

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7524548号
(P7524548)

(45)発行日 令和6年7月30日(2024.7.30)

(24)登録日 令和6年7月22日(2024.7.22)

(51)国際特許分類

FI

G 0 6 T 7/70 (2017.01) G 0 6 T 7/70 A

G 0 6 T 7/00 (2017.01) G 0 6 T 7/00 3 5 0 B

請求項の数 4 (全23頁)

(21)出願番号	特願2020-18551(P2020-18551)	(73)特許権者	000004237
(22)出願日	令和2年2月6日(2020.2.6)		日本電気株式会社
(65)公開番号	特開2021-125000(P2021-125000 A)	(74)代理人	東京都港区芝五丁目7番1号 100103894
(43)公開日	令和3年8月30日(2021.8.30)		弁理士 家入 健
審査請求日	令和5年1月6日(2023.1.6)	(72)発明者	藤若 雅也
			東京都港区芝五丁目7番1号 日本電気 株式会社内
		審査官	吉川 康男

最終頁に続く

(54)【発明の名称】 画像処理装置、検出方法、及びプログラム

(57)【特許請求の範囲】

【請求項1】

対象物を含む映像を用いて、対象空間における前記対象物の位置を推定する位置推定部と、

前記対象物が表示された映像フレーム及び前記対象物の位置を入力とし、前記対象物を認識するための学習モデルを構築する学習モデル構築部と、

UIデバイスに設けられたカメラがとらえている空間の領域を推定し、推定した前記空間に含まれている前記対象物の位置を示すデータを生成する空間推定部と、

前記カメラで撮影された映像と、前記データとを前記学習モデルに入力し、前記カメラで撮影された映像に含まれる前記対象物の検出を行う検出部と、を備える画像処理装置。

【請求項2】

前記空間推定部は、

前記カメラがとらえている空間の映像から前記対象物の切り出しを行い、切り出された前記対象物の前記映像における位置と、推定した前記空間をなす領域とを用いて、前記空間をなす領域内における、前記対象物の存在する位置を推定する、請求項1に記載の画像処理装置。

【請求項3】

対象物を含む映像を用いて、対象空間における前記対象物の位置を推定し、

前記対象物が表示された映像フレーム及び前記対象物の位置を入力とし、前記対象物を認識するための学習モデルを構築し、

UIデバイスに設けられたカメラがとらえている空間の領域を推定し、
推定した前記空間に含まれている前記対象物の位置を示すデータを生成し、
前記カメラで撮影された映像と、前記データとを前記学習モデルに入力し、前記カメラ
で撮影された映像に含まれる前記対象物の検出を行う、検出方法。

【請求項 4】

対象物を含む映像を用いて、対象空間における前記対象物の位置を推定し、
前記対象物が表示された映像フレーム及び前記対象物の位置を入力とし、前記対象物を
認識するための学習モデルを構築し、

UIデバイスに設けられたカメラがとらえている空間の領域を推定し、
推定した前記空間に含まれている前記対象物の位置を示すデータを生成し、
前記カメラで撮影された映像と、前記データとを前記学習モデルに入力し、前記カメラ
で撮影された映像に含まれる前記対象物の検出を行うことをコンピュータに実行させるプ
ログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、画像処理装置、検出方法、及びプログラムに関する。

【背景技術】

【0002】

AR (Augmented Reality) を用いて、場所もしくは物の詳細情報を仮想オブジェクト
として、その場所もしくは物にオーバレイ表示することによって、業務支援を行うことが
可能となる。効果的な業務支援を行うために、可能な限りARの情報提示にかかる時間を減
らすことが重要となる。

【0003】

ARを表示するために、映像認識技術を用いた手法が用いられることがある。具体的には
、ARを用いたサービス提供者が、AR表示箇所を、スマートフォンやARグラス等の端末に
搭載されるカメラ等を用いて事前に撮影し、映像解析装置等に、ARを用いて仮想オブジェ
クトを表示させる対象物の特徴点を登録する。その後、ARを表示可能なヘッドセット、A
RグラスなどのUI (User Interface: ユーザインタフェース) デバイスに表示される映
像が、対象物の特徴点と一致した際に、該UIデバイスのディスプレイは、当該映像と仮
想オブジェクトとを重ねた表示態様 (オーバレイ表示) によって表示する。

【0004】

非特許文献 1 及び 2 には、時系列な関係にある複数の映像フレーム (以下、単に「フレ
ーム」とも称する) のうち、前のフレームの分析結果から、次のフレームの対象物の位置
を予測することによって、イメージ全体の再スキャンを防ぎ、ARを表示するための計算量
を削減する処理が開示されている。

【先行技術文献】

【非特許文献】

【0005】

【文献】 Huizi Mao, Xiaodong Yang, William J. Dally, “A Delay Metric for Video
Object Detection: What Average Precision Fails to Tell”, November 6, 2019

【文献】 Huizi Mao, Taeyoung Kong, William J. Dally, “CATDET: CASCADED TRA
CKED DETECTOR FOR EFFICIENT OBJECT DETECTION FROM VIDEO”, February
19, 2019

【発明の概要】

【発明が解決しようとする課題】

【0006】

しかし、非特許文献 1 及び 2 に開示されている計算量の削減処理を実行した場合であっ
ても、フレームの映像品質が悪い場合には、フレームに含まれる対象物を認識するために
、フレームの映像品質が良い場合と比較して多くのフレームが必要となる。その結果、フ

10

20

30

40

50

レームの映像品質が悪くなるにつれて、ARを表示するための対象物の認識に時間がかかるという問題がある。

【0007】

本開示の目的は、ARを表示するための対象物の認識にかかる時間を削減することができる画像処理装置、検出方法、及びプログラムを提供することにある。

【課題を解決するための手段】

【0008】

本開示の第1の態様にかかる画像処理装置は、対象空間を1以上の空間に分割した分割空間を生成し、前記分割空間に含まれる少なくとも1以上の対象物であって、前記対象物を認識するための学習モデルを構築する学習モデル構築部と、前記学習モデルと、前記学習モデルによって認識される前記対象物を含む前記分割空間をなす領域とを関連付けて管理する学習モデル管理部と、UI (User Interface) デバイスに設けられたカメラがとらえているカメラ認識空間をなす領域を推定する空間推定部と、前記学習モデル管理部によって管理されている前記学習モデルの中から、前記推定されたカメラ認識空間をなす領域を含む前記分割空間をなす領域が関連付けられている特定の学習モデルを選択し、前記選択した特定の学習モデルを用いて、前記UIデバイスに表示された空間に含まれる前記対象物を検出する検出部と、を備える。

10

【0009】

本開示の第2の態様にかかる検出方法は、対象空間を1以上の空間に分割した分割空間を生成し、前記分割空間に含まれる少なくとも1以上の対象物であって、前記対象物を認識するための学習モデルを構築し、前記学習モデルと、前記学習モデルによって認識される前記対象物を含む前記分割空間をなす領域とを関連付けて管理し、UIデバイスに設けられたカメラがとらえているカメラ認識空間の領域を推定し、前記管理している前記学習モデルの中から、前記推定されたカメラ認識空間をなす領域を含む前記分割空間をなす領域が関連付けられている特定の学習モデルを選択し、前記選択した特定の学習モデルを用いて、前記UIデバイスに表示された空間に含まれる前記対象物の検出を行う。

20

【0010】

本開示の第3の態様にかかるプログラムは、対象空間を1以上の空間に分割した分割空間を生成し、前記分割空間に含まれる少なくとも1以上の対象物であって、前記対象物を認識するための学習モデルを構築し、前記学習モデルと、前記学習モデルによって認識される前記対象物を含む前記分割空間をなす領域とを関連付けて管理し、UIデバイスに設けられたカメラがとらえているカメラ認識空間をなす領域を推定し、前記管理している前記学習モデルの中から、前記推定されたカメラ認識空間をなす領域を含む前記分割空間をなす領域が関連付けられている特定の学習モデルを選択し、前記選択した特定の学習モデルを用いて、前記UIデバイスに表示された空間に含まれる前記対象物を検出することをコンピュータに実行させる。

30

【発明の効果】

【0011】

本開示により、ARを表示するための対象物の認識にかかる時間を削減することができる画像処理装置、検出方法、及びプログラムを提供することができる。

40

【図面の簡単な説明】

【0012】

【図1】実施の形態1にかかる画像処理装置の構成図である。

【図2】実施の形態2にかかる画像処理装置の構成図である。

【図3】実施の形態2にかかる分割空間に関連付けられた学習モデルを説明する図である。

【図4】実施の形態2にかかる空間推定部13が推定する空間を示す図である。

【図5】実施の形態2にかかる準備フェーズにおける処理を説明する図である。

【図6】実施の形態2にかかる実行フェーズにおける処理を説明する図である。

【図7】実施の形態2にかかる画像処理装置の変形例を示す図である。

【図8】実施の形態2にかかる画像処理装置の変形例を示す図である。

50

【図 9】実施の形態 3 にかかる学習モデルの概要を示す図である。

【図 10】実施の形態 3 にかかる準備フェーズにおける処理を説明する図である。

【図 11】実施の形態 3 にかかる実行フェーズにおける処理を説明する図である。

【図 12】実施の形態 4 にかかる準備フェーズにおける処理を説明する図である。

【図 13】実施の形態 4 にかかる対象物の位置を可視化した図である。

【図 14】実施の形態 4 にかかる対象物の位置を可視化した図である。

【図 15】実施の形態 4 にかかる対象物の位置を可視化した図である。

【図 16】実施の形態 4 にかかる実行フェーズにおける処理を説明する図である。

【図 17】それぞれの実施の形態にかかる画像処理装置の構成図である。

【発明を実施するための形態】

【0013】

(実施の形態 1)

以下、図面を参照して本発明の実施の形態について説明する。図 1 を用いて実施の形態 1 にかかる画像処理装置 10 の構成例について説明する。画像処理装置 10 は、プロセッサがメモリに格納されたプログラムを実行することによって動作するコンピュータ装置であってもよい。画像処理装置 10 は、例えば、サーバ装置であってもよい。

【0014】

画像処理装置 10 は、学習モデル構築部 11、学習モデル管理部 12、空間推定部 13、及び検出部 14 を有している。学習モデル構築部 11、学習モデル管理部 12、空間推定部 13、及び検出部 14 等の画像処理装置 10 の構成要素は、プロセッサがメモリに格納されたプログラムを実行することによって処理が実行されるソフトウェアもしくはモジュールであってもよい。または、画像処理装置 10 の構成要素は、回路もしくはチップ等のハードウェアであってもよい。

【0015】

学習モデル構築部 11 は、AR の対象空間を 1 以上の空間に分割した分割空間を生成し、分割空間に含まれる少なくとも 1 以上の対象物であって、AR により仮想オブジェクトが付加される対象物を認識するための学習モデルを構築する。

【0016】

対象空間は、例えば、工場内、店舗内、もしくは居室内等の、外部の領域から区別される特定の領域を有する空間であってもよい。分割空間は、任意の分割方法を適用して、分割された空間である。AR により仮想オブジェクトが付加される対象物は、例えば、移動可能な物体もしくは構造物であってもよく、特定の場所から移動することのない物体もしくは構造物であってもよい。仮想オブジェクトが付加される対象物は、人もしくは動物等であってもよい。学習モデルは、対象物の映像データを正解データもしくは教師データとし、対象物を含む映像データから、対象物を認識するために用いられる。

【0017】

学習モデル構築部 11 は、複数の分割空間が存在する場合、分割空間毎に、学習モデルを構築する。つまり、それぞれの学習モデルは、分割空間内に含まれる対象物を認識するために用いられる。

【0018】

学習モデル管理部 12 は、学習モデルと、学習モデルにおいて認識される対象物を含む分割空間の領域とを関連付けて管理する。学習モデル管理部 12 は、分割空間毎に構築されたそれぞれの学習モデルを、それぞれの分割空間と関連付けて管理する。

【0019】

空間推定部 13 は、対象物に仮想オブジェクトを付加した映像を表示する UI デバイス 15 搭載のカメラ 16 がとらえている（即ち、撮影している）カメラ認識空間がなす領域を推定する。対象物に仮想オブジェクトを付加した映像を表示する UI デバイス 15 は、対象物を認識した場合に、その対象物の補足情報として、仮想オブジェクトを映像に重ねた表示状態で表示させることができる画像処理装置と捉えることができる。UI デバイス 15 は、例えば、スマートフォン端末もしくはタブレット型端末であってもよく、スマートグラ

10

20

30

40

50

スもしくはARグラスであってもよい。UIデバイス15は、図1に示すように、画像処理装置10とは異なる装置である場合、無線通信もしくは固定通信等によって、映像を画像処理装置10へ送信してもよい。もしくは、UIデバイス15は、画像処理装置10と一体の装置として、画像処理装置10に内蔵もしくは画像処理装置10に外付けされてもよい。このようなUIデバイス15と画像処理装置10との構成のバリエーションは、後述する他の実施の形態においても同様に適用される。

【0020】

検出部14は、推定されたカメラ認識空間の領域を含む分割空間の領域と関連付けられた学習モデルを選択し、UIデバイス15に表示された空間に含まれる対象物の検出を行う。また、分割空間に含まれるカメラ認識空間の領域は、例えば、カメラ認識空間のすべての領域が分割空間の領域に含まれてもよく、カメラ認識空間の一部の領域が分割空間の領域に含まれることであってもよい。

10

【0021】

以上説明したように、画像処理装置10は、対象空間を分割した分割空間内に存在する対象物を認識するために用いられる学習モデルを構築することができる。このような学習モデルは、対象空間内に含まれるすべての対象物を認識するための学習モデルと比較して、精度が良く、認識に多くの映像フレームを必要としない。そのため、画像処理装置10において生成される学習モデルは、対象空間内に含まれるすべての対象物を認識するための学習モデルと比較して、短時間に対象物を認識することができる。このため、例えば、画像処理装置10を内蔵するUIデバイス15は、係る学習モデルによる認識結果を用いることによって、対象物に応じたARを迅速に表示することができる。

20

【0022】

また、画像処理装置10は、仮想オブジェクトを付加した映像を表示するUIデバイス15搭載のカメラ16がとらえているカメラ認識空間を推定することによって、対象物の検出を行うための学習モデルを選択することができる。これより、画像処理装置10は、カメラ16がとらえているカメラ認識空間内の対象物を認識するために必要な学習モデルを適切に選択することができる。その結果、画像処理装置10は、対象物に仮想オブジェクトを付加した映像を短時間に提供することができる。

【0023】

(実施の形態2)

30

続いて、図2を用いて実施の形態2にかかる画像処理装置20の構成例について説明する。画像処理装置20は、図1の画像処理装置10に、映像取得部21、センサ情報取得部22、及び位置推定部23が追加された構成である。以下においては、主に、画像処理装置10と異なる画像処理装置20の構成について説明する。

【0024】

映像取得部21は、例えば、撮像装置であるカメラであってもよい。映像取得部21は、ARの対象空間全体が表示された映像フレームを取得する。さらに、映像取得部21は、ARにより仮想オブジェクトが付加される対象物が主に表示された映像フレームを取得する。ARにより仮想オブジェクトが付加される対象物が主に表示された映像フレームは、例えば、ARにより仮想オブジェクトが付加される対象物と、その対象物の周辺の背景とが含まれる映像フレームであってもよい。ARにより仮想オブジェクトが付加される対象物が主に表示された映像フレームは、ARにより仮想オブジェクトが付加される対象物が一つのみ含まれる映像フレームであってもよい。

40

【0025】

センサ情報取得部22は、深度を検出するデプスセンサ、加速度を検出する加速度センサ、及び角速度を検出する角速度センサを含む。もしくは、センサ情報取得部22は、深度センサや、加速度及び角速度を検出するIMU (Inertial Measurement Unit) センサであってもよい。

【0026】

位置推定部23は、映像取得部21から受け取った映像フレームを用いて、ARの対象空

50

間内における、ARにより仮想オブジェクトが付加される対象物の位置を推定する。ARの対象空間、及びARにより仮想オブジェクトが付加される対象物の位置は、同一の3次元空間内の3次元座標を用いて示されてもよい。

【0027】

例えば、位置推定部23は、センサや映像フレームを用いて自己位置を推定するとともに周囲の環境地図を作成するSLAM (Simultaneous Localization and Mapping) を実行することによって、対象物の位置を推定してもよい。

【0028】

もしくは、位置推定部23は、映像フレームと、センサ情報取得部22から取得したセンサ情報とを組み合わせ、対象物の位置を推定してもよい。例えば、位置推定部23は、映像フレームを用いて推定した対象物の位置と、センサ情報とを用いて推定したカメラから対象物までの位置とを組み合わせることによって、推定精度を向上させてもよい。

【0029】

学習モデル構築部11は、対象空間を分割し、分割空間を生成する。例えば、学習モデル構築部11は、予め定められた分割数D1 (D1は、1以上の整数) に従って分割空間を生成する。言い換えると、学習モデル構築部11は、D1個の分割空間を生成する。例えば、学習モデル構築部11は、対象空間を格子状に分割してもよい。もしくは、学習モデル構築部11は、対象物が複数の分割空間に分割されないように分割空間を生成してもよい。また、それぞれの分割空間は、互いに重複する領域を有してもよい。つまり、分割空間同士のオーバーラップがあってもよい。分割空間は、1つの対象物を含んでもよく、2以上の対象物を含んでもよい。

【0030】

学習モデル構築部11は、分割空間に含まれる対象物を認識するための学習モデルを構築する。例えば、学習モデル構築部11は、映像取得部21から受け取った対象物が主に表示された映像フレームを正解データもしくは教師データとし、対象物を認識するための学習モデルを構築する。さらに、学習モデル構築部11は、それぞれの分割空間の背景を学習するために、対象空間全体が表示される映像フレームもしくは対象物の背景が表示される映像フレームを用いてもよい。

【0031】

学習モデル構築部11は、それぞれの分割空間に含まれる対象物を、位置推定部23において推定された対象物の位置情報を用いて特定してもよい。例えば、学習モデル構築部11は、推定された対象物Aの位置を含む分割空間Aに、対象物Aが含まれると判定してもよい。もしくは、学習モデル構築部11は、対象物の位置情報を利用しなくてもよい。例えば、画像処理装置20の管理者等が、分割空間Aに含まれる対象物Aを認識し、対象物Aが映された映像フレームを、分割空間Aの学習モデルを構築するための正解データとして、学習モデル構築部11へ入力してもよい。

【0032】

学習モデル構築部11は、ARにより仮想オブジェクトが付加される対象物が存在しない分割空間については、学習モデルを構築しない。

【0033】

学習モデル構築部11は、例えば、CNN (Convolutional Neural Network)、クラスタリング等を実行することによって学習モデルを構築してもよく、その他の手法を用いて学習モデルを構築してもよい。

【0034】

学習モデル構築部11は、分割数D1に従って分割空間を生成し、それぞれの分割空間に関連する学習モデルを構築した後に、分割数D1よりも小さい値である分割数D2に従って再度分割空間を生成してもよい。

【0035】

図3は、分割空間と、それぞれの分割空間に関連付けられた学習モデルとを示す図である。図3の左側の図は、対象空間が4分割され、4つの分割空間のそれぞれに学習モデル

10

20

30

40

50

が関連付けられていることを示している。また、図3の右側の図は、分割数が1であり、対象空間と分割空間とが一致しており、一つの学習モデルのみが存在していることを示している。

【0036】

学習モデル構築部11は、D2個の分割空間のそれぞれについて、学習モデルを構築する。学習モデル構築部11は、分割数を変えながら分割数が規定値に達するまで、対象空間の分割と学習モデルの構築を繰り返す。分割数D2は、例えば、D1個の分割空間のそれぞれについて、隣接した分割空間が結合されたような分割数となるように定められてもよい。例えば、分割数D2は、分割数D1の半分の値となるようにし、D1個の分割空間のそれぞれについて、隣り合う二つの分割空間が結合されてもよい。図3は、分割数の規定値が1であり、分割数が4から分割数が1になるまで、対象空間の分割が繰り返されたことを示している。

10

【0037】

学習モデル構築部11は、構築した学習モデルを、分割空間と関連付けて学習モデル管理部12に記録、もしくは格納する。具体的には、学習モデル構築部11は、分割空間Aと、分割空間Aに含まれる対象物を認識するために構築された学習モデルAとを関連付ける。

【0038】

位置推定部23及び学習モデル構築部11において実行される処理は、対象物の検出を行う前に実行される準備処理であり、準備フェーズにおける処理である。以下において、ARにより仮想オブジェクトを付加する対象物の検出を行うための実行フェーズにおける処理を実行する空間推定部13及び検出部14について説明する。

20

【0039】

空間推定部13は、ARにより仮想オブジェクトを付加した映像を表示するUIデバイス15（以下、「AR端末」と称する場合がある）に設けられたディスプレイにおいて表示される映像に対応する画像データを受け取る。AR端末から入手する映像は、例えば、動画データであってもよい。映像取得部21が、AR端末に設けられている場合、空間推定部13は、映像取得部21から映像を受け取る。もしくは、空間推定部13は、映像取得部21とは異なるAR端末から映像を受け取ってもよい。さらに、空間推定部13は、センサ情報取得部22から深度、加速度、角速度等のセンサ情報を受け取る。

30

【0040】

空間推定部13は、映像及びセンサ情報を用いて、AR端末の位置及び傾きを推定し、AR端末のカメラがとらえている空間を推定する。AR端末のカメラがとらえている空間は、AR端末のディスプレイ等の表示部に表示されている映像が示す空間であってもよい。AR端末がとらえている空間は、カメラ認識空間に相当する。また、空間推定部13が用いる映像及びセンサ情報は、静止画データである複数の映像フレーム及びセンサ情報と言い換えられてもよい。空間推定部13は、SLAMを用いてAR端末の周囲の環境地図を作製することによって、カメラ認識空間を推定してもよい。

【0041】

ここで、図4を用いて、空間推定部13が推定する空間について説明する。図4は、X軸、Y軸、及びZ軸を用いて位置を特定する3次元空間を示している。AR端末30は、映像取得部21であってもよい。空間推定部13が推定する空間は、AR端末30を頂点とする、図4に示される四角錐の空間であってもよい。

40

【0042】

空間推定部13は、AR端末30のスペックとして定められている視野角、もしくは画角等に基づいて、AR端末30のカメラがとらえている空間を推定してもよい。スペックは、仕様、カタログ値等と言い換えられてもよい。また、空間推定部13は、AR端末30の視野角等を用いて、SLAMを用いてAR端末の周囲の環境地図を作製することによって推定したカメラ認識空間の領域を絞り込んでもよい。

【0043】

50

検出部 14 は、空間推定部 13 において推定されたカメラ認識空間の領域が、複数の分割空間のうち、少なくとも 1 つに含まれるか否かを判定する。検出部 14 は、カメラ認識空間の領域が、一つの分割空間に含まれると判定した場合、カメラ認識空間の領域が含まれる分割空間に関連付けられた学習モデルを選択する。検出部 14 は、選択した学習モデルを用いて、AR 端末の映像に含まれる対象物を検出する。

【0044】

検出部 14 は、カメラ認識空間の領域を含む分割空間が複数あると判定した場合、分割空間の領域が、予め定められた基準値もしくは閾値よりも、小さい領域をなす分割空間に関連付けられた学習モデルを選択してもよい。また、検出部 14 は、カメラ認識空間の領域が、複数の分割空間に含まれると判定した場合、最も小さい領域をなす分割空間に関連付けられた学習モデルを選択してもよい。検出部 14 は、選択した学習モデルを用いて、AR 端末の映像に含まれる対象物を検出する。カメラ認識空間の領域を含む分割空間が複数ある場合とは、例えば、カメラ認識空間が、D1 個に分割されたうちの一つの分割空間に含まれ、さらに、D2 個に分割されたうちの一つの分割空間に含まれる場合がある。

【0045】

画像処理装置 20 は、検出部 14 において検出された対象物に、予め対象物に関連付けられた仮想オブジェクトを付加して、AR 端末に表示させる。例えば、画像処理装置 20 は、仮想オブジェクトが付加された映像を AR 端末へ送信する。もしくは、画像処理装置 20 内において生成された、仮想オブジェクトが付加された映像が、映像取得部 21 へ出力される。

【0046】

続いて、学習モデル構築部 11 が、学習モデルの再学習を行う再学習フェーズについて説明する。例えば、学習モデル構築部 11 は、AR により仮想オブジェクトが付加される対象物が追加された場合に、学習モデルの再学習を行う。

【0047】

再学習フェーズにおいて、位置推定部 23 は、映像取得部 21 から、追加された対象物を含む映像フレームを受け取った場合に、AR の対象空間内における、追加された対象物の位置を推定する。学習モデル構築部 11 は、追加された対象物の位置を含む分割空間を特定する。学習モデル構築部 11 は、特定した分割空間に関連付けられた学習モデルを破棄し、特定した分割空間毎に、追加された対象物を認識するための学習モデルを再構築する。

【0048】

また、学習モデル構築部 11 は、AR 端末の周囲の環境が変化した場合にも、学習モデルの再学習を行う。

【0049】

AR 端末が取得した映像と、準備フェーズにおいて映像取得部 21 が取得した映像とを比較し、それぞれの映像における特徴点の差分が大きい場合、学習モデル構築部 11 は、AR 端末が取得した映像に含まれる対象物の位置を含む分割空間を特定する。それぞれの映像における特徴点は、対象物以外の背景における特徴点であってもよく、対象物における特徴点であってもよい。学習モデル構築部 11 は、特定した分割空間に関連付けられた学習モデルを破棄し、特定した分割空間毎に、対象物を認識するための学習モデルを再構築する。

【0050】

続いて、図 5 を用いて実施の形態 2 にかかる画像処理装置 20 において実行される準備フェーズにおける処理の流れについて説明する。

【0051】

はじめに、位置推定部 23 は、映像取得部 21 から受け取った映像フレームを用いて、AR の対象空間内における、AR により仮想オブジェクトが付加される対象物の位置を推定する (S11)。ここで、例えば、画像処理装置 20 の管理者等は、AR により仮想オブジェクトが付加される対象物に対して、対象物の属性、名称等や映像フレーム上での対象物の位置を特定するためのメタデータを付与する。メタデータは、例えば、物の名称、動物

10

20

30

40

50

の名称、もしくは構造物の名称等や映像フレーム上での対象物の位置を四角形の4頂点のXY座標で示したものであってもよい。管理者等は、画像処理装置20に、メタデータを入力し、位置推定部23は、映像フレームと推定した対象物の位置およびメタデータとを関連づける。

【0052】

次に、学習モデル構築部11は、ARの対象空間の分割数Dの初期値を設定する(S12)。初期値は、予め定められた値が画像処理装置20内のメモリ等に格納されている場合、予め定められた値を初期値として設定する。もしくは、学習モデル構築部11は、分割数Dを決定する際に、画像処理装置20の管理者等から初期値の入力を受け付けてもよい。

【0053】

次に、学習モデル構築部11は、設定された分割数Dに従い、対象空間をD個に分割する(S13)。つまり、学習モデル構築部11は、D個の分割空間を生成する。

【0054】

次に、学習モデル構築部11は、D個の分割空間のそれぞれに含まれる対象物を認識するための、学習モデルを構築し、それぞれの分割空間と、それぞれの分割空間に含まれる対象物を認識するための学習モデルとを紐づけて記録する(S14)。学習モデル構築部11は、例えば、分割空間と学習モデルとを紐づけて、学習モデル管理部12へ記録する。例えば、学習モデル構築部11は、それぞれの分割空間に含まれる対象物の映像と、対象物のメタデータとを学習器へ入力し、対象物を認識するための学習モデルを構築する。学習モデル構築部11は、対象物の映像を、対象物の位置情報を含む分割空間の学習モデルを構築する学習器へ入力してもよい。

【0055】

次に、学習モデル構築部11は、対象空間の分割数Dが規定値であるか否かを判定する(S15)。例えば、規定値が1である場合、学習モデル構築部11は、分割数Dが1であるか否かを判定する。学習モデル構築部11は、分割数Dが規定値である場合、処理を終了する。学習モデル構築部11は、分割数Dが規定値ではない場合、つまり、規定値よりも大きい値である場合、現在の分割数Dの値よりも小さい正の整数値を決定し、決定した値を新たな分割数Dとして設定する(S16)。学習モデル構築部11は、ステップS16の処理の後、ステップS13以降の処理を繰り返す。

【0056】

ステップS15において、学習モデル構築部11が決定する、現在の分割数Dの値よりも小さい正の整数値は、予め定められていてもよく、画像処理装置20の管理者等から入力された値が設定されてもよい。

【0057】

続いて、図6を用いて実施の形態2にかかる画像処理装置20において実行される実行フェーズにおける処理の流れについて説明する。

【0058】

はじめに、空間推定部13は、AR端末のカメラ画像及びセンサ情報取得部22から受け取ったセンサ情報を用いて、AR端末の位置及び傾きを推定する(S21)。カメラ画像は、映像と称されてもよい。次に、空間推定部13は、AR端末のカメラに映っているカメラ認識空間を推定する(S22)。例えば、空間推定部13は、複数の静止画データを用いるSLAM、AR端末のスペックとして定められている視野角、もしくは画角等に基づいて、AR端末のカメラに映っているカメラ認識空間を推定してもよい。

【0059】

次に、検出部14は、空間推定部13において推定されたカメラ認識空間の領域が含まれる少なくとも1つの分割空間に関連付けられた学習モデルを選択する(S23)。ここでは、検出部14は、カメラ認識空間の領域が複数の分割空間に含まれ、複数の学習モデルを選択したとする。次に、検出部14は、最も小さい領域をなす分割空間に関連付けられた学習モデルを選択する(S24)。

【0060】

次に、検出部 14 は、選択した学習モデルを用いて、AR 端末の映像に含まれる対象物を検出する（S25）。

【0061】

以上説明したように、画像処理装置 20 は、分割空間に含まれる対象物を検出するための学習モデルを構築することができる。分割空間に含まれる対象物を検出するための学習モデルは、対象空間全体に含まれる対象物を検出するための学習モデルよりも、学習するエリアが小さい。そのため、分割空間に含まれる対象物を検出するための学習モデルは、対象空間全体に含まれる対象物を検出するための学習モデルよりも、精度がよく、対象物を検出するために必要な映像フレームの数を減少させることができる。その結果、画像処理装置 20 は、AR により仮想オブジェクトを付加する対象物を検出するまでの時間を、対象空間全体に含まれる対象物を検出する場合の時間と比較して、短くすることができる。

10

【0062】

さらに、画像処理装置 20 は、AR 端末に設けられたカメラによって撮影された画像に対応するカメラ認識空間が複数の分割空間に含まれる場合、最も小さい領域をなす分割空間に関連付けられた学習モデルを選択する。最も小さい領域をなす分割空間に関連付けられた学習モデルは、大きい領域をなす分割空間に関連付けられた学習モデルよりも、学習するエリアが小さい。そのため、最も小さい領域をなす分割空間に関連付けられた学習モデルは、大きい領域をなす分割空間に関連付けられた学習モデルよりも、精度がよく、対象物を検出するために必要な映像フレームの数を減少させることができる。そのため、画像処理装置 20 は、最も小さい領域をなす分割空間に関連付けられた学習モデルを選択することによって、対象物を検出するまでの時間をさらに短くすることができる。

20

【0063】

また、実施の形態 2 においては、画像処理装置 20 が、学習モデル構築部 11、学習モデル管理部 12、空間推定部 13、検出部 14、映像取得部 21、センサ情報取得部 22、及び位置推定部 23 を有する構成を説明した。しかし、画像処理装置 20 に含まれる構成要素を、AR 端末 30 がすべて有し、AR 端末 30 が学習モデルの構築及び学習モデルを用いた対象物の検出を行ってもよい。

【0064】

もしくは、画像処理装置 20 の構成要素は、図 7 に示されるように配置されてもよい。図 7 は、画像処理装置 20 の構成要素が、AR 端末 40 及び画像処理装置 50 に分かれて配置されているシステム構成を示している。図 7 は、AR 端末 40 が、映像取得部 21 及びセンサ情報取得部 22 を有し、画像処理装置 50 が、その他の構成要素を有する構成を示している。

30

【0065】

もしくは、画像処理装置 20 の構成要素は、図 8 に示されるように配置されてもよい。図 8 は、画像処理装置 70 が、準備フェーズにおける処理を実行し、学習モデル配信部 71 を介して、AR 端末 60 へ構築した学習モデルを配信することを示している。さらに、図 8 は、AR 端末 60 が、実行フェーズにおける処理を実行することを示している。図 8 は、学習モデル受信部 61 が学習モデル配信部 71 から学習モデルを受信し、学習モデル管理部 62 に、受信した学習モデルを記録もしくは格納する。これにより、AR 端末 60 は、分割空間毎に関連付けられた学習モデルを保持する。空間推定部 13 及び、検出部 14 は、実行フェーズにおける処理を実行し、対象物の検出を行う。

40

【0066】

（実施の形態 3）

続いて、図 9 を用いて、実施の形態 3 にかかる学習モデルの概要について説明する。図 3 は、AR の対象空間を分割し、分割空間毎に学習モデルが構築されていることを示している。一方、図 9 は、AR により仮想オブジェクトが付加される対象物毎に学習モデルが構築されていることを示している。図 9 における円形の領域は、対象空間内に、対象物が位置する領域を示している。例えば、それぞれの学習モデルは、図 9 の円形の領域内に存在する一つの対象物を検出するために用いられてもよく、複数の対象物をセットとして検出す

50

るために用いられてもよい。

【0067】

続いて、図10及び図11を用いて、実施の形態3にかかる準備フェーズ及び実行フェーズの処理の流れについて説明する。実施の形態3にかかる準備フェーズ及び実行フェーズは、図2の画像処理装置20と同様の構成の画像処理装置において実行される。

【0068】

図10を用いて準備フェーズにおける処理について説明する。図10のステップS31は、図5のステップS11と同様であるため詳細な説明を省略する。

【0069】

次に、学習モデル構築部11は、位置推定部23において位置が推定された対象物を認識するための学習モデルを構築する(S32)。学習モデル構築部11は、対象物毎に、それぞれの対象物を認識するための学習モデルを構築する。学習モデル構築部11は、映像取得部21から受け取った対象物が主に表示された映像フレームを正解データもしくは教師データとし、対象物を認識するための学習モデルを構築する。さらに、学習モデル構築部11は、対象物近傍の映像を背景情報として、対象物とともに学習してもよい。

10

【0070】

学習モデル構築部11は、ステップS32において、構築した学習モデルと、学習モデルが認識する対象物の位置とを関連付けて、学習モデル管理部12に記録もしくは格納する。対象物の位置は、例えば、対象物の任意の部分を示す位置であってもよい。例えば、対象物の位置は、対象物の重心の位置であってもよい。もしくは、対象物の位置は、対象物の任意の部分を示す複数の位置を含んでもよい。

20

【0071】

次に、図11を用いて、実行フェーズにおける処理について説明する。図11のステップS41及びS42は、図6のステップS21及びS22と同様であるため詳細な説明を省略する。

【0072】

次に、検出部14は、空間推定部13において推定されたカメラ認識空間の領域に含まれる位置と関連付けられた学習モデルを選択する(S43)。検出部14は、空間推定部13において推定されたカメラ認識空間の領域に含まれる位置と関連付けられた学習モデルが複数存在する場合、複数の学習モデルのすべてを選択してもよく、1以上の任意の数の学習モデルを選択してもよい。

30

【0073】

検出部14は、選択した学習モデルを用いて、AR端末の映像に含まれる対象物を検出する(S44)。検出部14は、2以上の学習モデルを選択し、それぞれの学習モデルを用いて対象物を検出した場合、それぞれの学習モデルが検出結果と共に出力する信頼度が高い検出結果を採用する(S45)。信頼度が数値によって示される場合、検出部14は、信頼度の数値が閾値以上の学習モデルの検出結果を採用してもよい。もしくは、検出部14は、それぞれの学習モデルを用いて対象物の検出を行った際に、検出結果の信頼度が、閾値以上である信頼度の検出結果を採用してもよい。

【0074】

40

検出部14は、2以上の学習モデルを選択した場合、それぞれの学習モデルを順番に実行して、AR端末の映像に含まれる対象物を検出してもよい。もしくは、検出部14は、2以上の学習モデルを選択した場合、それぞれの学習モデルを並列に実行して、AR端末の映像に含まれる対象物を検出してもよい。

【0075】

画像処理装置20は、検出部14において選択された対象物に、予め対象物に関連付けられた仮想オブジェクトを付加して、AR端末に表示させる。例えば、画像処理装置20は、仮想オブジェクトが付加された映像をAR端末へ送信する。もしくは、画像処理装置20内において生成された、仮想オブジェクトが付加された映像が、映像取得部21へ出力される。

50

【0076】

以上説明したように、実施の形態3においては、AR端末のカメラがとらえている空間であるカメラ認識空間に含まれる位置に関連付けられた学習モデルを選択することができる。実施の形態3にかかる学習モデルは、対象空間内の特定の対象物を認識するために用いられる。そのため、特定の対象物を認識するために用いられる学習モデルを用いた検出処理は、対象空間内のすべての対象物を認識するために用いられる学習モデルを用いた検出処理と比較して、処理時間を短くすることができる。その結果、画像処理装置20は、ARにより仮想オブジェクトを付加する対象物を検出するまでの時間を、対象空間全体に含まれる対象物を検出する場合の時間と比較して、短くすることができる。

【0077】

(実施の形態4)

続いて、実施の形態4にかかる準備フェーズ及び実行フェーズの処理の流れについて説明する。実施の形態4にかかる準備フェーズ及び実行フェーズは、図2の画像処理装置20と同様の構成の画像処理装置において実行される。

【0078】

図12を用いて準備フェーズの処理の流れについて説明する。図12のステップS51は、図5のステップS11と同様であるため詳細な説明を省略する。

【0079】

次に、位置推定部23は、ARの対象空間におけるARにより仮想オブジェクトが付加される対象物の位置を可視化した2次元もしくは3次元データを作成する(S52)。図13に、対象物の位置を可視化した2次元データの例を示す。図13は、対象空間内における対象物の位置を黒丸にて示している。

【0080】

次に、位置推定部23は、ARにより仮想オブジェクトが付加される対象物の映像及び可視化したデータに対してオーグメンテーションを行う(S53)。具体的には、位置推定部23は、対象物が様々な方向、もしくは様々な距離離れた位置から撮影される可能性を考慮し、対象物が存在する可能性のある範囲を拡張して可視化する。例えば、図14は、対象空間内の対象物を、右上方向から見た場合に、対象物が存在する範囲として推定される位置を示している。また、図15は、対象空間内の対象物を右下方向から見た場合に、対象物が存在する範囲として推定される位置を示している。対象物が存在する範囲は、対象物を見る位置が、対象物から離れるにつれて、広がってもよい。対象物を見る位置が対象物に近い場合、対象物を見る位置が対象物から遠い場合よりも、より正確に対象物の位置を推定することができる。そのため、対象物を見る位置が対象物から離れるにつれて、対象物の位置の正確性が低くなり、対象物が存在する範囲が広がっていく。

【0081】

位置推定部23は、図14及び図15に示されている方向以外にも、様々な方向から対象物を見た場合における、対象物が存在する範囲を推定し、推定した範囲を可視化する。

【0082】

図12に戻り、学習モデル構築部11は、対象物の映像及び可視化したデータのセットと、対象物のメタデータとを学習器へ入力し、対象物及び対象物の位置を認識するための学習モデルを構築する(S54)。

【0083】

続いて、図16を用いて、実行フェーズの処理の流れについて説明する。図16のステップS61及びS62は、図6のステップS21及びS22と同様であるため詳細な説明を省略する。

【0084】

次に、空間推定部13は、空間に含まれている物の位置を可視化したデータを作成する(S63)。空間に含まれている物は、ARにより仮想オブジェクトが付加される対象物及び非対象物を含む。

【0085】

10

20

30

40

50

次に、検出部 14 は、AR 端末の映像及び空間に含まれている物の位置を可視化したデータのセットを学習モデルへ入力する（S64）。このようにして、検出部 14 は、AR 端末の映像に含まれる対象物を検出する。

【0086】

以上説明したように、実施の形態 4 にかかる学習モデルは、準備フェーズにおいて、対象物の映像とともに、対象物の対象空間における位置を入力して構築されている。そのため、実行フェーズにおいて、画像処理装置 20 は、AR 端末の映像とともに、可視化した対象物の位置を学習モデルに入力することによって、対象物の検出を行うことができる。つまり、画像処理装置 20 は、学習モデルの選択処理を行うことなく、対象物の検出を行うことができる。

【0087】

（実施の形態 5）

続いて、空間推定部 13 が実行する、実施の形態 5 にかかる方角推定処理について説明する。空間推定部 13 は、AR 端末の映像及びセンサ情報を用いて、AR 端末の位置及び傾きを推定し、AR 端末がとらえている空間を推定する。実施の形態 5 においては、空間推定部 13 は、さらに、AR 端末の映像のフレームから、物の切り出しを実行する。物の切り出しは、例えば、フレーム内の対象物及び非対象物を特定するとともに、フレーム内における物の相対位置を特定することであってもよい。フレーム内における切り出された物の相対位置は、例えば、フレーム内の物を囲った四角形の 4 つの頂点座標であってもよい。

【0088】

空間推定部 13 は、フレーム内の物の相対位置の情報を用いて、推定した空間内の物の方角もしくは位置を絞り込む。言い換えると、空間推定部 13 は、フレーム内の物の相対位置と、推定したカメラ認識空間の領域とから、物が存在する可能性のある空間の領域を推定する。

【0089】

検出部 14 は、空間推定部 13 において絞り込まれた方角の領域もしくは位置に関連付けられた学習モデルを選択する。

【0090】

以上説明したように、実施の形態 5 にかかる方角推定処理を行った場合、空間推定部 13 において推定した AR 端末がとらえている空間の領域を、さらに絞り込むことができる。例えば、空間推定部 13 において推定された AR 端末がとらえている空間に、同様の特徴を有する二つの対象物が含まれている場合であっても、空間の領域を絞り込むことによって、同様の特徴を有する二つの対象物を区別することができる。つまり、実施の形態 5 にかかる方角推定処理を行うことによって、検出部 14 は、同様の特徴を有する二つの対象物であっても、正しく対象物を検出することができる。

【0091】

実施の形態 5 にかかる方角推定処理は、例えば、図 6 の実行フェーズにおけるステップ S23、図 11 の実行フェーズにおけるステップ S43、及び図 16 のステップ S63 において実行されてもよい。

【0092】

上述したそれぞれの実施の形態においては、主に AR により仮想オブジェクトが付加される対象物が静止している場合について説明したが、対象物が移動する場合にもそれぞれの実施の形態を適用することができる。

【0093】

例えば、対象物が移動する場合、移動するエリアを指定し、移動するエリアに対応付けられた学習モデルを構築する際に、対象物を学習する。また、移動する対象物を検出する場合、移動する対象物を学習した学習モデルを常に用いてもよい。

【0094】

図 17 は、画像処理装置 10、画像処理装置 20、画像処理装置 50、及び画像処理装置 70（以下、画像処理装置 10 等と称する）の構成例を示すブロック図である。図 17

10

20

30

40

50

を参照すると、画像処理装置 10 等は、ネットワークインタフェース 1201、プロセッサ 1202、及びメモリ 1203を含む。ネットワークインタフェース 1201は、ネットワークノード（e.g., eNB、MME、P-GW、）と通信するために使用される。ネットワークインタフェース 1201は、例えば、IEEE 802.3 seriesに準拠したネットワークインタフェースカード（NIC）を含んでもよい。ここで、eNBはevolved Node B、MMEはMobility Management Entity、P-GWはPacket Data Network Gatewayを表す。IEEEは、Institute of Electrical and Electronics Engineersを表す。

【0095】

プロセッサ 1202は、メモリ 1203からソフトウェア（コンピュータプログラム）を読み出して実行することで、上述の実施形態においてフローチャートを用いて説明された画像処理装置 10等の処理を行う。プロセッサ 1202は、例えば、マイクロプロセッサ、MPU、又はCPUであってもよい。プロセッサ 1202は、複数のプロセッサを含んでもよい。

10

【0096】

メモリ 1203は、揮発性メモリ及び不揮発性メモリの組み合わせによって構成される。メモリ 1203は、プロセッサ 1202から離れて配置されたストレージを含んでもよい。この場合、プロセッサ 1202は、図示されていないI/O（Input/Output）インタフェースを介してメモリ 1203にアクセスしてもよい。

【0097】

図 17 の例では、メモリ 1203は、ソフトウェアモジュール群を格納するために使用される。プロセッサ 1202は、これらのソフトウェアモジュール群をメモリ 1203から読み出して実行することで、上述の実施形態において説明された画像処理装置 10等の処理を行うことができる。

20

【0098】

図 17 を用いて説明したように、上述の実施形態における画像処理装置 10等が有するプロセッサの各々は、図面を用いて説明されたアルゴリズムをコンピュータに行わせるための命令群を含む 1 又は複数のプログラムを実行する。

【0099】

上述の例において、プログラムは、様々なタイプの非一時的なコンピュータ可読媒体（non-transitory computer readable medium）を用いて格納され、コンピュータに供給することができる。非一時的なコンピュータ可読媒体は、様々なタイプの実体のある記録媒体（tangible storage medium）を含む。非一時的なコンピュータ可読媒体の例は、磁気記録媒体（例えばフレキシブルディスク、磁気テープ、ハードディスクドライブ）、光磁気記録媒体（例えば光磁気ディスク）、CD-ROM（Read Only Memory）、CD-R、CD-R/W、半導体メモリ（例えば、マスクROM、PROM（Programmable ROM）、EPROM（Erasable PROM）、フラッシュROM、RAM（Random Access Memory））を含む。また、プログラムは、様々なタイプの一時的なコンピュータ可読媒体（transitory computer readable medium）によってコンピュータに供給されてもよい。一時的なコンピュータ可読媒体の例は、電気信号、光信号、及び電磁波を含む。一時的なコンピュータ可読媒体は、電線及び光ファイバ等の有線通信路、又は無線通信路を介して、プログラムをコンピュータに供給できる。

30

40

【0100】

なお、本開示は上記実施の形態に限られたものではなく、趣旨を逸脱しない範囲で適宜変更することが可能である。

【0101】

上記の実施形態の一部又は全部は、以下の付記のようにも記載されうるが、以下には限られない。

（付記 1）

対象空間を 1 以上の空間に分割した分割空間を生成し、前記分割空間に含まれる少なくとも 1 以上の対象物であって、前記対象物を認識するための学習モデルを構築する学習モ

50

デル構築部と、

前記学習モデルと、前記学習モデルによって認識される前記対象物を含む前記分割空間をなす領域とを関連付けて管理する学習モデル管理部と、

UIデバイスに設けられたカメラがとらえているカメラ認識空間をなす領域を推定する空間推定部と、

前記学習モデル管理部によって管理されている前記学習モデルの中から、前記推定されたカメラ認識空間をなす領域を含む前記分割空間をなす領域が関連付けられている特定の学習モデルを選択し、前記選択した特定の学習モデルを用いて、前記UIデバイスに表示された空間に含まれる前記対象物を検出する検出部と、を備える画像処理装置。

(付記 2)

前記学習モデル構築部は、

前記対象物の位置情報を用いて、前記対象物が含まれる前記分割空間を特定する、付記 1 に記載の画像処理装置。

(付記 3)

前記学習モデル構築部は、

前記対象空間を N (N は 1 以上の整数) 分割して、 N 分割された前記分割空間に含まれる対象物を認識するための学習モデルを構築した後に、前記対象空間を M (1 以上であって N よりも小さい整数) 分割し、 M 分割された前記分割空間に含まれる対象物を認識するための学習モデルを構築する、付記 1 又は 2 に記載の画像処理装置。

(付記 4)

前記検出部は、

前記推定された領域を含む前記分割空間が複数存在する場合、前記分割空間をなす領域が、予め定められた基準値よりも小さい分割空間を選択する、付記 3 に記載の画像処理装置。

(付記 5)

前記検出部は、

前記推定された領域を含む前記分割空間が複数存在する場合、前記分割空間の領域が最も小さい分割空間を選択する、付記 3 又は 4 に記載の画像処理装置。

(付記 6)

前記空間推定部は、

センサによって検出された前記カメラの加速度及び傾きに関する情報を用いて、前記カメラで撮影された空間をなす領域を推定する、付記 1 乃至 5 のいずれか 1 項に記載の画像処理装置。

(付記 7)

対象物を含む映像を用いて、対象空間における前記対象物の位置を推定する位置推定部と、

前記対象物を認識するための学習モデルを構築する学習モデル構築部と、

前記学習モデル、前記学習モデルにおいて認識される前記対象物、及び前記推定された対象物の位置を関連付けて管理する学習モデル管理部と、

UIデバイスに設けられたカメラがとらえている空間の領域を推定する空間推定部と、

前記学習モデル管理部によって管理されている前記学習モデルの中から、前記領域に含まれる位置が関連付けられている特定の学習モデルを選択し、前記選択した特定の学習モデルを用いて、前記領域に含まれる前記対象物の検出を行う検出部と、を備える画像処理装置。

(付記 8)

前記空間推定部は、

前記UIデバイスに表示された空間の映像から前記対象物の切り出しを行い、切り出された前記対象物の前記映像における位置と、推定した前記空間の領域とを用いて、前記空間の領域内における、前記対象物の存在する位置を推定する、付記 1 乃至 7 のいずれか 1 項に記載の画像処理装置。

10

20

30

40

50

(付記 9)

対象物を含む映像を用いて、対象空間における前記対象物の位置を推定する位置推定部と、

前記対象物及び前記対象物の位置を入力とし、前記対象物を認識するための学習モデルを構築する学習モデル構築部と、

UIデバイスに設けられたカメラにより取得した映像と、センサによって検出された前記カメラの位置情報とを前記学習モデルに入力し、前記カメラで撮影された映像に含まれる前記対象物の検出を行う検出部と、を備える画像処理装置。

(付記 10)

UIデバイスに設けられた前記カメラがとらえている空間の領域を推定する空間推定部をさらに備え、

前記空間推定部は、

前記カメラがとらえている空間の映像から前記対象物の切り出しを行い、切り出された前記対象物の前記映像における位置と、推定した前記空間をなす領域とを用いて、前記空間をなす領域内における、前記対象物の存在する位置を推定する、付記 9 に記載の画像処理装置。

(付記 11)

前記空間推定部は、

前記カメラがとらえている空間の映像から前記対象物の切り出しを行い、切り出された前記対象物の前記映像における位置と、推定した前記空間をなす領域とを用いて、前記空間をなす領域内における、前記対象物の存在する位置を推定する、付記 1 乃至 8 のいずれか 1 項に記載の画像処理装置。

(付記 12)

対象空間を 1 以上の空間に分割した分割空間を生成し、前記分割空間に含まれる少なくとも 1 以上の対象物であって、前記対象物を認識するための学習モデルを構築し、

前記学習モデルと、前記学習モデルによって認識される前記対象物を含む前記分割空間をなす領域とを関連付けて管理し、

UIデバイスに設けられたカメラがとらえているカメラ認識空間の領域を推定し、

前記管理している前記学習モデルの中から、前記推定されたカメラ認識空間をなす領域を含む前記分割空間をなす領域が関連付けられている特定の学習モデルを選択し、前記選択した特定の学習モデルを用いて、前記UIデバイスに表示された空間に含まれる前記対象物の検出を行う、検出方法。

(付記 13)

対象物を含む映像を用いて、対象空間における前記対象物の位置を推定し、

前記対象物を認識するための学習モデルを構築し、

前記学習モデル、前記学習モデルにおいて認識される前記対象物、及び前記推定された対象物の位置を関連付けて管理し、

UIデバイスに設けられたカメラがとらえている空間の領域を推定し、

前記管理している前記学習モデルの中から、前記領域に含まれる位置と関連付けられた前記学習モデルを選択し、前記選択した特定の学習モデルを用いて、前記領域に含まれる前記対象物の検出を行う、検出方法。

(付記 14)

対象物を含む映像を用いて、対象空間における前記対象物の位置を推定し、

前記対象物及び前記対象物の位置を入力とし、前記対象物を認識するための学習モデルを構築し、

UIデバイスに設けられたカメラで撮影された映像と、センサによって検出された前記カメラの位置情報とを前記学習モデルに入力し、前記UIデバイスに表示された映像に含まれる前記対象物の検出を行う、検出方法。

(付記 15)

対象空間を 1 以上の空間に分割した分割空間を生成し、前記分割空間に含まれる少なく

10

20

30

40

50

とも1以上の対象物であって、前記対象物を認識するための学習モデルを構築し、

前記学習モデルと、前記学習モデルによって認識される前記対象物を含む前記分割空間をなす領域とを関連付けて管理し、

UIデバイスに設けられたカメラがとらえているカメラ認識空間をなす領域を推定し、

前記管理している前記学習モデルの中から、前記推定されたカメラ認識空間をなす領域を含む前記分割空間をなす領域が関連付けられている特定の学習モデルを選択し、前記選択した特定の学習モデルを用いて、前記UIデバイスに表示された空間に含まれる前記対象物を検出することをコンピュータに実行させるプログラム。

(付記16)

対象物を含む映像を用いて、対象空間における前記対象物の位置を推定し、

前記対象物を認識するための学習モデルを構築し、

前記学習モデル、前記学習モデルにおいて認識される前記対象物、及び推定された前記対象物の位置を関連付けて管理し、

UIデバイスに設けられたカメラがとらえている空間の領域を推定し、

前記管理されている前記学習モデルの中から、前記領域に含まれる位置が関連付けられている特定の学習モデルを選択し、前記領域に含まれる前記対象物を検出することをコンピュータに実行させるプログラム。

(付記17)

対象物を含む映像を用いて、対象空間における前記対象物の位置を推定し、

前記対象物及び前記対象物の位置を入力とし、前記対象物を認識するための学習モデルを構築し、

UIデバイスに設けられたカメラで撮影された映像と、センサによって検出された前記カメラの位置情報とを前記学習モデルに入力し、前記カメラで撮影された映像に含まれる前記対象物の検出を行うことをコンピュータに実行させるプログラム。

【0102】

なお、本発明は上記実施の形態に限られたものではなく、趣旨を逸脱しない範囲で適宜変更することが可能である。

【符号の説明】

【0103】

10 画像処理装置

11 学習モデル構築部

12 学習モデル管理部

13 空間推定部

14 検出部

15 UIデバイス

16 カメラ

20 画像処理装置

21 映像取得部

22 センサ情報取得部

23 位置推定部

30 AR端末

40 AR端末

50 画像処理装置

60 AR端末

61 学習モデル受信部

62 学習モデル管理部

70 画像処理装置

71 学習モデル配信部

10

20

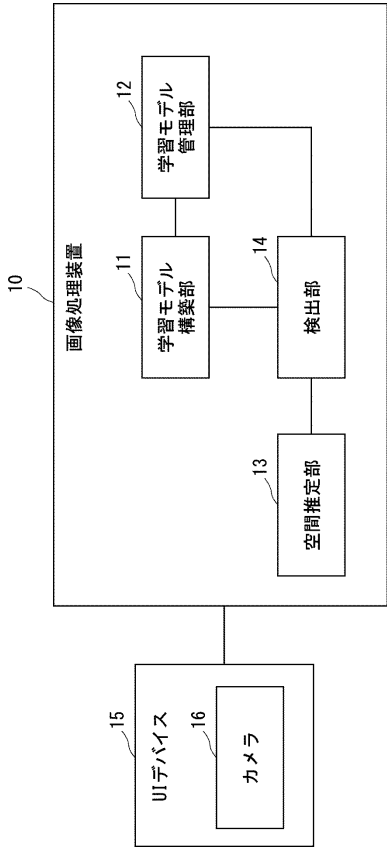
30

40

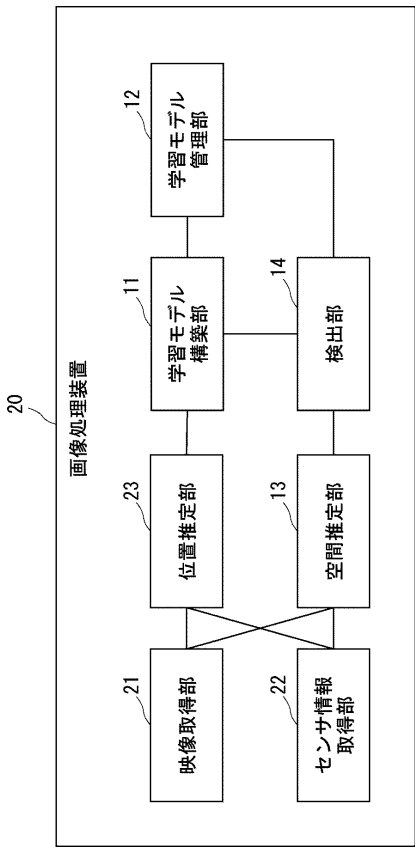
50

【図面】

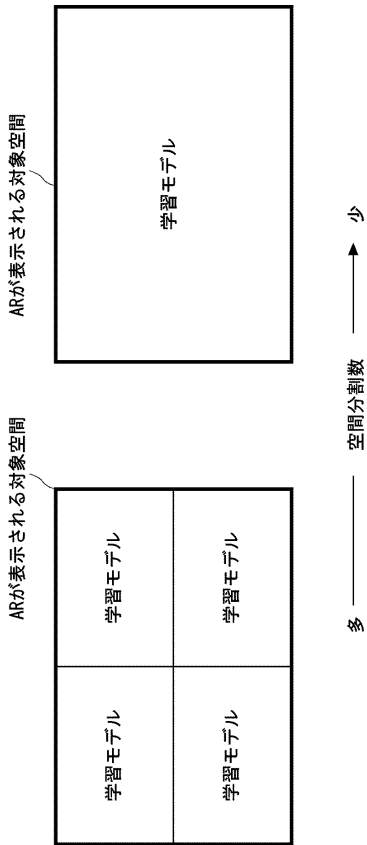
【図 1】



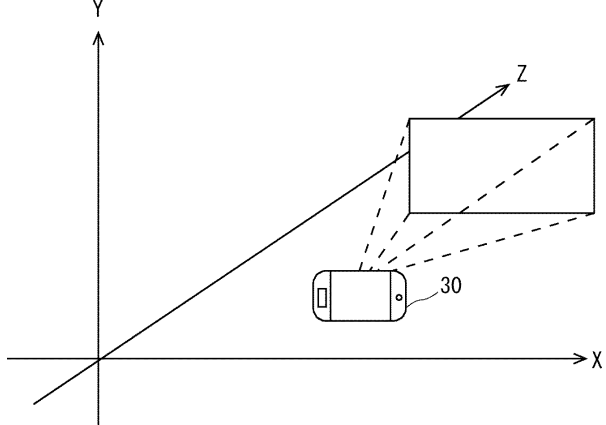
【図 2】



【図 3】



【図 4】



10

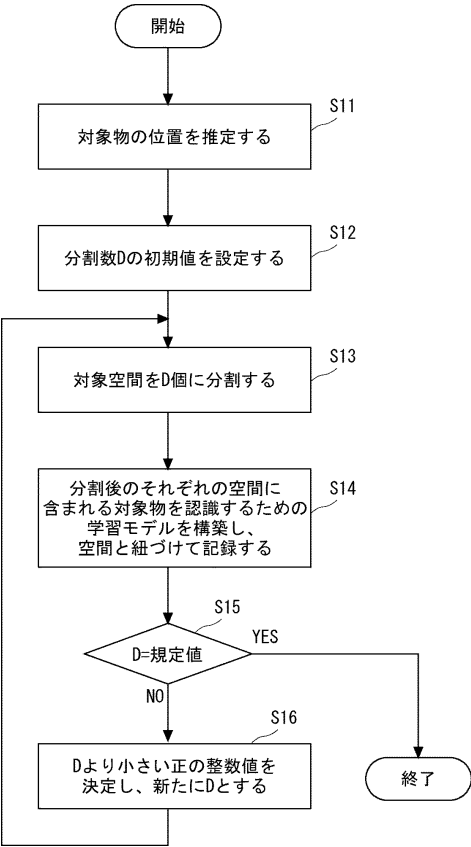
20

30

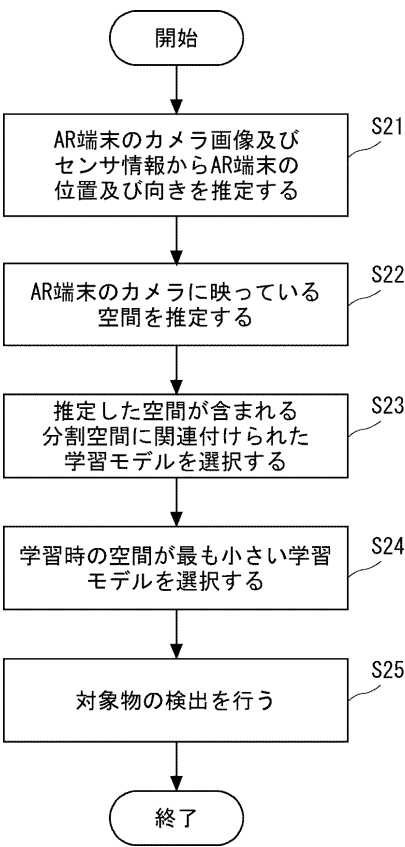
40

50

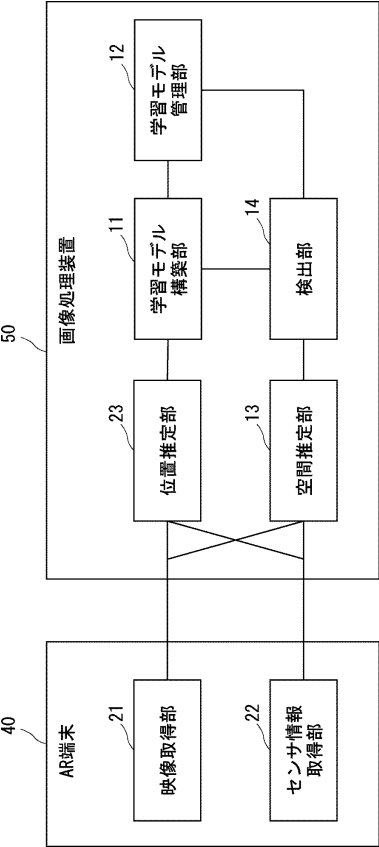
【図 5】



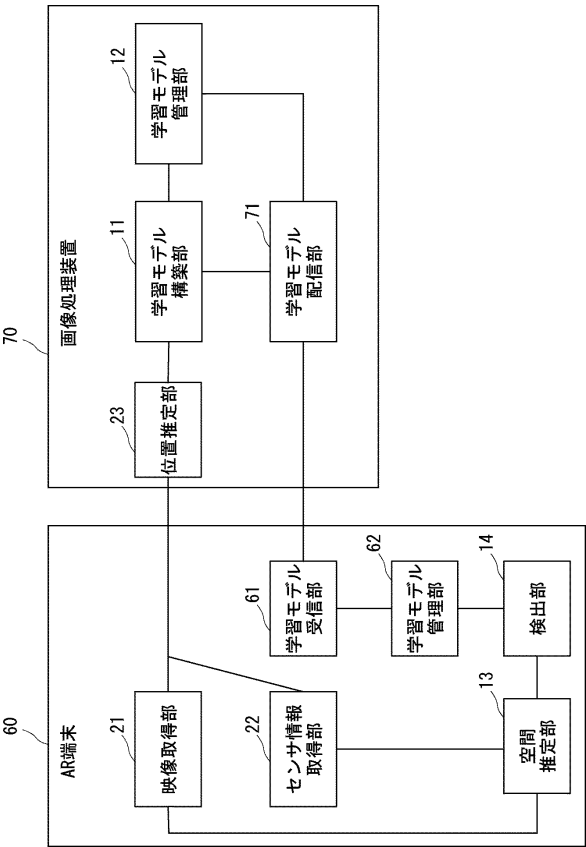
【図 6】



【図 7】



【図 8】



10

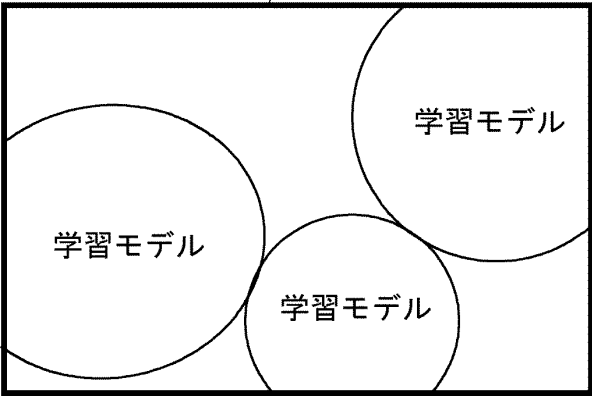
20

30

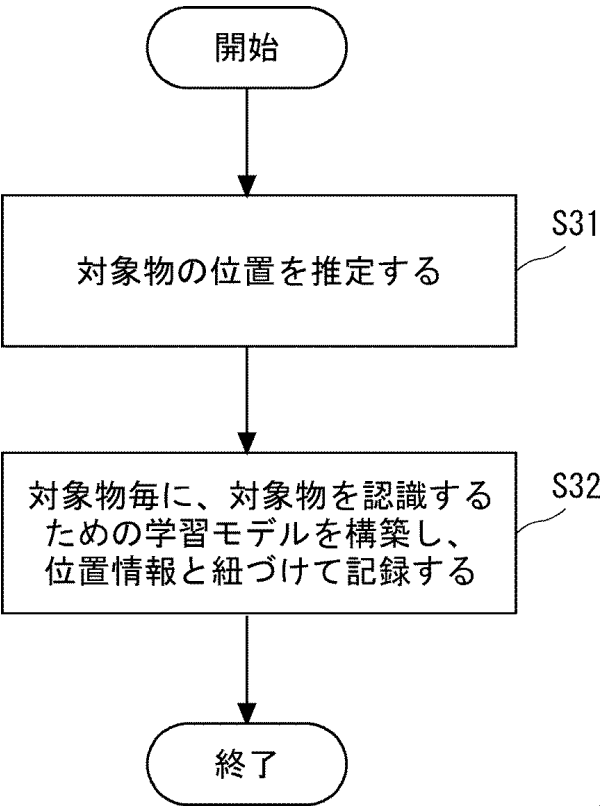
40

50

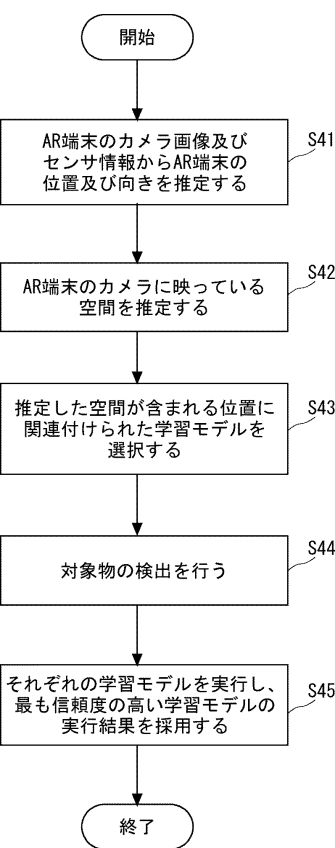
【図 9】
ARが表示される対象空間



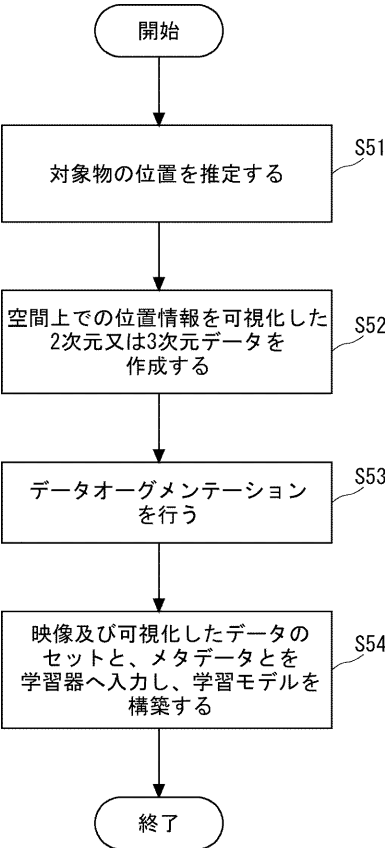
【図 10】



【図 11】



【図 12】



10

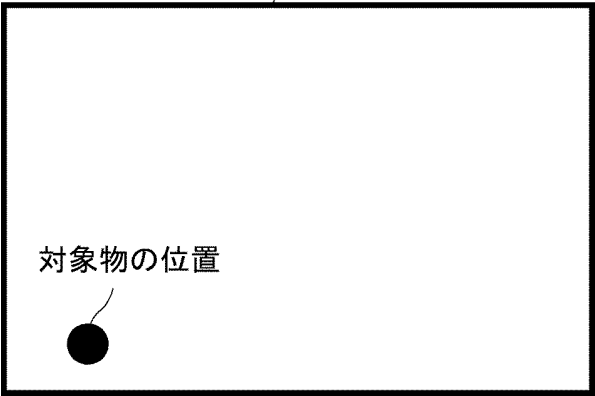
20

30

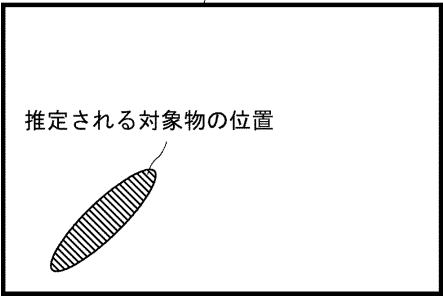
40

50

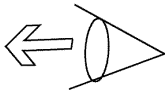
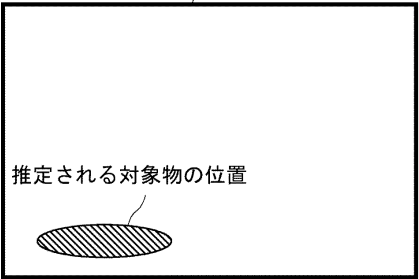
【図 1 3】
ARが表示される対象空間



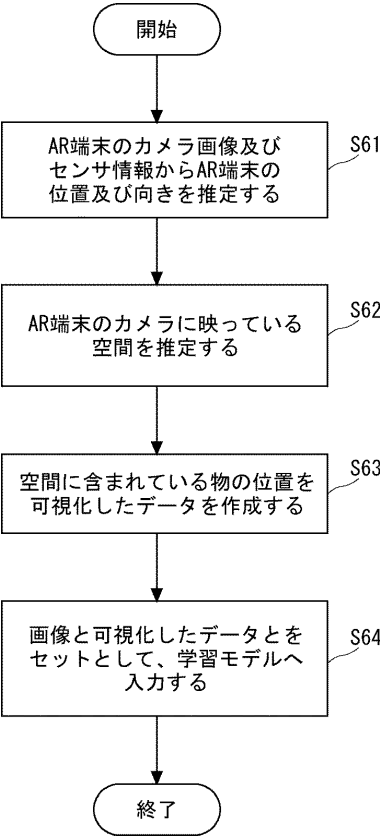
【図 1 4】
ARが表示される対象空間



【図 1 5】
ARが表示される対象空間



【図 1 6】



10

20

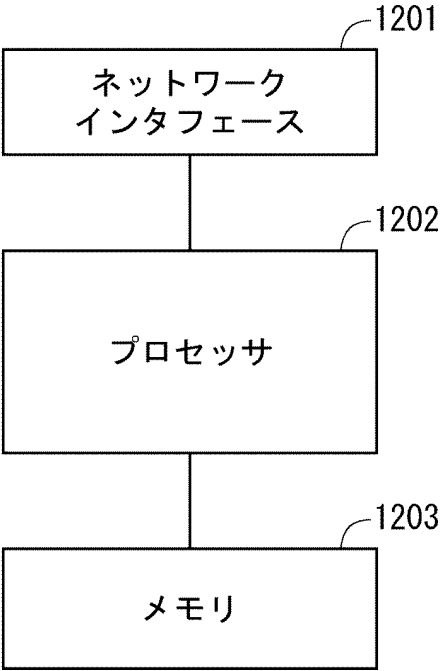
30

40

50

【図 17】

10, 20, 50, 70



10

20

30

40

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 2 0 1 8 / 0 0 1 2 4 1 1 (U S , A 1)
特開 2 0 1 9 - 0 2 9 0 2 1 (J P , A)
特開 2 0 1 9 - 1 3 9 3 4 6 (J P , A)
特開 2 0 1 6 - 1 5 7 2 5 8 (J P , A)
特開 2 0 2 1 - 0 6 0 2 9 4 (J P , A)
Kuan-Hsun Wang;Shang-Hong Lai, Object Detection in Curved Space for 360-Degree Camera, 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2019年05月12日, 3642-3646, <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8683093>
Joris Guerin;Olivier Gibaru;Eric Nyiri;Stephane Thieryl;Byron Boots, Semantically Meaningful View Selection, 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), IEEE, 2018年10月01日, 1061-1066, <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8593524>
視覚マーカを用いた物体検出のための偏りのない学習データ自動収集システム, ロボティクスメカトロニクス講演会 2 0 1 8 講演会論文集 2 A2-J18, 一般社団法人日本機械学会, 2018年06月01日

(58)調査した分野 (Int.Cl., D B 名)
G 0 6 T 7 / 7 0
G 0 6 T 7 / 0 0