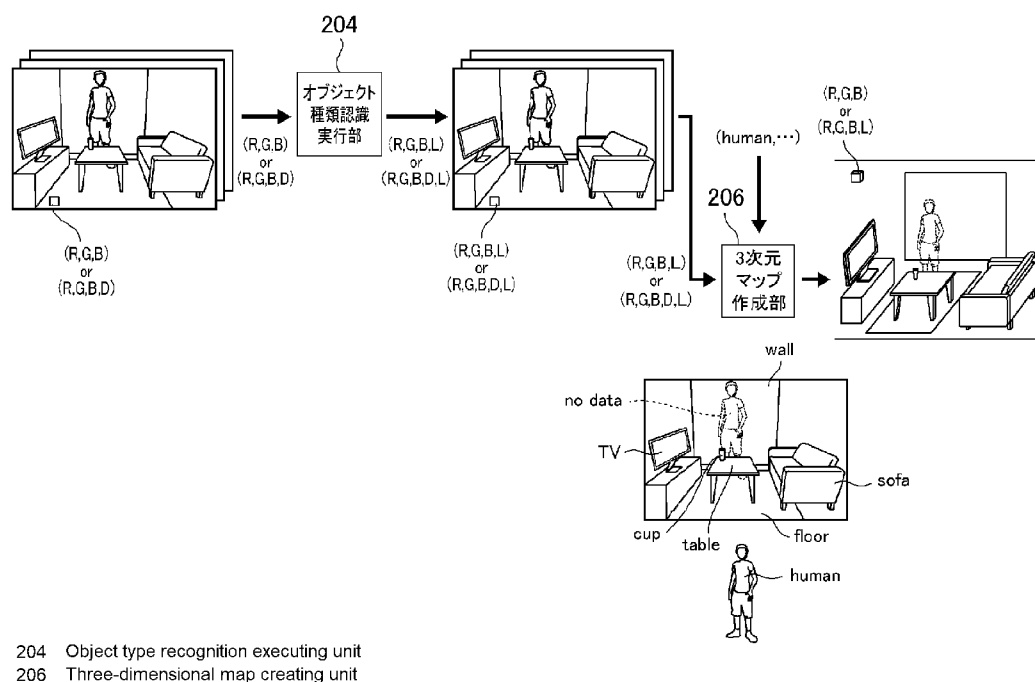
会社ソニー・インタラクティブエンタテインメント内 Tokyo (JP). 小野 大地(ONO, Daichi); 〒1080075 東京都港区港南1丁目7番1号 株式会社ソニー・インタラクティブエンタテインメント内 Tokyo (JP). 矢部 博之(YABE, Hiroyuki); 〒1080075 東京都港区港南1丁目7番1号 株式会社ソニー・インタラクティブエンタテインメント内 Tokyo (JP).

(74) 代理人: 特許業務法人はるか国際特許事務所 (HARUKA PATENT & TRADEMARK ATTORNEYS); 〒1020085 東京都千代田区六番町3六番町SKビル5階 Tokyo (JP).

(81) 指定国(表示のない限り、全ての種類の国内保護が可能): AE, AG, AL, AM, AO, AT, AU, AZ,

(54) Title: IMAGE PROCESSING DEVICE, IMAGE PROCESSING METHOD, AND PROGRAM

(54) 発明の名称: 画像処理装置、画像処理方法、及び、プログラム



(57) Abstract: Provided are an image processing device, an image processing method, and a program with which an accurate three-dimensional map can be formed even when a moving object is included in an input image. This image processing device includes: an image acquiring unit that sequentially acquires two-dimensional photographic images; an object type recognition executing unit that applies, to each pixel in the sequentially acquired photographic images, a label indicating the type of an object represented by said pixel; and a three-dimensional map creating unit that executes three-di-

BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

- (84) 指定国(表示のない限り、全ての種類の広域保護が可能): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), ユーラシア (AM, AZ, BY, KG, KZ, RU, TJ, TM), ヨーロッパ (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

添付公開書類:

- 一 国際調査報告 (条約第21条(3))

mensional position recognition for each pixel in the photographic images to create a three-dimensional map on the basis of the sequentially acquired photographic images. The three-dimensional map creating unit limits the three-dimensional position recognition for each pixel in the photographic images in accordance with the label applied to said pixel.

(57) 要約: 入力画像に移動するオブジェクトが含まれる場合であっても、精度よく3次元マップを形成することができる画像処理装置、画像処理方法、及び、プログラムを提供する。画像処理装置は、2次元の撮像画像を順次取得する画像取得部と、順次取得された前記撮像画像の各ピクセルに、該ピクセルにより表されたオブジェクトの種類を表すラベルを付すオブジェクト種類認識実行部と、順次取得された前記撮像画像に基づいて、前記撮像画像の各ピクセルの3次元位置認識を実行して3次元マップを作成する3次元マップ作成部と、を含み、前記3次元マップ作成部は、前記撮像画像の各ピクセルの3次元位置認識を、該ピクセルに付されたラベルに応じて制限する。

明 細 書

発明の名称：画像処理装置、画像処理方法、及び、プログラム
技術分野

[0001] 本発明は、画像処理装置、画像処理方法、及び、プログラムに関する。

背景技術

[0002] 入力画像に基づいて、自己位置推定と入力画像に表されたオブジェクトの３次元位置認識を同時に実行することで３次元マップ形成を行う技術（いわゆる、スラム(SLAM: Simultaneous Localization and Mapping))が知られている。

[0003] また、入力画像の各ピクセルに対して、該ピクセルに表された人物、道路、空等のオブジェクトの種類を認識し、当該種類を表すラベルを各ピクセルに付すオブジェクト種類認識技術（いわゆる、セマンティックセグメンテーション(Semantic segmentation))が知られている。当該オブジェクト種類認識は、学習データを用いて学習済である機械学習モデルを用いて行われる。

発明の概要

発明が解決しようとする課題

[0004] 自己位置推定と入力画像に表されたオブジェクトの３次元位置認識を同時に実行することで３次元マップを形成する技術は、当該入力画像に表されたオブジェクトが静止していることを前提としている。そのため、入力画像に表されたオブジェクトに人物や動物等の動く（位置、姿勢、形状等が時間に変化する）オブジェクトが含まれている場合、３次元マップを正しく形成することができない。また、入力画像に表されたオブジェクトがコップ等の静止しているものであっても、３次元マップを形成後に当該コップを動かした場合、３次元マップを再度作成する必要がある。

[0005] 本発明は、上記問題点に鑑みてなされたものであり、その目的は、入力画像に移動するオブジェクトが含まれる場合であっても、精度よく３次元マップを形成することができる画像処理装置、画像処理方法、及び、プログラム

を提供することである。

課題を解決するための手段

- [0006] 上記課題を解決するために、本発明に係る画像処理装置は、2次元の撮像画像を順次取得する画像取得部と、順次取得された前記撮像画像の各ピクセルに、該ピクセルにより表されたオブジェクトの種類を表すラベルを付すオブジェクト種類認識実行部と、順次取得された前記撮像画像に基づいて、前記撮像画像の各ピクセルの3次元位置認識を実行して3次元マップを作成する3次元マップ作成部と、を含み、前記3次元マップ作成部は、前記撮像画像の各ピクセルの3次元位置認識を、該ピクセルに付されたラベルに応じて制限する、ことを特徴とする。
- [0007] 本発明の一態様では、前記3次元マップ作成部は、ピクセルに付されたラベルが、移動するオブジェクトの種類を示す場合に、該ピクセルの3次元位置認識を制限する、ことを特徴とする。
- [0008] 本発明の一態様では、前記3次元マップ作成部は、前記撮像画像の各ピクセルに付されたラベルに基づいて、前記撮像画像を複数の領域に分割し、領域ごとに該領域に含まれるピクセルの3次元位置認識を実行する、ことを特徴とする。
- [0009] 本発明の一態様では、前記3次元マップ作成部は、領域ごとの3次元位置認識結果を合成することで前記3次元マップを作成する、ことを特徴とする。
- [0010] 本発明の一態様では、前記撮像画像の撮像位置を視点として前記3次元マップをレンダリングすることにより2次元画像を作成するレンダリング部と、前記2次元画像に対して、前記撮像画像のピクセルのうち、3次元位置認識が制限されたピクセルを合成する合成部と、をさらに有することを特徴とする。
- [0011] 本発明の一態様では、前記3次元マップ作成部は、さらに、前記ラベルに応じて予め設定された重み係数を用いて、前記3次元位置認識を制限することを特徴とする。

[0012] 本発明に係る画像処理方法は、２次元の撮像画像を順次取得する画像取得ステップと、順次取得された前記撮像画像の各ピクセルに、該ピクセルにより表されたオブジェクトの種類を表すラベルを付すオブジェクト種類認識実行ステップと、順次取得された前記撮像画像に基づいて、前記撮像画像の各ピクセルの３次元位置認識を実行して３次元マップを作成する３次元マップ作成ステップと、を含み、前記３次元マップ作成ステップにおいて、前記撮像画像の各ピクセルの３次元位置認識は、該ピクセルに付されたラベルに応じて制限される、ことを特徴とする。

[0013] 本発明に係るプログラムは、２次元の撮像画像を順次取得する画像取得手順と、順次取得された前記撮像画像の各ピクセルに、該ピクセルにより表されたオブジェクトの種類を表すラベルを付すオブジェクト種類認識実行手順と、順次取得された前記撮像画像に基づいて、前記撮像画像の各ピクセルの３次元位置認識を実行して３次元マップを作成する３次元マップ作成手順と、をコンピュータに実行させるプログラムであって、前記３次元マップ作成手順において、前記撮像画像の各ピクセルの３次元位置認識を、該ピクセルに付されたラベルに応じて制限する、ことを特徴とする。

図面の簡単な説明

[0014] [図1]本発明の一実施形態に係る画像処理装置の構成図である。

[図2]本発明の一実施形態に係る画像処理装置で実装される機能の一例を示す機能ブロック図である。

[図3]本発明の一実施形態に係る画像処理装置で行われる画像処理の流れの一例を示す図である。

発明を実施するための形態

[0015] 以下、本発明の一実施形態について図面に基づき詳細に説明する。図１は、本実施形態に係る画像処理装置１０の構成図である。

[0016] 本実施形態に係る画像処理装置１０は、例えば、ゲームコンソールやパーソナルコンピュータなどのコンピュータである。また、画像処理装置１０は、デジタルカメラ等の撮像装置であってもよい。図１に示すように、本実施

形態に係る画像処理装置１０は、例えば、プロセッサ１２、記憶部１４、操作部１６、表示部１８を含んでいる。

[0017] プロセッサ１２は、例えば画像処理装置１０にインストールされるプログラムに従って動作するＣＰＵ等のプログラム制御デバイスである。

[0018] 記憶部１４は、ＲＯＭやＲＡＭ等の記憶素子やハードディスクドライブなどである。記憶部１４には、プロセッサ１２によって実行されるプログラムなどが記憶される。

[0019] 操作部１６は、キーボード、マウス、ゲームコンソールのコントローラ等のユーザインタフェースであって、ユーザの操作入力を受け付けて、その内容を示す信号をプロセッサ１２に出力する。

[0020] 表示部１８は、液晶ディスプレイ等の表示デバイスであって、プロセッサ１２の指示に従って各種の画像を表示する。

[0021] なお、画像処理装置１０は、ネットワークボードなどの通信インタフェース、ＤＶＤ－ＲＯＭやＢｌｕ－ｒａｙ（登録商標）ディスクなどの光ディスクを読み取る光ディスクドライブ、ＵＳＢ（Universal Serial Bus）ポートなどを含んでいてもよい。

[0022] 続いて、本実施形態に係る画像処理装置１０の機能、及び、画像処理装置１０で実行される処理についてさらに説明する。

[0023] 図２に示すように、本実施形態に係る画像処理装置１０には、機能的には例えば、画像取得部２０２、オブジェクト種類認識実行部２０４、３次元マップ作成部２０６、レンダリング部２０８、合成部２１０、が含まれる。以上の要素はプロセッサ１２、及び、記憶部１４を主として実装される。

[0024] 画像取得部２０２は、２次元の撮像画像を順次取得する。具体的には、例えば、画像取得部２０２は、一般的なビデオカメラで撮影された動画像に含まれる各フレームの静止画像を順次取得する。ここで、静止画像は、例えば赤、緑、及び、青の各色の階調情報を含む２次元画像である。また、入力画像を撮影するビデオカメラは、時間に応じて位置及び方向が変化する。従って、画像取得部２０２は、時間によって連続的に視点及び画角が変化しなが

ら撮影された２次元画像を取得する。

[0025] なお、画像取得部２０２は、撮像画像とともに深度情報を取得してもよい。具体的には、例えば、画像取得部２０２は、ビデオカメラのレンズ近傍に配置された測距センサから、各ピクセルに表されたオブジェクトとレンズの距離を、当該入力画像の各ピクセルに対応する深度情報として取得してもよい。また、例えば、２眼のビデオカメラが撮像した２枚の撮像画像に基づいて、画像取得部２０２は、赤、緑、及び、青の各色の階調情報とともに、各ピクセルの深度情報を含む２次元画像を取得してもよい。

[0026] オブジェクト種類認識実行部２０４は、順次取得された撮像画像の各ピクセルに、該ピクセルにより表されたオブジェクトの種類を表すラベルを付す。具体的には、例えば、オブジェクト種類認識実行部２０４は、畳み込みニューラルネットワーク（CNN）により実装された機械学習モデルである。オブジェクト種類認識実行部２０４は、画像取得部２０２が１フレームの入力画像を取得するたびに、入力画像に含まれる階調情報に基づいて、当該入力画像に対するセマンティックセグメンテーションを実行する。これにより、フレームごとに、取得された入力画像の各ピクセルに対して、該ピクセルにより表されたオブジェクトの種類を表すラベルが付される。ラベルは、例えば、人物を表す「human」、テーブルを表す「table」、ソファを表す「sofa」、コップを表す「cup」、床を表す「floor」、壁を表す「wall」等である。

[0027] ３次元マップ作成部２０６は、順次取得された撮像画像に基づいて、撮像画像の各ピクセルの３次元位置認識を実行して３次元マップを作成する。具体的には、例えば、３次元マップ作成部２０６は、画像取得部２０２が順次取得した入力画像に基づいて、ビジュアルスラム(Visual SLAM)を実行する。３次元マップ作成部２０６は、カメラで撮影された２次元画像からカメラが撮影した環境の３次元情報とカメラの位置姿勢を同時に推定することができる。これにより、３次元マップ作成部２０６は、３次元マップを作成できる。

[0028] なお、3次元マップ作成部206は、GPS（図示なし）から取得した位置情報と、画像取得部202が取得した深度情報と、を用いて3次元位置認識を行うことで、3次元マップを作成してもよい。

[0029] ここで、3次元マップ作成部206は、撮像画像の各ピクセルの3次元位置認識を、該ピクセルに付されたラベルに応じて制限する。具体的には、3次元マップ作成部206は、ピクセルに付されたラベルが、移動するオブジェクトの種類を示す場合に、該ピクセルの3次元位置認識を制限する。例えば、1枚の撮像画像に含まれる各ピクセルに対して、人物を表す「human」、コップを表す「cup」、テレビを表す「TV」、テーブルを表す「table」、ソファを表す「sofa」、床を表す「floor」、または、壁を表す「wall」のいずれかが付されているとする。この場合、3次元マップ作成部206は、移動するオブジェクトの種類を示す「human」というラベルが付されたピクセルの3次元位置認識を制限する。すなわち、3次元マップ作成部206は、取得された撮像画像に含まれる複数のピクセルのうち、「cup」、「TV」、「table」、「sofa」、「floor」、及び、「wall」というラベルが付されたピクセルのみを用いて、3次元マップを作成する。これにより、実空間において、移動する対象である人物が存在する位置以外の位置と対応するボクセルによって構成される3次元マップが作成される。人物が存在する位置に対応するボクセルは、当該3次元マップに含まれない。

[0030] なお、上記において、移動するオブジェクトの種類が「human」のみである場合について説明したが、移動するオブジェクトの種類は、適宜設定されてよい。具体的には、例えば、コップは人物が頻繁に移動させる対象であるため、移動するオブジェクトの種類として、「human」及び「cup」が設定されてよい。また、例えば、床及び壁以外の対象はすべて移動させることができるため、移動するオブジェクトの種類として、「human」、「cup」、「TV」、「table」及び「sofa」が設定されてもよい。

[0031] また、3次元マップ作成部206は、撮像画像の各ピクセルに付されたラベルに基づいて、撮像画像を複数の領域に分割し、領域ごとに該領域に含ま

れるピクセルの3次元位置認識を実行してもよい。具体的には上記例では、3次元マップ作成部206は、撮像画像を「human」というラベルが付された領域と、「cup」、「TV」、「table」、「sofa」、「floor」、及び、「wall」というラベルが付された領域に分割する。そして、3次元マップ作成部206は、「cup」、「TV」、「table」、「sofa」、「floor」、及び、「wall」というラベルが付されたピクセルのみを用いて3次元位置認識を実行する。また、3次元マップ作成部206は、「human」というラベルが付されたピクセルのみを用いて3次元位置認識を実行する。そして、3次元マップ作成部206は、領域ごとの3次元位置認識結果を合成することで3次元マップを作成してもよい。これにより、人物が存在する位置に対応するボクセルを含む3次元マップを作成できる。

[0032] なお、3次元マップ作成部206は、人物が動かない状況下でのみ、領域ごとの3次元位置認識結果を合成することで3次元マップを作成することができる。そのため、人物が動いた時点で、3次元マップ作成部206は、領域ごとの3次元位置認識結果の合成を停止し、移動するオブジェクトの種類を示すラベル以外のラベルが付されたピクセルのみを用いて、3次元マップを作成する。そして、人物の動きが停止した時点で、移動するオブジェクトの種類を示すラベルが付されたピクセルに対する3次元位置認識を再開し、再び領域ごとの3次元位置認識結果が合成された3次元マップを作成することができる。

[0033] レンダリング部208は、撮像画像の撮像位置を視点として3次元マップをレンダリングすることにより2次元画像を作成する。具体的には、例えば、画像取得部202は、2次元の撮像画像をフレームごとに順次取得する。このとき、ビデオカメラ等の撮像装置は時間とともに移動するため、撮像時の視点はフレームごとに変化する。レンダリング部208は、現フレームにおける撮像画像の撮像位置を視点として、3次元マップに対してレンダリングを行う。

[0034] 合成部210は、レンダリング部208の作成した2次元画像に対して、

撮像画像のピクセルのうち、３次元位置認識が制限されたピクセルを合成する。具体的には、上記例のように、レンダリング部２０８は、「cup」、「TV」、「table」、「sofa」、「floor」、及び、「wall」というラベルが付されたピクセルのみを用いて作成された３次元マップに対して、現フレームにおける撮像画像の撮像位置を視点として、２次元画像を作成したとする。当該３次元マップには、「human」というラベルが付されたピクセルと対応するボクセルが含まれないため、作成された２次元画像は、人物が表された領域に階調情報を含まない。一方、画像取得部２０２が取得した撮像画像の全てのピクセルは階調情報を含む。そのため、撮像画像のピクセルのうち、人物を表すピクセルは階調情報を含む。合成部２１０は、人物が表された領域に階調情報を含まない２次元画像に対して、現フレームの撮像画像のピクセルのうち、人物を表すラベルが付されたピクセルの階調情報を合成する。これにより、精度よく３次元マップを作成しつつ、３次元マップから作成された２次元画像と、３次元位置認識が制限されたために、３次元マップに存在しない移動するオブジェクトと、を重ねて表示することができる。

[0035] なお、３次元マップ作成部２０６が撮像画像を複数の領域に分割し、領域ごとに該領域に含まれるピクセルの３次元位置認識を実行する場合には、３次元マップからレンダリングによって作成された２次元画像は、移動するオブジェクトの種類を示すラベルが付与されたオブジェクトと、それ以外のラベルが付与されたオブジェクトと、どちらのオブジェクトも含む。この場合、画像処理装置１０は、合成部２１０を含まない構成であってもよい。

[0036] 以上の機能は、コンピュータである画像処理装置１０にインストールされた、以上の機能に対応する指令を含むプログラムをプロセッサ１２で実行することにより実装されてもよい。このプログラムは、例えば、光ディスク、磁気ディスク、磁気テープ、光磁気ディスク、フラッシュメモリ等のコンピュータ読み取り可能な情報記憶媒体を介して、あるいは、インターネットなどを介して画像処理装置１０に供給されてもよい。

[0037] 本実施形態では例えば、一般的なビデオカメラで撮影される２次元画像に

基づいて、3次元マップが作成される。以下、3次元マップの作成方法について図3を参照しながら説明する。

[0038] まず、画像取得部202によってフレームごとに順次取得された2次元の入力画像は、オブジェクト種類認識実行部204に入力される。具体的には、例えば、2眼のビデオカメラによって撮影された2次元の入力画像は、フレームごとに順次オブジェクト認識実行部に入力される。ここで、例えば、入力画像の各ピクセルは、赤、緑及び青を表す階調情報と、ビデオカメラからの距離を表す深度情報と、を含む。

[0039] オブジェクト種類認識実行部204は、フレームごとに取得した入力画像の各ピクセルに、該ピクセルにより表されたオブジェクトの種類を表すラベルを付す。具体的には、オブジェクト種類認識実行部204は、入力された2次元の入力画像の各ピクセルの階調情報に基づいて、該ピクセルにより表されたオブジェクトの種類を判定する。そして、オブジェクト種類認識実行部204は、入力画像の各ピクセルに対して、判定されたオブジェクトの種類を表すラベルを付す。例えば、オブジェクト種類認識実行部204は、あるピクセルが表すオブジェクトが人物であると判定した場合、当該ピクセルに対して「human」というラベルを付す。また、例えば、オブジェクト種類認識実行部204は、あるピクセルが表すオブジェクトが壁面であると判定した場合、当該ピクセルに対して「wall」というラベルを付す。従って、オブジェクト種類認識実行部204が出力した画像の各ピクセルは、階調情報及び深度情報に加えて、ラベルを表す情報を含む。

[0040] 次に、出力された画像は、フレームごとに順次3次元マップ作成部206に入力され、3次元マップが作成される。具体的には、上記のように、各ピクセルが階調情報、深度情報、及び、ラベルを表す情報を含む2次元の画像は、3次元マップ作成部206に入力される。3次元マップ作成部206は、順次入力された入力画像の階調情報及び深度情報に基づいて、入力画像に表されたオブジェクトの3次元位置認識を実行して3次元マップを作成する。

- [0041] ここで、3次元マップ作成部206は、撮像画像の各ピクセルの3次元位置認識を、該ピクセルに付されたラベルに応じて制限する。図3の例では、3次元マップ作成部206は、制限の対象となるラベルとして、「human」というラベルを取得する。そして、3次元マップ作成部206は、各ラベルが付された画像を、「human」というラベルが付された領域と、「cup」、「TV」、「table」、「sofa」、「floor」、及び、「wall」というラベルが付された領域と、に分割する（図3の3次元マップ作成部206の下側参照）。そして、3次元マップ作成部206は、「cup」、「TV」、「table」、「sofa」、「floor」、及び、「wall」というラベルが付されたピクセルのみを用いて3次元位置認識を実行する。
- [0042] 静止しているオブジェクトについてのみスラムを実行することにより、3次元マップ作成部206は、入力画像に移動するオブジェクトが含まれる場合であっても、精度よく3次元マップを形成することができる。また、時間が経過するごとに、入力画像中の移動するオブジェクトが位置を変更するとともに、画像取得部202が異なる視点で撮像された入力画像を取得する。これにより、3次元マップ作成部206は、図3の3次元マップにおいて、移動するオブジェクトの奥にあったために欠けてしまった領域のボクセルを含む3次元マップを作成できる。
- [0043] 上記のように作成された3次元マップは、例えば、拡張現実（AR：Augmented Reality）技術に適用することができる。テーブル、テレビ、ソファ等が配置された部屋の内部の3次元マップを作成した場合、ユーザは、現実世界の部屋にしながら、現実世界の部屋が反映された仮想空間を体験できる。この時、撮影のために部屋から人物を退去させたり、コップなどの小物や動く物体を片づけたりする必要がなく、ユーザはより手軽に高精度な3次元マップを作成することができる。
- [0044] なお、本発明は上述の実施形態に限定されるものではない。また、上記の具体的な文字列や数値及び図面中の具体的な文字列や数値は例示であり、これらの文字列や数値には限定されない。

- [0045] 例えば、3次元マップ作成部206は、さらに、ラベルに応じて予め設定された重み係数を用いて、3次元位置認識を制限してもよい。具体的には、 n フレーム目の入力画像が入力されたときに、3次元マップ作成部206は、 $n-1$ フレームで作成された3次元マップと、 n フレーム目の入力画像と、からそれぞれ特徴点を抽出する。3次元マップ作成部206は、当該特徴点の誤差の総和が最小となるように、入力画像の位置認識を行う。
- [0046] ここで、3次元マップ作成部206は、特徴点ごとに、当該特徴点の誤差に対して重み付け係数を乗算してもよい。乗算される重み付け係数は、特徴点が現れたピクセルに付されたラベルに対して予め設定される。
- [0047] 予め設定される重みづけ係数は、ラベルが示すオブジェクトの種類が移動しやすいほど小さい値が設定される。具体的には、例えば、「floor」及び「wall」というラベルに対して、「1.0」という重み係数が設定される。「table」、「TV」及び「sofa」というラベルに対して、「0.5」という重み係数が設定される。「cup」というラベルに対して、「0.1」という重み係数が設定される。「human」というラベルに対して、「0.0」という重み係数が設定される。
- [0048] この場合、「0.0」という重み係数が乗算される特徴点の誤差は「0.0」であるため、3次元マップ作成部206が位置認識する際に、当該特徴点の誤差は計算対象から除外される。「0.1」、「0.5」及び「1.0」という重み係数が乗算される特徴点の誤差は、当該係数に応じて誤差の総和に対する影響の大きさが異なる。これにより、オブジェクトの移動する可能性の高さに応じて、特徴点ごとに3次元位置認識に与える影響の大きさを変化させることができる。
- [0049] 上記重み付け係数は、撮影環境や使用用途に応じて、適宜設定される。従って、移動する可能性は低いがゼロではないオブジェクト（上記例では「table」、「TV」及び「sofa」が表すオブジェクト）が3次元位置認識の精度に与える悪影響を軽減することができる。一方、移動する可能性が高いオブジェクト（上記例では「cup」が表すオブジェクト）も必要に応じて特徴点とし

て利用することが可能となる。

請求の範囲

- [請求項1] 2次元の撮像画像を順次取得する画像取得部と、
 順次取得された前記撮像画像の各ピクセルに、該ピクセルにより表
 されたオブジェクトの種類を表すラベルを付すオブジェクト種類認識
 実行部と、
 順次取得された前記撮像画像に基づいて、前記撮像画像の各ピクセル
 の3次元位置認識を実行して3次元マップを作成する3次元マップ
 作成部と、
 を含み、
 前記3次元マップ作成部は、前記撮像画像の各ピクセルの3次元位
 置認識を、該ピクセルに付されたラベルに応じて制限する、ことを特
 徴とする画像処理装置。
- [請求項2] 前記3次元マップ作成部は、ピクセルに付されたラベルが、移動す
 るオブジェクトの種類を示す場合に、該ピクセルの3次元位置認識を
 制限する、ことを特徴とする請求項1に記載の画像処理装置。
- [請求項3] 前記3次元マップ作成部は、前記撮像画像の各ピクセルに付された
 ラベルに基づいて、前記撮像画像を複数の領域に分割し、領域ごとに
 該領域に含まれるピクセルの3次元位置認識を実行する、ことを特徴
 とする請求項1または2に記載の画像処理装置。
- [請求項4] 前記3次元マップ作成部は、領域ごとの3次元位置認識結果を合成
 することで前記3次元マップを作成する、ことを特徴とする請求項3
 に記載の画像処理装置。
- [請求項5] 前記撮像画像の撮像位置を視点として前記3次元マップをレンダリ
 ングすることにより2次元画像を作成するレンダリング部と、
 前記2次元画像に対して、前記撮像画像のピクセルのうち、3次元
 位置認識が制限されたピクセルを合成する合成部と、
 をさらに有することを特徴とする請求項1に記載の画像処理装置。
- [請求項6] 前記3次元マップ作成部は、さらに、前記ラベルに応じて予め設定

された重み係数を用いて、前記３次元位置認識を制限することを特徴とする請求項１に記載の画像処理装置。

[請求項7]

２次元の撮像画像を順次取得する画像取得ステップと、

順次取得された前記撮像画像の各ピクセルに、該ピクセルにより表されたオブジェクトの種類を表すラベルを付すオブジェクト種類認識実行ステップと、

順次取得された前記撮像画像に基づいて、前記撮像画像の各ピクセルの３次元位置認識を実行して３次元マップを作成する３次元マップ作成ステップと、

を含み、

前記３次元マップ作成ステップにおいて、前記撮像画像の各ピクセルの３次元位置認識は、該ピクセルに付されたラベルに応じて制限される、ことを特徴とする画像処理方法。

[請求項8]

２次元の撮像画像を順次取得する画像取得手順と、

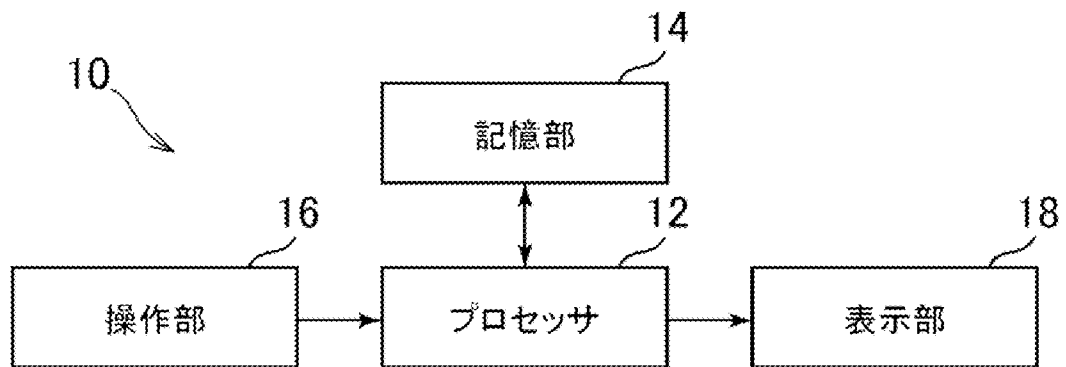
順次取得された前記撮像画像の各ピクセルに、該ピクセルにより表されたオブジェクトの種類を表すラベルを付すオブジェクト種類認識実行手順と、

順次取得された前記撮像画像に基づいて、前記撮像画像の各ピクセルの３次元位置認識を実行して３次元マップを作成する３次元マップ作成手順と、

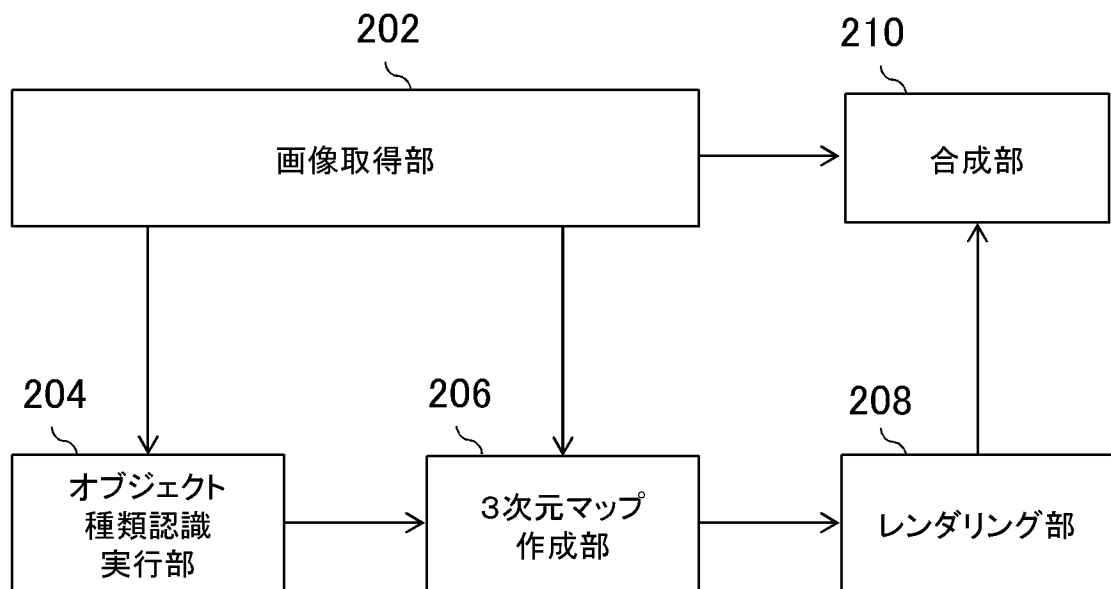
をコンピュータに実行させるプログラムであって、

前記３次元マップ作成手順において、前記撮像画像の各ピクセルの３次元位置認識を、該ピクセルに付されたラベルに応じて制限する、ことを特徴とするプログラム。

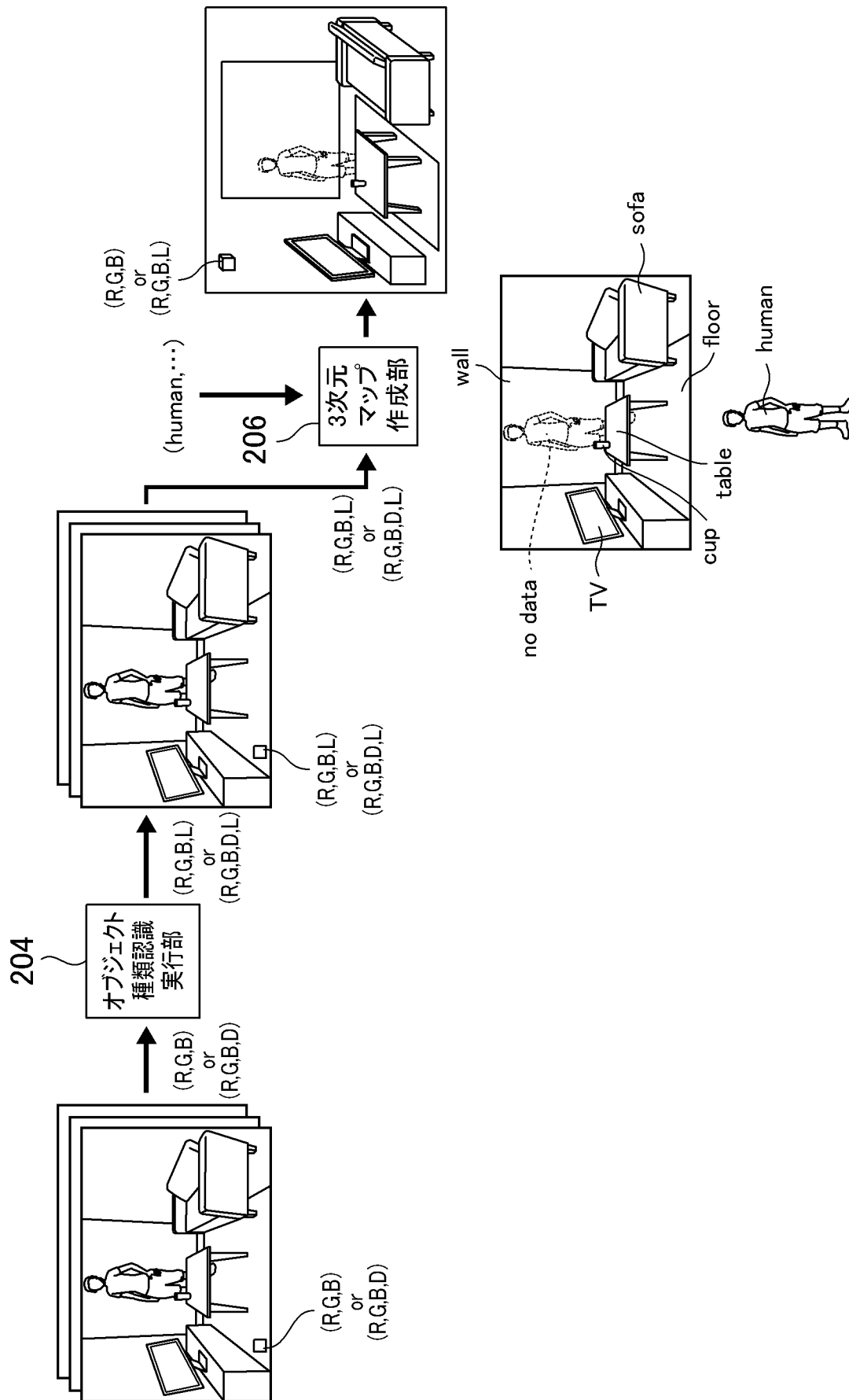
[図1]



[図2]



[図3]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2019/035972

A. CLASSIFICATION OF SUBJECT MATTER

Int. Cl. G06T7/70 (2017.01) i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

Int. Cl. G06T7/70

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Published examined utility model applications of Japan 1922-1996

Published unexamined utility model applications of Japan 1971-2019

Registered utility model specifications of Japan 1996-2019

Published registered utility model applications of Japan 1994-2019

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y A	JP 2019-28861 A (SONY CORP.) 21 February 2019, paragraphs [0081]-[0094], [0120]-[0154], [0160]-[0169], fig. 4, 5 & WO 2019/026832 A1	1-4, 6-8 5
Y	TATENO, Keisuke et al., CNN-SLAM: Real-time dense monocular SLAM with learned depth prediction, 2017 IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 6565-6574	1-4, 6-8
A	WO 2015/111283 A1 (SONY CORP.) 30 July 2015, paragraphs [0065]-[0071] & US 2016/0379413 A1, paragraphs [0082]-[0088] & EP 3098689 A1 & CN 105992986 A & KR 10-2016-0111904 A	5



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance

“E” earlier application or patent but published on or after the international filing date

“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

“O” document referring to an oral disclosure, use, exhibition or other means

“P” document published prior to the international filing date but later than the priority date claimed

“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

“&” document member of the same patent family

Date of the actual completion of the international search
07.10.2019

Date of mailing of the international search report
15.10.2019

Name and mailing address of the ISA/
Japan Patent Office
3-4-3, Kasumigaseki, Chiyoda-ku,
Tokyo 100-8915, Japan

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORTInternational application No.
PCT/JP2019/035972

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	JP 6-18221 A (DIRECTOR GENERAL, AGENCY OF INDUSTRIAL SCIENCE AND TECHNOLOGY) 25 January 1994, paragraphs [0065]-[0071] (Family: none)	6
A	MARTIN, Runz, et al., Co-Fusion: Real-time Segmentation, Tracking and Fusion of Multiple Objects, 2017 IEEE International Conference on Robotics and Automation, 2017, pp. 4471-4478	1-8

A. 発明の属する分野の分類（国際特許分類（IPC））

Int.Cl. G06T7/70(2017.01)i

B. 調査を行った分野

調査を行った最小限資料（国際特許分類（IPC））

Int.Cl. G06T7/70

最小限資料以外の資料で調査を行った分野に含まれるもの

日本国実用新案公報	1922-1996年
日本国公開実用新案公報	1971-2019年
日本国実用新案登録公報	1996-2019年
日本国登録実用新案公報	1994-2019年

国際調査で使用した電子データベース（データベースの名称、調査に使用した用語）

C. 関連すると認められる文献

引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求項の番号
Y A	JP 2019-28861 A（ソニー株式会社） 2019.02.21, 段落[0081]-[0094], [0120]-[0154], [0160]-[0169], 図 4-5 & WO 2019/026832 A1	1-4, 6-8 5
Y	TATEN0, Keisuke et al., CNN-SLAM: Real-time dense monocular SLAM with learned depth prediction, 2017 IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 6565-6574	1-4, 6-8

C欄の続きにも文献が列挙されている。

パテントファミリーに関する別紙を参照。

* 引用文献のカテゴリー

「A」特に関連のある文献ではなく、一般的技術水準を示すもの

「E」国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの

「L」優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献（理由を付す）

「O」口頭による開示、使用、展示等に言及する文献

「P」国際出願日前で、かつ優先権の主張の基礎となる出願

の日の後に公表された文献

「T」国際出願日又は優先日後に公表された文献であって出願と矛盾するものではなく、発明の原理又は理論の理解のために引用するもの

「X」特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの

「Y」特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの

「&」同一パテントファミリー文献

国際調査を完了した日

07.10.2019

国際調査報告の発送日

15.10.2019

国際調査機関の名称及びあて先

日本国特許庁（ISA/J P）

郵便番号100-8915

東京都千代田区霞が関三丁目4番3号

特許庁審査官（権限のある職員）

新井 則和

5H

7889

電話番号 03-3581-1101 内線 3531

C (続き) . 関連すると認められる文献		
引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求項の番号
A	WO 2015/111283 A1 (ソニー株式会社) 2015.07.30, 段落[0065]-[0071] & US 2016/0379413 A1, 段落[0082]-[0088] & EP 3098689 A1 & CN 105992986 A & KR 10-2016-0111904 A	5
Y	JP 6-18221 A (工業技術院長) 1994.01.25, 段落[0065]-[0071] (ファミリーなし)	6
A	MARTIN Runz, et al., Co-Fusion: Real-time Segmentation, Tracking and Fusion of Multiple Objects, 2017 IEEE International Conference on Robotics and Automation, 2017, pp. 4471-4478	1-8