

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7379065号
(P7379065)

(45)発行日 令和5年11月14日(2023.11.14)

(24)登録日 令和5年11月6日(2023.11.6)

(51)国際特許分類

G 0 6 T	7/70 (2017.01)	G 0 6 T	7/70	Z
G 0 6 T	7/246(2017.01)	G 0 6 T	7/246	

F I

請求項の数 12 (全17頁)

(21)出願番号 特願2019-185352(P2019-185352)
 (22)出願日 令和1年10月8日(2019.10.8)
 (65)公開番号 特開2021-60868(P2021-60868A)
 (43)公開日 令和3年4月15日(2021.4.15)
 審査請求日 令和4年10月5日(2022.10.5)

(73)特許権者 000001007
 キヤノン株式会社
 東京都大田区下丸子3丁目30番2号
 (74)代理人 100090273
 弁理士 國分 孝悦
 林 将之
 東京都大田区下丸子3丁目30番2号
 (72)発明者 キヤノン株式会社内
 審査官 德 田 賢二

最終頁に続く

(54)【発明の名称】 情報処理装置、情報処理方法、およびプログラム

(57)【特許請求の範囲】**【請求項1】**

追跡対象物を含む現実空間を撮影した画像から複数の特徴情報を抽出する特徴抽出手段と、

前記現実空間を撮影した画像から前記追跡対象物に付されている指標を検出し、該指標の位置姿勢を推定する指標推定手段と、

前記指標の位置姿勢と追跡対象モデルとに基づいて、前記追跡対象物の位置姿勢を推定する対象物推定手段と、

前記現実空間を前記撮影したカメラの位置姿勢および、前記追跡対象物の位置姿勢に基づいて、前記複数の特徴情報を分類する分類手段と、

前記分類手段が前記追跡対象物に属していると判定した特徴情報を、前記追跡対象モデルに追加するモデル構築手段と、

を有することを特徴とする情報処理装置。

【請求項2】

前記現実空間を撮影した前記画像は、奥行きの情報を含む画像であり、

前記複数の特徴情報は、前記現実空間の画像中の複数の特徴に関して、前記カメラの座標系で表される第一の3次元座標を含み、

前記分類手段は、

前記複数の特徴情報のうちで前記判定の対象となる特徴情報に対し、前記第一の3次元座標を、前記指標の位置姿勢を基準とする3次元座標系に投影した第二の3次元座標を算

出し、

さらに、異なる時刻に撮影された1つ以上の前記現実空間の画像において前記第二の3次元座標が一定であった場合、前記判定の対象となる特徴情報は前記追跡対象物に属していると判定することを特徴とする請求項1に記載の情報処理装置。

【請求項3】

前記奥行きの情報は、ステレオカメラまたはデプスカメラによる計測、もしくは、1台の可視光カメラを構成した上で機械学習された学習モデルによる奥行き推定により求められることを特徴とする請求項2に記載の情報処理装置。

【請求項4】

前記複数の特徴情報は、前記現実空間の画像中の複数の特徴に関する画像座標および、
前記カメラの座標系で表される第一の3次元座標を含み、

前記分類手段は、

前記複数の特徴情報のうちで前記判定の対象となる特徴情報に対し、前記第二の3次元座標を前記現実空間の画像上に投影した座標と前記画像座標との誤差である再投影誤差を算出し、

さらに、異なる時刻に撮影された1つ以上の前記現実空間の画像において前記再投影誤差が閾値以内であった場合、前記判定の対象となる特徴情報は前記追跡対象物に属していると判定することを特徴とする請求項1に記載の情報処理装置。

【請求項5】

前記複数の特徴情報は、前記現実空間の画像中の複数の特徴に関する第一の光線を含み、

前記第一の光線は、前記現実空間を前記撮影したカメラの光学中心から、前記現実空間の特徴の画像座標を通る前記カメラの座標系における光線であり、

前記分類手段は、

前記複数の特徴情報のうちで前記判定の対象となる特徴情報に対し、前記第一の光線を前記指標の位置姿勢を基準とする3次元座標系に投影した第二の光線を算出し、

さらに、異なる時刻に撮影された1つ以上の前記現実空間の画像において前記第二の光線が交差した場合、前記判定の対象となる特徴情報は前記追跡対象物に属していると判定し、

前記モデル構築手段は、前記交差した点の3次元座標を前記追跡対象モデルに蓄積することを特徴とする請求項1に記載の情報処理装置。

【請求項6】

前記分類手段は、前記現実空間の画像のうち、前記追跡対象物ではないと判定される領域をマスク領域として算出し、前記複数の特徴情報のうちで前記マスク領域内の特徴は前記追跡対象物に属していると判定しないことを特徴とする請求項1から5のいずれか1項に記載の情報処理装置。

【請求項7】

前記分類手段は、異なる時刻に撮影された1つ以上の前記現実空間の画像において、前記第二の3次元座標が一定であると判定された画像数に基づいて、前記特徴が前記追跡対象物に属している確からしさを算出して、前記特徴情報に含めることを特徴とする請求項2または3に記載の情報処理装置。

【請求項8】

前記分類手段は、異なる時刻に撮影された1つ以上の前記現実空間の画像において、前記再投影誤差が閾値以内であった画像数に基づいて、前記特徴が前記追跡対象物に属している確からしさを算出して、該確からしさに基づいて前記特徴情報に含めることを特徴とする請求項4に記載の情報処理装置。

【請求項9】

前記分類手段は、異なる時刻に撮影された1つ以上の前記現実空間の画像において、前記第二の光線が同じ点で交差した画像数に基づいて、前記特徴が前記追跡対象物に属している確からしさを算出して、該確からしさに基づいて前記特徴情報に含めることを特徴とする請求項5に記載の情報処理装置。

【請求項10】

10

20

30

40

50

前記指標推定手段は、前記追跡対象物の一部をモデル化した情報を、前記追跡対象物に設けられている前記指標として用いることを特徴とする請求項 1 から 9 のいずれか 1 項に記載の情報処理装置。

【請求項 1 1】

情報処理装置が実行する情報処理方法であって、

追跡対象物を含む現実空間を撮影した画像から複数の特徴情報を抽出する特徴抽出工程と、

前記現実空間を撮影した画像から前記追跡対象物に付されている指標を検出し、該指標の位置姿勢を推定する指標推定工程と、

前記指標の位置姿勢と追跡対象モデルとに基づいて、前記追跡対象物の位置姿勢を推定する対象物推定工程と、

前記現実空間を前記撮影したカメラの位置姿勢および、前記追跡対象物の位置姿勢に基づいて、前記複数の特徴情報を分類する分類工程と、

前記分類工程で前記追跡対象物に属していると判定した特徴情報を、前記追跡対象モデルに追加するモデル構築工程と、

を有することを特徴とする情報処理方法。

【請求項 1 2】

コンピュータを請求項 1 から 10 のいずれか 1 項に記載の情報処理装置の各手段として機能させるためのプログラム。

【発明の詳細な説明】

【技術分野】

【0 0 0 1】

本発明は、撮像された画像と物体との位置合わせを行う情報処理技術に関する。

【背景技術】

【0 0 0 2】

近年、設計・製造分野において、プロトタイプを用いた評価の期間短縮、費用削減が求められている。例えば、C A D (コンピュータ支援設計) システムで作成した設計 (形状・デザイン) データを用いて、組み立てやすさやメンテナンス性の評価をするための複合現実感 (M R : M i x e d R e a l i t y) システムが導入されている。

【0 0 0 3】

M R システムの画像表示装置には、ビデオシースルーワー方式と光学シースルーワー方式がある。ビデオシースルーワー方式の画像表示装置は、撮像装置によって撮影された現実空間の画像に、該撮像装置の位置及び姿勢に応じて生成された仮想空間 (コンピュータグラフィックスにより描画された仮想物体や文字情報等) の画像を重畳描画した合成画像を表示する。光学シースルーワー方式の画像表示装置は、観察者の頭部に装着された光学シースルーワー型ディスプレイに、観察者の視点の位置及び姿勢に応じて生成された仮想空間の画像を表示する。

【0 0 0 4】

M R 技術において最も重要な課題の一つは、現実空間と仮想空間の間の位置合わせをいかに正確に行うかということであり、従来から、多くの取り組みが行われている。M R における位置合わせの問題には、シーンを撮影するカメラの位置及び姿勢を求める問題と、C G モデルを重畳表示したい対象物 (追跡対象物) の位置及び姿勢を求める問題とが含まれる。

【0 0 0 5】

例えば、機械のメンテナンス性を評価するためのM R システムでは、メンテナンス対象のC G モデルを空間中の固定された位置に表示するとともに、ユーザーが手を持つ実物体の工具にも工具C G モデルを配置するというケースがある。これを実現するためには、シーン中に規定された基準座標系に対するカメラ座標系の位置姿勢と、カメラ座標系に対する工具座標系の位置姿勢とをリアルタイムに求める必要がある。また、実物体の工具に対して工具C G モデルがずれないようにするために、実物体の工具と工具座標系の位置関係は、予め決められた位置関係が再現されなければならない。

10

20

30

40

50

【0006】

カメラの位置及び姿勢を求める問題を解決する方法の一つとして、現実空間のマップを構築し、そのマップに基づいて計測対象物の位置及び姿勢を求める方法がある。その一例として、Visual SLAM (Simultaneous Localization and Mapping) と呼ばれる手法がある（非特許文献1参照）。Visual SLAMは、カメラで撮影された画像から現実空間におけるカメラの位置姿勢を求めるときに同時に、必要に応じてその画像からキーフレームの情報を抽出し、その情報を蓄積したマップを構築する手法である。キーフレームの情報には、画像と、その画像を撮影したカメラの位置姿勢、およびその画像から抽出された特徴点の3次元情報が含まれる。

【0007】

またシーン中に規定された基準座標系に対するカメラ座標系の位置姿勢を正確に推定するためには、シーンに固定された（すなわち3次元座標が変わらない）特徴点に基づくことが望ましい。そのため、ユーザーが手に持つ工具のように、シーン中を動く物体から検出された特徴点の情報は除去することが一般的であり、動く物体である追跡対象物の位置及び姿勢を求める問題へ応用するには、別 の方法が必要となる。

10

【0008】

一方、CGモデルを重畳表示したい対象物（追跡対象物）の位置及び姿勢を求める問題を解決する方法の一つとして、モデルベーストラッキングと呼ばれる手法がある（非特許文献2参照）。モデルベーストラッキングは、事前に作成した追跡対象物の3次元モデルを用いて、カメラで撮影した画像中の追跡対象物の位置姿勢を求める手法である。その具体的方法の一つとして、カメラで撮影した画像からエッジ情報を抽出し、CADや3Dスキャナーを用いて予め作成した追跡対象物の形状モデルのエッジ情報をあてはめることで、追跡対象物の位置姿勢を推定する方法がある。エッジ情報以外にも、オプティカルフレーム情報やテクスチャ情報を用いる方法などがあるが、いずれにおいても、CADや3Dスキャナーを用いて予め追跡対象物のモデルを作成する必要がある。

20

【0009】

カメラの位置姿勢と追跡対象物の位置姿勢とを、同じプロセスによって求められる方法として、マーカーを使った位置合わせがある（非特許文献3参照）。この手法では、シーン中の動かない場所に固定したマーカーからカメラの位置及び姿勢を求めることができ、追跡対象物に固定したマーカーから追跡対象物の位置及び姿勢を求めることができる。また、それぞれのマーカーは、マーカーに埋め込まれたIDによって識別することができる。

30

【先行技術文献】**【非特許文献】****【0010】**

【文献】G. Klein and D. Murray : "Parallel Tracking and Mapping for Small AR Workspaces", Proceedings of Sixth IEEE and ACM International Symposium on Mixed and Augmented Reality, 2007.

【文献】J. Lima and F. Simoes and L. Figueiredo and J. Kelner : "Model Based Markerless 3D Tracking applied to Augmented Reality", SBC Journal on 3D Interactive Systems, volume 1, 2010.

40

【文献】H. Kato and M. Billinghurst : "Marker Tracking and hmd calibration for a video-based augmented reality conferencing system", Proceedings of 2nd IEEE and ACM International Workshop on Augmented Reality, 1999.

【発明の概要】

50

【発明が解決しようとする課題】

【0011】

しかし、前述した方法では、マーカーがカメラの視野外に出た場合には位置姿勢を求めることができず、またCADや3Dスキャナーなどを用いた事前準備が必要となり、どのような状況でもカメラ及び追跡物体の位置姿勢を高精度に取得できるとは言い難い。

【0012】

そこで本発明は、カメラ及び追跡物体の位置および姿勢を高精度に取得可能にすることを目的とする。

【課題を解決するための手段】

【0013】

本発明は、追跡対象物を含む現実空間を撮影した画像から複数の特徴情報を抽出する特徴抽出手段と、前記現実空間を撮影した画像から前記追跡対象物に付されている指標を検出し、該指標の位置姿勢を推定する指標推定手段と、前記指標の位置姿勢と追跡対象モデルとに基づいて、前記追跡対象物の位置姿勢を推定する対象物推定手段と、前記現実空間を前記撮影したカメラの位置姿勢および、前記追跡対象物の位置姿勢に基づいて、前記複数の特徴情報を分類する分類手段と、前記分類手段が前記追跡対象物に属していると判定した特徴情報を、前記追跡対象モデルに追加するモデル構築手段と、を有することを特徴とする。

【発明の効果】

【0014】

本発明によれば、カメラ及び追跡物体の位置および姿勢を高精度に取得可能となる。

【図面の簡単な説明】

【0015】

【図1】実施形態に係る情報処理装置の機能構成例を示すブロック図である。

【図2】情報処理装置のハードウェア構成例を示す図である。

【図3】カメラで撮影する現実空間の例を示す模式図である。

【図4】情報処理装置が利用するデータ構造を示す模式図である。

【図5】特徴分類処理の流れを示すフローチャートである。

【図6】情報処理装置における情報処理の流れを示すフローチャートである。

【図7】第2の実施形態に係る特徴分類処理の説明に用いる模式図である。

20

【発明を実施するための形態】

【0016】

以下、実施形態を、添付の図面に基づいて詳細に説明する。なお、以下の実施形態は本発明を必ずしも限定するものではない。

<第1の実施形態>

図1は第1の実施形態に係る情報処理装置の機能構成例を示すブロック図である。

本実施形態の情報処理装置100は、画像入力部101、特徴抽出部102、指標推定部103、カメラ推定部104、対象物推定部105、特徴分類部106、モデル構築部107、モデル保持部108、および画像生成部109を有する。本実施形態では、シーンを撮影するカメラ110および表示装置111を備えたビデオシースルー型のMR(複合現実感)システムを例に挙げている。なお、本実施形態に係るMRシステムは、光学シースルー型のシステムや、ユーザーの視線に応じた表示画像を生成するアイトラッキング型のMRシステムであってもよい。

40

【0017】

本実施形態では、マーカーとカメラ110にて撮影されるシーンの画像の特徴を用いて、当該シーンを撮影するカメラ110の位置及び姿勢と、CG(コンピュータグラフィクス)モデルを重畳表示したい対象物(追跡対象物)の位置及び姿勢とを求める。特に、本実施形態では、マーカーがカメラ110の視野外となる状況でも位置及び姿勢を求ることを可能とし、CAD(コンピュータ支援設計)や3D(3次元)スキャナーを用いて事前に追跡対象物のモデルを作成する必要のない方法を実現可能とする。

50

【 0 0 1 8 】

本実施形態において、カメラ 110 は、左目用と右目用を有するステレオカメラ構成として説明するが、画素の奥行きを計測可能なカメラであればよく、例えばデブスカメラでもよい。また本実施形態のシステムは、例えば 1 台の可視光カメラを構成した上で、機械学習された学習モデルによる奥行推定を行う構成であってもよい。また本実施形態の情報処理装置 100 は、カメラ 110 と表示装置 111 をユーザーの目の付近に配置した頭部装着型として説明するが、各部の物理的な配置はこれに限定されるものではなく、タブレットやスマートフォン等のハンドヘルド型であってもよい。

【 0 0 1 9 】

また実施形態の情報処理装置 100 は、画像からカメラ 110 の位置および姿勢の推定と空間モデル（マップ）の作成とを行うものとして説明する。ただし、カメラ 110 の位置姿勢推定と空間モデル（マップ）作成はこれに限定されるものではなく、加速度センサ、角速度センサ、超音波センサ、磁気センサ、奥行画像センサ、レーザー距離センサなどの情報を用いて行われてもよい。

10

【 0 0 2 0 】

図 2 は、情報処理装置 100 のハードウェア構成例を示した図である。

図 2 において、CPU 201 は、装置全体を制御する。RAM 202 は、CPU 201 が各部を制御しながら処理を行う時に作業領域として用いられる。ROM 203 は、本実施形態に係る情報処理プログラム、各種アプリケーションプログラム、およびデータなどを記憶する。CPU 201 は、ROM 203 に記憶されている情報処理プログラムを RAM 202 に展開して実行する。これにより、CPU 201 は、図 1 の画像入力部 101、特徴抽出部 102、指標推定部 103、カメラ推定部 104、対象物推定部 105、特徴分類部 106、モデル構築部 107、画像生成部 109、およびモデル保持部 108 等の各機能を実現する。また、入力 I/F 204 は、カメラ 110 により撮影された画像信号を、情報処理装置 100 で処理可能な形式の信号として入力するインターフェイス部である。出力 I/F 205 は、外部装置に対して、当該外部装置が処理可能な形式の信号を出力するインターフェイス部である。例えば、外部装置が表示装置 111 である場合、出力 I/F 205 は、画像生成部 109 にて生成された MR 画像信号を、表示装置 111 で表示可能な信号にして出力する。

20

【 0 0 2 1 】

上述したように、図 1 に示す情報処理装置 100 の各部の機能は、CPU 201 が本実施形態に係る情報処理プログラムを実行することで実現することができる。ただし、図 1 に示す情報処理装置 100 の各部のうち、少なくとも一部は不図示の専用のハードウェアや GPU によって実行されるようになされてもよい。この場合、専用のハードウェアや GPU は、CPU 201 の制御に基づいて動作する。

30

【 0 0 2 2 】

以下、情報処理装置 100 の各部について説明する。

画像入力部 101 は、カメラ 110 で撮影した現実空間の画像を取得する。カメラ 110 は動画像を撮像および記録可能な撮像装置であり、画像入力部 101 は、カメラ 110 から動画像のフレーム毎に画像を取得する。カメラ 110 がステレオカメラなど、複数の撮像装置からなるカメラである場合、画像入力部 101 が複数の撮像装置から取得する複数の画像のフレームは同期している（撮影時刻が一致している）ものとする。

40

【 0 0 2 3 】

本実施形態の場合、図 3 に示すように、現実空間には CG モデルを重畠表示したい対象物（追跡対象物 302）が存在しており、この追跡対象物 302 には、予め 1 つ以上の追跡対象マーカー 301 が設置されているものとする。本実施形態では、追跡対象物が 1 つであるとして説明するが、追跡対象物は複数あってもよく、その場合、それぞれに対して少なくとも 1 つの追跡対象マーカーが設置される。以下の説明では、それぞれの追跡対象マーカーに対して設定された 3 次元座標系を追跡対象座標系と呼ぶ。

【 0 0 2 4 】

50

また本実施形態では、現実空間の基準となる3次元座標系（すなわち世界座標系）を定義するための1つ以上の世界座標系マーカー300が配置されているものとする。ただし、現実空間の基準となる3次元座標系を定義する方法は、これに限らず、カメラの初期位置姿勢を基準にする方法や、モーションキャプチャー装置を用いた方法などでもよい。

【0025】

特徴抽出部102は、画像入力部101より入力された画像から、画像中の特徴情報を抽出する。特徴抽出部102にて抽出される特徴情報は、図4(a)に示すように、特徴の識別子、特徴の画像座標(2次元座標)、特徴のカメラ座標系における3次元座標を含む。また、特徴情報には、後述する特徴分類処理によって求められる分類情報も含まれる。なお、特徴情報はこの例に限定されるものではなく、特徴周辺の画像情報や、SIFTなどに代表される特徴記述子、エッジ情報などが含まれてもよい。特徴抽出部102は、画像から抽出した特徴情報を、カメラ推定部104および対象物推定部105へ送出する。

10

【0026】

ここで、特徴情報を抽出する具体的な方法の一例として、Kanade-Lucas-Tomasi法(KLT法)と呼ばれる特徴点の検出と、検出した特徴点をフレーム間で追跡する手法がある。KLT法を用いた場合、特徴抽出部102は、図3に示すように、まず画像から局所的に見てコーナー点と判定される点を特徴点303として検出する。このとき特徴抽出部102は、1つの画像から特徴点303を複数検出する。次に特徴抽出部102は、続くフレームについて、特徴点303の座標周辺の局所領域の見た目に基づく探索によって、新しい画像における座標を算出する。KLT法は、このようにして特徴点を追跡する方法である。

20

【0027】

また本実施形態の場合、カメラ110は画素の奥行きを計測可能なカメラであるため、特徴抽出部102は、特徴点の座標に対する奥行きを計測して、カメラ座標系における3次元座標に変換する。その際に必要となるカメラ110のパラメータなどは、予めキャリブレーションされているものとする。

【0028】

指標推定部103は、画像入力部101より入力された画像から、既知の指標を検出して、指標の位置及び姿勢を推定する。本実施形態の場合、指標推定部103は、視覚マーカーと画像処理を用いた指標の位置姿勢推定を行う。なお、指標推定部103は、視覚マーカーではなく、追跡対象物302の一部をモデル化した情報や、環境設置型の光学センサ、超音波センサ、磁気センサ、加速度センサ、角速度センサなどの検出信号を指標としてその位置姿勢を計測しても良い。

30

【0029】

ここで、視覚マーカーは、画像中に写るマーカーの見え方から、マーカーのIDの特定と、マーカーの位置姿勢の推定を可能にするものである。本実施形態の場合、図3に示す世界座標系マーカー300および追跡対象マーカー301が、既知の指標として予め登録されているものとする。

【0030】

指標推定部103は、入力された画像から、それぞれのマーカーに対し、マーカーのID、そのマーカーを検出したか否か、マーカーの位置姿勢の情報を算出する。そして、指標推定部103は、世界座標系マーカー300に関する情報をカメラ推定部104へ送出し、追跡対象マーカー301に関する情報を対象物推定部105へ送出する。なお、追跡対象マーカーが複数ある場合、指標推定部103は、それぞれの情報を対象物推定部105へ送出する。

40

【0031】

カメラ推定部104は、特徴抽出部102から送出された情報と、指標推定部103から送出された世界座標系マーカー300の情報とを用いて、画像入力部101より入力された画像におけるカメラ110の位置および姿勢を推定する。本実施形態の場合、カメラ

50

推定部 104 が、視覚マーカーの情報と Visual SLAM を併用したカメラの位置姿勢推定アルゴリズムを用いる例を説明する。なお、カメラの位置姿勢推定方法はこれに限らず、その他のアルゴリズムや外部のセンサを用いた方法でもよい。

【0032】

以下、カメラ推定部 104 が用いる、カメラの位置姿勢推定アルゴリズムの具体的手順の例を説明する。

まず、カメラ推定部 104 は、視覚マーカーを用いてカメラの位置姿勢を推定する。指標推定部 103 にて世界座標系マーカー 300 が検出されている場合、カメラ推定部 104 は、その世界座標系マーカー 300 を基に、世界座標系におけるカメラの位置姿勢を取得する。

10

【0033】

次に、カメラ推定部 104 は、特徴抽出部 102 から送出された特徴情報を用いて、特徴情報の世界座標系における 3 次元座標を推定する。ここで、モデル保持部 108 に特徴情報に対応する空間モデル（マップ）の情報が保持されている場合、カメラ推定部 104 は、その情報を参照する。それ以外の場合、カメラ推定部 104 は、特徴情報に含まれるカメラ座標系における 3 次元座標と、世界座標系マーカー 300 によって求めたカメラの位置姿勢の情報から、特徴情報の世界座標系における 3 次元座標を求める。

【0034】

ここでカメラ推定部 104 は、世界座標系における特徴点 303 の 3 次元座標と、特徴点 303 の画像座標との対応から、PnP (Perspective-n-Point) 問題と呼ばれる連立方程式を解くことで、カメラの位置と姿勢を 6 自由度で求める。なお、PnP 問題を解くために必要な、カメラの内部パラメータは予め求めているものとする。そして、カメラ推定部 104 は、PnP 法によって解が得られた場合には、カメラの位置姿勢を更新する。

20

【0035】

対象物推定部 105 は、特徴抽出部 102 から送出された情報と、指標推定部 103 から送出された追跡対象マーカー 301 の情報を用いて、画像入力部 101 より入力された画像における追跡対象物 302 の位置および姿勢を推定する。その際、対象物推定部 105 は、後述するモデル構築部 107 における処理によって追跡対象物 302 に属する情報が保持されているれば、それを参照する。

30

【0036】

本実施形態において、追跡対象物 302 の位置姿勢は、追跡対象マーカー 301 の位置姿勢によって表されるものとする。すなわち追跡対象物 302 の位置姿勢を求める問題は、世界座標系における追跡対象座標系の位置姿勢を求める問題に帰結する。本実施形態の場合、カメラ推定部 104 によって世界座標系におけるカメラ座標系の位置姿勢が既知であるため、カメラ座標系に対する追跡対象座標系の位置姿勢を求める問題として説明する。

【0037】

以下、カメラ座標系に対する追跡対象座標系の位置姿勢を求める処理の手順について説明する。

まず、指標推定部 103 にて追跡対象マーカー 301 が検出された場合、カメラ座標系に対する追跡対象座標系の位置姿勢が得られる（これを初期位置姿勢と呼ぶ）。そして、モデル構築部 107 に特徴情報が保持されていない場合、対象物推定部 105 は、初期位置姿勢を追跡対象座標系の位置姿勢として、対象物の位置姿勢推定処理を終了する。

40

【0038】

一方、モデル構築部 107 に特徴情報が保持されている場合、対象物推定部 105 は、その特徴情報を用いて、追跡対象物の、より安定した位置姿勢を求める。具体的には、対象物推定部 105 は、まずモデル構築部 107 に保持されている特徴情報（以下、保持された特徴情報と呼ぶ）と、特徴抽出部 102 で抽出された特徴情報（以下、現フレームの特徴情報）とを対応付ける。

【0039】

50

ここで、特徴情報を対応付ける方法として、例えば、保持された特徴情報の投影座標を用いる方法がある。この方法を用いる場合、対象物推定部 105 は、まず、保持された特徴情報の追跡対象座標系における 3 次元座標を、初期位置姿勢を用いて現フレームの画像上に投影して投影座標を得る。そして、対象物推定部 105 は、現フレームの特徴情報のうち、該投影座標との距離が最も小さい画像座標を持つ特徴情報を対応付ける。

【 0 0 4 0 】

また、追跡対象マーク 301 が検出されず、初期位置姿勢が得られない場合、対象物推定部 105 は、過去の動きから予測した現フレームにおける追跡対象座標系の位置姿勢を初期位置姿勢としてもよい。また、対象物推定部 105 は、保持された特徴情報と過去のフレームの特徴情報との対応付けを基に、KLT 法など、画像情報を用いて特徴情報のフレーム間対応を得る方法によって現フレームの特徴情報を対応付けてもよい。ただし、特徴情報の対応を得る方法はこれらに限らない。例えば、特徴周辺および投影座標周辺の画像情報や、特徴記述子の類似度によって対応付けが行われてもよいし、特徴情報付近のエッジ情報を用いて対応付けが行われてもよい。

10

【 0 0 4 1 】

対象物推定部 105 は、前述の方法により保持された特徴情報と現フレームの特徴情報の対応付けを得た後は、対応付けが得られた現フレームの特徴情報のみを用いて、追跡対象座標系の位置姿勢を求める。これは、保持された特徴情報の追跡対象座標系における 3 次元座標と、現フレームの特徴情報の画像座標を用いた PnP (Perspective - n - Point) 問題を解くことで求めることができる。ただし、追跡対象座標系の位置姿勢を求める方法はこれに限らない。例えば、保持された特徴情報の追跡対象座標系における 3 次元座標の点群と、現フレームの特徴情報の世界座標系における 3 次元座標の点群から、点同士の距離の総和を最小化するように、追跡対象座標系の位置姿勢が求められてもよい。

20

【 0 0 4 2 】

特徴分類部 106 は、カメラ推定部 104 から送出された情報と、対象物推定部 105 から送出された情報を用いて、特徴抽出部 102 で抽出された特徴情報を図 4 (b) に示すように分類する。

【 0 0 4 3 】

特徴分類部 106 は、まず、特徴情報に含まれるカメラ座標系における 3 次元座標を世界座標系に投影する。そして、特徴分類部 106 は、複数のフレームにおいて、その世界座標系における 3 次元座標が一定であった場合、その特徴情報は空間に属すると分類する。なお、それ以外の場合、特徴分類部 106 は、特徴情報に含まれるカメラ座標系における 3 次元座標を追跡対象座標系に投影する。そして、特徴分類部 106 は、複数のフレームにおいて、その追跡対象座標系における 3 次元座標が一定であった場合、その特徴情報は、その追跡対象座標系を有する追跡対象物 302 に属すると分類する。

30

【 0 0 4 4 】

特徴情報が追跡対象物 302 に属すると分類した場合、特徴分類部 106 は、その特徴情報をモデル構築部 107 に送出する。一方、特徴情報が空間に属すると分類した場合、特徴分類部 106 は、その特徴情報をモデル保持部 108 に送出する。

40

【 0 0 4 5 】

図 5 は、特徴分類部 106 における分類の具体的な方法の一例を示すフローチャートである。

まずステップ S501において、特徴分類部 106 は、判定対象となる特徴点について、N フレーム分の特徴情報が得られているかどうかを判定する (N は 2 以上の整数)。そして、特徴分類部 106 は、N フレーム分の特徴情報が得られていると判定した場合にはステップ S502 に処理を進める。一方、特徴情報が N フレームに満たないと判定した場合、特徴分類部 106 は、ステップ S508 に進んで特徴点は未分類と判定して分類処理を終了する。

【 0 0 4 6 】

50

ステップ S 502 に進むと、特徴分類部 106 は、分類の対象となる特徴情報から、世界座標系における特徴点の 3 次元座標を取得する。この時の特徴分類部 106 は、特徴情報に含まれるカメラ座標系における 3 次元座標を、カメラ推定部 104 で推定したカメラの位置姿勢を用いて、世界座標系における 3 次元座標に変換して取得する。

【0047】

次にステップ S 503において、特徴分類部 106 は、ステップ S 502 で取得した世界座標系における特徴点の 3 次元座標が N フレームに渡って一定であるか否かを判定する。具体的には、特徴分類部 106 は、N フレーム分の 3 次元座標の平均座標を求め、その平均座標と各フレームにおける座標との距離が一定値以内であるか否かによって判定する。ただし、判定の方法はこれに限らない。そして、特徴分類部 106 は、世界座標系における特徴点の 3 次元座標が N フレームに渡って一定であると判定した場合には、ステップ S 504 に進んで特徴点は空間に属すると判定して分類処理を終了する。それ以外の場合、特徴分類部 106 は、ステップ S 505 へ処理を進める。

【0048】

ステップ S 505 に進むと、特徴分類部 106 は、分類の対象となる特徴情報から、追跡対象座標系における 3 次元座標を取得する。この時の特徴分類部 106 は、特徴情報に含まれるカメラ座標系における 3 次元座標を、対象物推定部 105 で推定したカメラ座標系に対する追跡対象座標系の位置姿勢を用いて、追跡対象座標系における 3 次元座標に変換して取得する。

【0049】

次にステップ S 506 において、特徴分類部 106 は、ステップ S 505 で取得した追跡対象座標系における特徴点の 3 次元座標が N フレームに渡って一定であるか否かを判定する。具体的な判定方法は、ステップ S 503 と同等である。特徴分類部 106 は、追跡対象座標系における特徴点の 3 次元座標が N フレームに渡って一定であると判定した場合、ステップ S 507 に進んで、特徴点はその追跡対象物に属すると判定して分類処理を終了する。それ以外の場合、特徴分類部 106 は、ステップ S 508 に進んで、特徴点は未分類と判定して分類処理を終了する。ただし、追跡対象物が複数ある場合、特徴分類部 106 は、それぞれの追跡対象物に対してステップ S 505、ステップ S 506 の処理を実行する。

【0050】

モデル構築部 107 は、特徴分類部 106 で追跡対象に属すると判定された特徴情報に基づく情報を、追跡対象物ごとに蓄積することで追跡対象モデルを構築する。図 4 (c) は、モデル構築部 107 が保持する情報の一例である。すなわち、モデル構築部 107 は、図 4 (c) に示すように、特徴分類部 106 で追跡対象物 302 に属すると判定された特徴の識別子を記録する。また、モデル構築部 107 は、その特徴のカメラ座標系における 3 次元座標を、追跡対象物 302 の追跡対象座標系における 3 次元座標に変換し、記録する。ここで、ある特徴に対し、複数のフレームに渡って追跡対象座標系における 3 次元座標が得られている場合、モデル構築部 107 は、それらの結果を統合して記録してもよいが、その統合の方法については限定しない。

【0051】

モデル保持部 108 は、特徴分類部 106 で空間に属すると判定された特徴情報に対し、世界座標系における 3 次元座標を計算して保持する。その際、モデル保持部 108 は、カメラ推定部 104 で求められた情報を用いる。図 4 (d) は、モデル保持部 108 が保持する情報の一例である。すなわち、モデル保持部 108 は、図 4 (d) に示すように、特徴の識別子、および世界座標系における 3 次元座標の各情報を保持する。なお、世界座標系における特徴の 3 次元座標を高精度に求める問題は、非特許文献 1 を含む様々な Vi sual - SLAM 技術にて取り組まれている問題であり、本発明においてはその方法は限定しない。

【0052】

画像生成部 109 は、カメラ推定部 104 で推定されたカメラの位置姿勢と、対象物推

10

20

30

40

50

定部 105 で推定された追跡対象物 302 の位置姿勢とを用いて、仮想画像を生成する。そして、画像生成部 109 は、生成した仮想画像を、画像入力部 101 から入力された画像と合成することで複合現実画像を生成し、表示装置 111 に表示可能な信号として出力する。

【0053】

ここでは、作業対象の機械の CG モデルを現実空間中の固定された位置に重畠表示するとともに、ユーザーが手に持つ実物体の工具（追跡対象物）にも工具の CG モデルを重畠表示するような、作業シミュレーションシステムを例として説明する。

【0054】

画像生成部 109 は、まず、カメラ推定部 104 で推定された世界座標系に対するカメラ座標系の位置姿勢を仮想カメラに反映することで、現実空間の固定された位置に作業対象の機械の CG モデルが存在するかのような仮想画像を生成する。さらに、画像生成部 109 は、対象物推定部 105 で推定された世界座標系に対する追跡対象座標系の位置姿勢を CG モデルの位置姿勢に反映することで、ユーザーが手に持つ工具（追跡対象物）上に工具の CG モデルが存在するかのような仮想画像を生成する。

10

【0055】

ここで、ユーザーが手に持つ工具（追跡対象物）は、先端が尖っているなどの都合により、視覚マーカーのような指標を設置することが難しく、例えば持ち手付近などに視覚マーカーが設置されることがある。また、視覚マーカーのみを用いて追跡対象物の位置姿勢を求めた場合、ユーザーが工具の先端をよく見ようとして近づくと、視覚マーカーがカメラの観察領域外となり、追跡対象物の位置姿勢を求めることができなくなる。これにより、工具の CG モデルが正しい位置に表示されず、作業シミュレーションの実施が阻害される可能性がある。

20

【0056】

これに対し、本実施形態における方式を用いると、視覚マーカーと工具の先端とがカメラ 110 の観察領域内にある際に、工具の先端の特徴情報をモデル構築部 107 に保持される。したがって、その後に、視覚マーカーが検出できなくなった場合でも、工具の先端の特徴情報を基に、追跡対象物の位置姿勢を求めることができる。これにより、本実施形態では、工具の先端付近に近づいて観察しながら作業が行える作業シミュレーションシステムを実現できる。

30

【0057】

図 6 は、本実施形態の情報処理装置 100 が複合現実画像を生成するために行う処理の流れを示したフローチャートである。

まずステップ S601において、画像入力部 101 は画像を取得する。すなわち画像入力部 101 は、カメラ 110 で撮影された現実空間の画像を、入力 I/F 204 を通して取り込む。

40

【0058】

次にステップ S602において、特徴抽出部 102 は、ステップ S601 で画像入力部 101 が取得した画像から、特徴情報を抽出する。

次にステップ S603において、指標推定部 103 は、ステップ S601 で画像入力部 101 が取得した現実空間の画像から指標を検出し、各指標の位置姿勢を推定する。

【0059】

次にステップ S604において、カメラ推定部 104 は、カメラ 110 の位置姿勢を推定する。すなわちカメラ推定部 104 は、ステップ S601 で取得された現実空間の画像と、ステップ S602 で抽出された特徴情報を、ステップ S603 で推定された指標の位置姿勢とを基に、カメラ 110 の位置姿勢を推定する。

【0060】

次にステップ S605において、対象物推定部 105 は、追跡対象物の位置姿勢を推定する。すなわち対象物推定部 105 は、ステップ S601 で取得された現実空間の画像と、ステップ S602 で抽出された特徴情報を、ステップ S603 で推定された指標の位置

50

姿勢とを基に、追跡対象物 302 の位置姿勢を推定する。

【0061】

次にステップ S606において、特徴分類部 106 は、図 5 で説明したようにして特徴情報の分類処理を行う。すなわち特徴分類部 106 は、ステップ S602 で抽出された特徴情報と、ステップ S604 で推定されたカメラの位置姿勢と、ステップ S605 で推定された追跡対象物の位置姿勢とを基に、特徴情報を分類する。

【0062】

次にステップ S607において、モデル構築部 107 は、特徴情報の 3 次元座標を、追跡対象座標系上の座標に変換して保持することで、追跡対象のモデルを構築する。

そして、ステップ S608において、画像生成部 109 は複合現実画像を生成する。その後、図 6 のフローチャートの処理は終了する。

10

【0063】

<第 1 の変形例>

第 1 の実施形態における特徴分類部 106 では、複数のフレームにおいて特徴点 303 の追跡対象座標系における 3 次元座標が一定であることを分類の基準として説明した。しかし、画像から計測できる特徴点の 3 次元座標は、画像の奥行方向の誤差が生じる場合がある。そこで、特徴分類部 106 は、奥行方向の誤差が影響しにくい判定方法として、追跡対象座標系における 3 次元座標を、追跡対象座標系におけるカメラの位置姿勢を用いて画像上に投影し、特徴点の検出座標との誤差（再投影誤差）を用いる。

【0064】

具体的には、ステップ S506において、特徴分類部 106 は、追跡対象座標系における再投影誤差が N フレームに渡って予め定めた閾値以内である場合に、特徴点はその追跡対象に属すると判定してステップ S507 に進んで分類処理を終了する。

20

同様に、ステップ S503において、特徴分類部 106 は、世界座標系における再投影誤差が N フレームに渡って予め定めた閾値以内である場合に、特徴点はその空間に属すると判定してステップ S504 に進んで分類処理を終了する。

【0065】

<第 2 の実施形態>

第 1 の実施形態の場合、カメラ 110 は、画素の奥行きを計測可能なカメラであるとして説明した。また第 1 の実施形態において、画素の奥行きの情報は、カメラ座標系における特徴点の 3 次元座標を計算するために用いられている。また第 1 の実施形態において、特徴点の 3 次元座標の情報は、カメラ推定部 104、対象物推定部 105、および特徴分類部 106 にて用いられている。ここで、カメラ推定部 104 における処理は、第 1 の実施形態で説明した方法に限るものではなく、単眼カメラによる Visual SLAM の技術などで代替してよい。第 2 の実施形態では、カメラ 110 が、単眼カメラのように、1 フレームの画像からでは画素の奥行きを求められないカメラである場合の例について説明する。なお、第 2 の実施形態における情報処理装置 100 の構成は前述した図 1、図 2 と同様であるため、それらの図示と説明は省略する。以下、第 2 の実施形態において第 1 の実施形態とは異なる構成および動作について説明する。

30

【0066】

第 2 の実施形態において、対象物推定部 105 は、モデル構築部 107 に保持されている特徴点のみ、特徴点の 3 次元座標を必要とする。そのため、本実施形態における対象物推定部 105 における処理は、第 1 の実施形態と同じでよい。

40

【0067】

ここで、第 2 の実施形態のように、単眼のカメラを用いて、特徴点の画像座標からその 3 次元座標を求める方法として、モーションステレオ法がある。モーションステレオ法は、カメラで撮影された複数のフレームにおいて空間中のある点 X の画像座標が得られており、ある座標系 C に対する各フレームのカメラの位置姿勢が求まっている場合に、点 X の 3 次元座標を求める方法である。

【0068】

50

図7(a)および図7(b)を用いてその具体例について説明する。図7(a)および図7(b)中の三角印702は、異なる時刻(フレーム)における、追跡対象マーカー301に対するカメラの位置姿勢を表している。

図7(a)に示すように、追跡対象物302上の点が特徴点701として検出された場合、特徴点701は追跡対象座標系において一定の位置にある。そのため、各フレームにおける特徴点701のレイ(カメラの光学中心から、特徴点の画像座標を通る3次元空間における光線)は1点で交差し、当該交点の座標が特徴点701の追跡対象座標系における3次元座標となる。

【0069】

一方、図7(b)に示すように、追跡対象座標系において一定の位置にない特徴点703の場合、各フレームにおける特徴点703のレイは1点で交差しない。そのことから特徴点703が追跡対象物とは違う物体上の点であると判定できる。

第2の実施形態の場合、この性質を利用することにより、特徴分類部106における図5に示したフローチャートの処理は以下のようになる。

【0070】

第2の実施形態の場合、ステップS505において、特徴分類部106は、追跡対象座標系におけるカメラの位置姿勢と特徴点の画像座標から、追跡対象座標系におけるレイを計算する。

【0071】

次にステップS506において、特徴分類部106は、Nフレームのレイが交差するかどうかを判定する。そして、特徴分類部106は、交差していると判定した場合には、ステップS507に進み、特徴点はその追跡対象に属すると判定して分類処理を終了する。一方、それ以外の場合、特徴分類部106は、ステップS508に進み、特徴点は未分類と判定して分類処理を終了する。ただし、追跡対象物が複数ある場合、特徴分類部106は、それぞれの追跡対象物に対してステップS505、ステップS506を実行する。

【0072】

第2の実施形態の場合、特徴分類部106は、世界座標系におけるカメラ位置姿勢を用いて同様の方法を用いることで、特徴点が空間に属するか否かを分類することができる。ステップS502において、特徴分類部106は、世界座標系におけるカメラの位置姿勢と特徴点の画像座標とを基に、世界座標系におけるレイを計算する。そして、ステップS503において、特徴分類部106は、Nフレームのレイが交差するかどうかを判定し、交差している場合にはステップS504に進んで、特徴点は空間に属すると判定して分類処理を終了する。一方、それ以外の場合、特徴分類部106は、ステップS505へ移行する。

【0073】

<第2の変形例>

第1の実施形態や第2の実施形態において、CGモデルを重畠表示したい対象物(追跡対象物302)は、人が手で掴んで動かされる場合がある。その場合、人の手の動きが追跡対象物302と一致するため、人の手に対して検出された特徴情報が、特徴分類部106にて、追跡対象に属する特徴情報として分類される可能性がある。これを回避するために、特徴分類部106は、画像入力部101より入力される画像のうち、人体など、追跡対象物ではないと判定される領域をマスク領域として推定し、そのマスク領域内の特徴情報は空間にも追跡対象にも属しない点と判定してもよい。なお、マスク領域の推定の具体的な方法として、各画素の色が予め登録された肌色情報と一致している場合、マスク領域とする方法がある。マスク領域の推定方法はこの例に限定されるものではない。

【0074】

<第3の変形例>

第1の実施形態や第2の実施形態における前述した図4(a)および図4(b)に示した特徴情報には、分類の確からしさの項目が加えられてもよい。ある特徴情報に対し、特徴分類部106にて同じ判定が繰り返されるにつれて、その分類の確からしさは上がって

10

20

30

40

50

いくと考えられる。そのため、特徴分類部 106 は、その繰り返し数に基づく値を計算し、それを分類の確からしさの項目に記録する。

【0075】

具体例を説明すると、第1の実施形態や第2の実施形態において、特徴分類部 106 は、N フレームに渡って、3 次元座標もしくは投影誤差もしくはレイの交差を確認することで分類を行っている。ここで、N の数（つまり画像数）が小さい場合、特徴情報の抽出誤差や、位置姿勢の検出誤差などの影響による誤判定の影響が生じやすくなる。一方、N の数（画像数）が大きいほど、分類の確からしさは向上するが、N フレーム分のデータが得られるまで判定できないため、特徴情報の分類に時間を要する。

【0076】

そこで、特徴分類部 106 は、3 次元座標が一定である、もしくは投影誤差が閾値以内、もしくはレイが同じ点で交差する画像数をカウントし、この画像数に準じた分類の確からしさを特徴情報に記録する。これにより、フレーム数が少ない時には暫定的な分類結果が得られ、かつ、時間経過に応じてより精度の高い分類結果を得ることができる。

【0077】

以上説明したように、各実施形態の情報処理装置 100 によれば、マーカーを設置する以外の事前準備を必要とせず、カメラおよび追跡物体の位置姿勢を高い精度で推定することが可能なMRシステムを実現することができる。

【0078】

以上、本発明の各実施形態について説明したが、本発明はこれら特定の実施形態に限られるものではなく、この発明の要旨を逸脱しない範囲の様々な形態も本発明に含まれる。上述の各実施形態の一部を適宜組み合わせることも可能である。

【0079】

前述した実施形態の図1に示した構成において、例えば画像入力部 101 やモデル保持部 108 を除く各構成部では、人工知能（AI : artificial intelligence）を適用した処理が行われてもよい。例えば、それらの各部の代わりとして、機械学習された学習済みモデルを用いてもよい。その場合には、それら各部への入力データと出力データとの組合せを学習データとして複数個準備し、それらから機械学習によって知識を獲得し、獲得した知識に基づいて入力データに対する出力データを結果として出力する学習済みモデルを生成する。この学習済みモデルは、例えばニューラルネットワークモデルで構成できる。そして、それらの学習済みモデルは、前述の各部と同等の処理をするためのプログラムとして、CPU あるいは GPU などと協働で動作することにより、前述した各部の処理を行う。また、前述の学習済みモデルは、必要に応じて一定のデータを処理する毎に更新する等も可能である。

【0080】

本発明は、上述の実施形態の1以上の機能を実現するプログラムを、ネットワーク又は記憶媒体を介してシステム又は装置に供給し、そのシステム又は装置のコンピュータにおける一つ以上のプロセッサがプログラムを読み出し実行する処理でも実現可能である。また、1以上の機能を実現する回路（例えば、ASIC）によっても実現可能である。

上述の実施形態は、何れも本発明を実施するにあたっての具体化の例を示したものに過ぎず、これらによって本発明の技術的範囲が限定的に解釈されてはならないものである。即ち、本発明は、その技術思想、又はその主要な特徴から逸脱することなく、様々な形で実施することができる。

【符号の説明】

【0081】

100：情報処理装置、110：カメラ、101：画像入力部、102：特徴抽出部、103：指標推定部、104：カメラ推定部、105：対象物推定部、106：特徴分類部、107：モデル構築部、108：モデル保持部、109：画像生成部

10

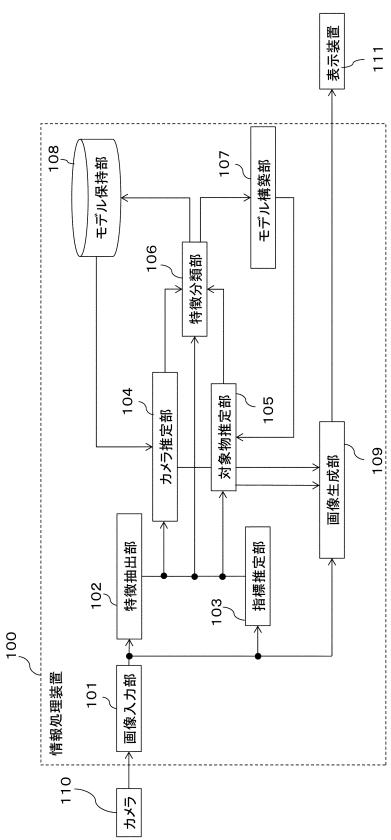
20

30

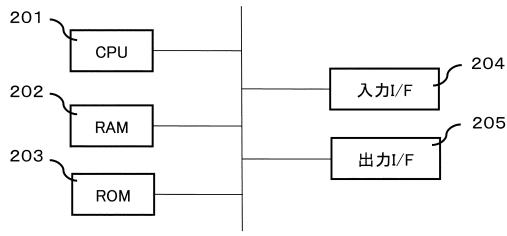
40

50

【図面】
【図 1】



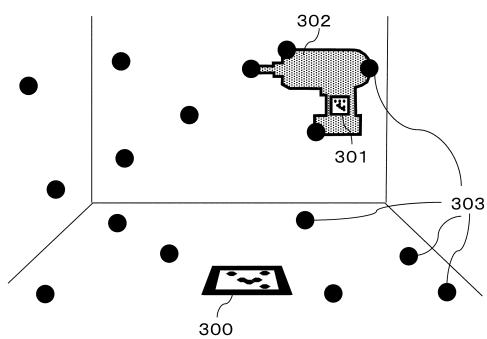
【図 2】



10

20

【図 3】



【図 4】

特徴の識別子	0	1	2
画像座標	u_0, v_0	u_1, v_1	u_2, v_2
カメラ座標系における3次元座標	X_{c0}, Y_{c0}, Z_{c0}	X_{c1}, Y_{c1}, Z_{c1}	X_{c2}, Y_{c2}, Z_{c2}
分類			

(a)

特徴の識別子	0	1	2
画像座標	u_0, v_0	u_1, v_1	u_2, v_2
カメラ座標系における3次元座標	X_{c0}, Y_{c0}, Z_{c0}	X_{c1}, Y_{c1}, Z_{c1}	X_{c2}, Y_{c2}, Z_{c2}
分類	未分類	空間	追跡対象物 3 0 2

(b)

追跡対象物	追跡対象物 3 0 2
特徴の識別子	2
追跡対象座標系における3次元座標	X_{c2}, Y_{c2}, Z_{c2}
分類	未分類

(d)

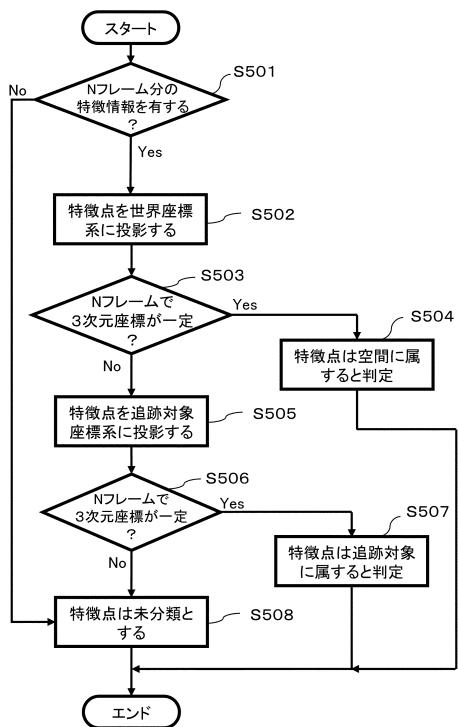
特徴の識別子	1	2
世界座標系における3次元座標	X_{w1}, Y_{w1}, Z_{w1}	
分類		

30

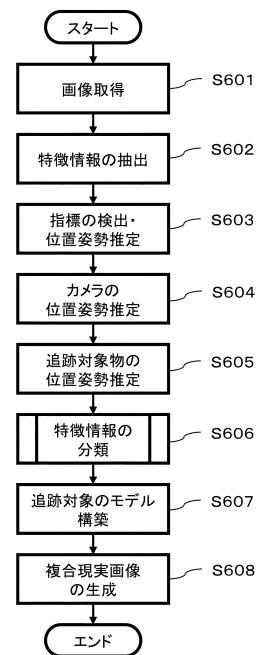
40

50

【図 5】



【図 6】



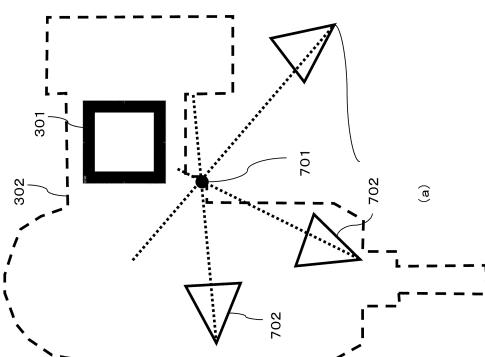
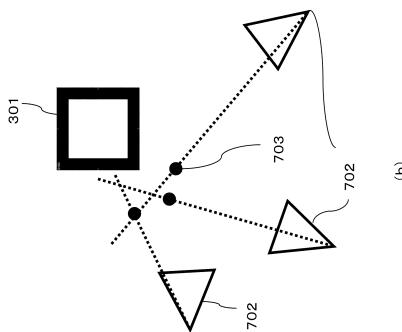
10

20

30

40

【図 7】



50

フロントページの続き

(56)参考文献 米国特許出願公開第2019/0050668(US,A1)

特開2015-153035(JP,A)

特開2019-008460(JP,A)

(58)調査した分野 (Int.Cl., DB名)

G 06 T 7 / 70

G 06 T 7 / 246