

(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
2023年9月21日(21.09.2023)



(10) 国際公開番号

WO 2023/175727 A1

(51) 国際特許分類:

G06N 20/00 (2019.01)

(21) 国際出願番号: PCT/JP2022/011645

(22) 国際出願日: 2022年3月15日(15.03.2022)

(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: 佐藤 祥悟 (SATO, Shogo); 〒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). 稲田 徹悟 (INADA, Tetsugo); 〒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). 勢川 博之 (SEGAWA, 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).

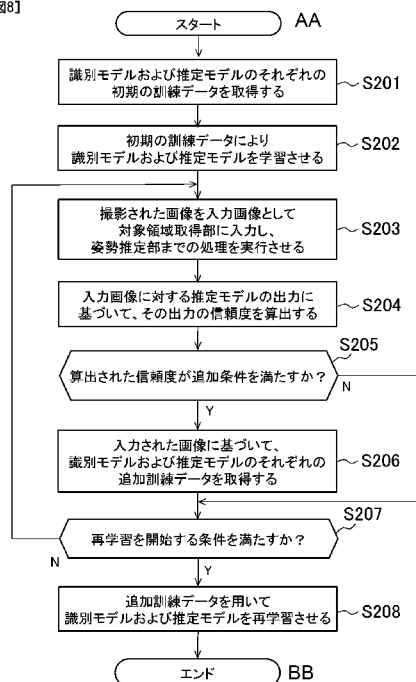
(74) 代理人: 弁理士法人はるか国際特許事務所 (HARUKA PATENT & TRADEMARK ATTORNEYS); 〒1020085 東京都千代田区六番町3六番町SKビル5階 Tokyo (JP).

(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, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL,

(54) Title: INFORMATION PROCESSING SYSTEM, INFORMATION PROCESSING METHOD, AND PROGRAM

(54) 発明の名称: 情報処理システム、情報処理方法及びプログラム

〔図8〕



S201... Acquire initial training data of each of discrimination model and estimation model
S202... Train discrimination model and estimation model by initial training data
S203... Input captured image to target region acquisition unit as input image to execute processing up to posture estimation unit
S204... Use output of estimation model for input image to calculate reliability of output
S205... Does calculated reliability satisfy additional condition?
S206... Acquire additional training data of each of discrimination model and estimation model on basis of input image
S207... Is condition for starting retraining satisfied?
S208... Retrain discrimination model and estimation model using additional training data
AA ... Start
BB ... End

(57) Abstract: The present invention improves the accuracy of a machine learning model while facilitating organization of training data. This information processing system refers to an output, which is produced by a machine learning model trained by training data in response to input of input data to the machine learning model, so as to output the reliability of the output for the input data, generates new training data on the basis of the input data if the reliability satisfies a predetermined condition, and trains the machine learning model by the new training data.

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, IT, JM, 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, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, 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, SC, 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:

- 国際調査報告 (条約第21条(3))
 - 補正された請求の範囲及び説明書 (条約第19条(1))
-

(57) 要約: 訓練データの整備を容易にしつつ機械学習モデルの精度を向上させる。情報処理システムは、訓練データにより学習された機械学習モデルに入力データが入力された際の当該機械学習モデルの出力に基づいて、前記入力データに対する当該出力の信頼度を出力し、前記信頼度が所定の条件を満たす場合に、前記入力データに基づいて新たな訓練データを生成し、前記新たな訓練データにより機械学習モデルを学習させる。

明 細 書

発明の名称： 情報処理システム、情報処理方法及びプログラム
技術分野

[0001] 本発明は、情報処理システム、情報処理方法及びプログラムに関する。

背景技術

[0002] 一般的な機械学習モデルは、予め準備された訓練データにより学習される。

[0003] Sida Peng et alは、2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)において、論文PVNet: Pixel-Wise Voting Network for 6DoF Pose Estimationを発表している。この論文では、入力画像と、正解の出力画像とを含む訓練データにより機械学習モデルを学習させ、さらにその機械学習モデルに撮影された画像が入力された際の出力に基づいて姿勢推定に用いるキーポイントの画像上の位置を算出することが開示されている。

発明の概要

発明が解決しようとする課題

[0004] 機械学習モデルを学習させるためには多量の訓練データが必要となるが、そのデータを準備するには多くの手間がかかる。一方、訓練データの量を減らすと、機械学習モデルの精度が確保できない恐れがある。

[0005] 本発明は上記実情に鑑みてなされたものであって、その目的は、訓練データの整備にかかる手間を抑制しつつ機械学習モデルの精度を向上させる技術を提供することにある。

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、本発明に係る情報処理システムは、訓練データにより学習された機械学習モデルと、入力データが入力された際の前記機械学習モデルの出力に基づいて、前記入力データに対する当該出力の信頼度を出力する信頼度出力手段と、前記信頼度が所定の条件を満たす場合に、前

記入入力データに基づいて新たな訓練データを生成する生成手段と、前記新たな訓練データにより機械学習モデルを学習させる学習制御手段と、を含む。

[0007] 本発明の一態様では、情報処理システムは、前記機械学習モデルの出力に基づいて推定結果を出力する学習済の推定モデルを含み、前記信頼度出力手段は、前記推定モデルの出力に基づいて前記入力データに対する前記機械学習モデルの出力の信頼度を出力してよい。

[0008] 本発明の一態様では、前記入力データは、対象オブジェクトが撮影された画像を含み、前記推定モデルは、前記機械学習モデルの出力に基づいて、前記対象オブジェクトの姿勢推定のためのキーポイントを示す画像を出力し、前記信頼度出力手段は、前記画像に基づいて前記信頼度を出力してよい。

[0009] 本発明の一態様では、前記推定モデルは、各点がキーポイントとの位置関係を示す画像を出力し、前記信頼度出力手段は、複数のキーポイントの位置の候補であってそれぞれ前記推定モデルが出力した画像に含まれる互いに異なる点から生成されるキーポイントの位置の候補のばらつきに基づいて、前記信頼度を出力してよい。

[0010] 本発明の一態様では、前記信頼度出力手段は、対象オブジェクトが撮影された入力画像が前記推定モデルに入力された際の当該推定モデルの出力と、前記入力画像が所定の加工処理により加工された加工画像が前記推定モデルに入力された際の出力との相違を示す情報に基づいて、前記信頼度を出力してよい。

[0011] 本発明の一態様では、前記機械学習モデルは前記入力データが前記対象オブジェクトを含むか否かを示す情報を出力してよい。

[0012] 本発明の一態様では、前記推定モデルは推定訓練データにより学習され、前記生成手段は、前記信頼度が前記所定の条件を満たす場合に、前記入力データに基づいて新たな推定訓練データを生成し、前記学習制御手段は、前記新たな訓練データにより機械学習モデルを学習させてよい。

[0013] 本発明の一態様では、前記入力データは、対象オブジェクトが撮影された画像を含み、前記機械学習モデルは、前記入力データに基づいて、対象オブ

ジェクトの姿勢推定のためのキーポイントを示す画像を出力し、前記信頼度出力手段は、前記画像に基づいて前記信頼度を出力させてよい。

[0014] 本発明の一態様では、前記訓練データは3次元形状モデルからレンダリングされた複数の学習画像とそれぞれが前記学習画像に対する正解データである正解画像とを含んでよい。

[0015] 本発明の一態様では、前記生成手段は、前記入力データが第1の加工処理により加工された第1の追加画像と、前記入力データが前記第1の加工処理と異なる第2の加工処理により加工された第2の追加画像とを含む新たな訓練データを生成し、前記学習制御手段は、前記機械学習モデルに前記第1の追加画像を入力した際の出力と、前記機械学習モデルに前記第2の追加画像を追加した際の出力との違いに基づいて、機械学習モデルを学習させてよい。

[0016] また、本発明に係る情報処理方法は、訓練データにより学習された機械学習モデルに入力データが入力された際の当該機械学習モデルの出力に基づいて、前記入力データに対する当該出力の信頼度を出力するステップと、前記信頼度が所定の条件を満たす場合に、前記入力データに基づいて新たな訓練データを生成するステップと、前記新たな訓練データにより機械学習モデルを学習させるステップと、を含む。

[0017] また、本発明に係るプログラムは、訓練データにより学習された機械学習モデルに入力データが入力された際の当該機械学習モデルの出力に基づいて、前記入力データに対する当該出力の信頼度を出力し、前記信頼度が所定の条件を満たす場合に、前記入力データに基づいて新たな訓練データを生成し、前記新たな訓練データにより機械学習モデルを学習させる、処理をコンピュータに実行させる。

発明の効果

[0018] 本発明によれば、訓練データの整備にかかる手間を抑制しつつ機械学習モデルの精度を向上させることができる。

図面の簡単な説明

[0019] [図1]本発明の一実施形態に係る情報処理システムの構成の一例を示す図である。

[図2]本発明の一実施形態に係る情報処理システム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]入力画像の一例を示す図である。

[図4]対象オブジェクトのキーポイントの一例を示す図である。

[図5]対象領域における位置画像の一例を模式的に示す図である。

[図6]主に対象領域取得部および姿勢推定部の処理の一例を示すフロー図である。

[図7]検出された対象オブジェクトの姿勢を説明する図である。

[図8]識別モデルおよび推定モデルの学習を概略的に説明するフロー図である。

[図9]初期の訓練データを生成する処理の一例を示すフロー図である。

[図10]対象オブジェクトの撮影を説明する図である。

[図11]推定モデルの再学習の処理の一例を示すフロー図である。

発明を実施するための形態

[0020] 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。本実施形態では、オブジェクトが撮影された画像を入力し、その姿勢を推定する情報処理システムに発明を適用した場合について説明する。

[0021] この情報処理システムは、撮影された画像の少なくとも一部がオブジェクトを含むか否かを判定する機械学習モデルと、オブジェクトを含む画像からそのオブジェクトの推定される姿勢を示す情報を出力する機械学習モデルとを含んでいる。また情報処理システムはその学習を短時間で完了するように構成されている。所要時間は、例えば、オブジェクトを把持して回転させるのに数十秒、機械学習に数分程度が想定されている。

[0022] 図1は、本発明の一実施形態にかかる情報処理システムの構成の一例を示す図である。本実施形態にかかる情報処理システムは、情報処理装置10を含む。情報処理装置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。図 1 に示すように、情報処理装置 10 は、例えば、プロセッサ 11、記憶部 12、通信部 14、操作部 16、表示部 18、撮影部 20を含んでいる。情報処理システムは 1 台の情報処理装置 10 により構成されてもよいし、情報処理装置 10 を含む複数の装置により構成されてもよい。

[0023] プロセッサ 11 は、例えば情報処理装置 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。

[0024] 記憶部 12 は、ROM や RAM 等の記憶素子やソリッドステートドライブのような外部記憶装置のうち少なくとも一部からなる。記憶部 12 には、プロセッサ 11 によって実行されるプログラムなどが記憶される。

[0025] 通信部 14 は、例えばネットワークインタフェースカードのような、有線通信又は無線通信の通信インタフェースであり、インターネット等のコンピュータネットワークを介して、他のコンピュータや端末との間でデータを授受する。

[0026] 操作部 16 は、例えば、キーボード、マウス、タッチパネル、ゲームコンソールのコントローラ等の入力デバイスであって、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 11 に出力する。

[0027] 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 11 の指示に従って各種の画像を表示する。表示部 18 は、外部の表示デバイスに対して映像信号を出力するデバイスであってもよい。

[0028] 撮影部 20 は、デジタルカメラ等の撮影デバイスである。本実施形態にかかる撮影部 20 は、例えば動画像の撮影が可能なカメラである。撮影部 20 は、可視の RGB 画像を取得可能なカメラであってよい。撮影部 20 は、可視の RGB 画像と、その RGB 画像と同期した深度情報とを取得可能なカメラであってもよい。撮影部 20 は情報処理装置 10 の外部にあってもよく、この場合は情報処理装置 10 と撮影部 20 とが、通信部 14 または後述の入出力部を介して接続されてよい。

[0029] なお、情報処理装置 10 は、マイクやスピーカなどといった音声入出力デ

バイスを含んでいてもよい。また、情報処理装置 10 は、例えば、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、外部機器とデータの入出力をするための入出力部 (USB (Universal Serial Bus) ポート) を含んでいてもよい。

[0030] 図 2 は、本発明の一実施形態に係る情報処理システム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 2 に示すように、情報処理システムは、機能的に、対象領域取得部 21、姿勢推定部 25、撮影画像取得部 33、識別訓練データ生成部 34、識別学習部 35、形状モデル取得部 36、推定訓練データ生成部 37、推定学習部 38、信頼度取得部 39 を含む。対象領域取得部 21 は、機能的に、領域抽出部 22、特徴抽出部 23、および識別モデル 24 を含む。姿勢推定部 25 は、機能的に、推定モデル 26、キーポイント決定部 27、および姿勢算出部 28 を含む。識別モデル 24 および推定モデル 26 は、どちらも機械学習モデルの一種である。

[0031] これらの機能は、主にプロセッサ 11 及び記憶部 12 により実装される。より具体的には、これらの機能は、コンピュータである情報処理装置 10 にインストールされた、以上の機能に対応する実行命令を含むプログラムをプロセッサ 11 で実行することにより実装されてよい。また、このプログラムは、例えば、光学的ディスク、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情報処理装置 10 に供給されてもよい。

[0032] なお、本実施形態にかかる情報処理システムに、必ずしも図 2 に示す機能のすべてが実装されていなくてもよく、また、図 2 に示す機能以外の機能が実装されていてもよい。

[0033] 対象領域取得部 21 は、撮影部 20 により撮影された入力画像を取得し、その取得された入力画像に含まれる 1 または複数の候補領域 56 (図 3 参照) のそれぞれが対象オブジェクト 51 の画像を含むか否か判定する。領域抽出部 22 は、この 1 または複数の候補領域 56 を抽出し、特徴抽出部 23 は

候補領域 5 6 のそれぞれから画像の特徴を示す特徴量を抽出する。識別モデル 2 4 には、その候補領域 5 6 の画像としてその特徴量が入力され、その候補領域 5 6 が対象オブジェクト 5 1 の画像を含むか否かを示す情報を出力する。

[0034] 対象領域取得部 2 1 は、候補領域 5 6 が対象オブジェクト 5 1 を含む場合に、入力画像から抽出される、対象オブジェクト 5 1 の画像を含む対象領域 5 5 を取得する。対象オブジェクト 5 1 は、情報処理装置 1 0 において姿勢の推定の対象となるオブジェクトである。対象オブジェクト 5 1 は事前の学習の対象となる。

[0035] 図 3 は、入力画像の一例を示す図である。図 3 の例では、対象オブジェクト 5 1 は電動工具であり、以降の図においても特に説明のない場合は対象オブジェクト 5 1 の例は電動工具であるとする。入力画像は、撮影部 2 0 により撮影されており、対象領域 5 5 は対象オブジェクト 5 1 およびその近傍を含む矩形の領域である。なお、対象領域 5 5 の取得の過程において、対象オブジェクト 5 1 を含む領域の候補として、対象オブジェクト 5 1 を含まない領域も含む 1 または複数の候補領域 5 6 も抽出される。

[0036] 領域抽出部 2 2 は、入力画像から、識別モデル 2 4 の判定の対象となる候補領域 5 6 の画像を抽出する。より具体的には、領域抽出部 2 2 は、公知の Region Proposal 技術により、入力画像から、何らかのオブジェクトが撮影された 1 または複数の候補領域 5 6 を識別し、その 1 または複数の候補領域 5 6 のそれぞれを抽出する。

[0037] 識別モデル 2 4 は機械学習モデルであり、訓練データにより学習され、学習済の識別モデル 2 4 は、入力データが入力されると、識別の結果としてデータを出力する。識別モデル 2 4 に入力される入力データは、候補領域 5 6 の画像を示す情報であり、例えば、特徴抽出部 2 3 がその画像から抽出した特徴量である。また、識別モデル 2 4 は、入力データが入力されると、その候補領域 5 6 の画像が対象オブジェクト 5 1 の画像を含むか否かを示す情報を出力する。

- [0038] 識別モデル24の訓練データは、対象オブジェクト51が撮影された画像を含む複数の正例画像と、対象オブジェクト51を含まない複数の負例画像とを含む学習画像のそれぞれを示すデータを含む。識別モデル24およびその学習の詳細については後述する。学習画像のそれぞれは、撮影された画像のうち対象オブジェクト51が存在する領域の画像であってよい。その領域の抽出は、領域抽出部22と同様の手法で行われてよい。なお、識別モデル24は、上記の訓練データだけでなく、追加の訓練データによっても学習される。
- [0039] なお、特徴抽出部23を介さずに候補領域56の画像が識別モデル24に直接的に入力されてもよい。精度が低下する恐れはあるが、領域抽出部22が存在しなくてもよい。この場合、特徴抽出部23が入力画像そのものから特徴を抽出し、識別モデル24がその入力画像に対象オブジェクト51が存在するか判定してもよいし、入力画像が直接的に識別モデル24に入力されてもよい。
- [0040] 姿勢推定部25は、推定モデル26に対象領域55が入力された際に出力される情報に基づいて、対象オブジェクト51の姿勢を推定する。推定モデル26は、機械学習モデルであり、訓練データにより学習され、学習済の推定モデル26は、入力データが入力されると、推定結果としてデータを出力する。訓練データは、対象オブジェクト51の3次元形状モデルによりレンダリングされた複数の学習画像とその学習画像における対象オブジェクト51の姿勢に関する情報である正解データとを含む。
- [0041] 学習済の推定モデル26には、対象領域55の画像を示す情報が入力され、推定モデル26は対象オブジェクトの姿勢推定のためのキーポイントの位置を示す情報を出力する。対象領域55は、識別モデル24の出力に基づいて選択された候補領域56に基づく画像である。推定モデル26の訓練データは、対象オブジェクト51の3次元形状モデルによりレンダリングされた複数の学習画像と、学習画像における前記対象オブジェクト51のキーポイントの位置を示す正解データとを含む。キーポイントは、対象オブジェクト

5 1 内にある仮想的な点であって、姿勢の算出に用いる点である。なお、推定モデル 2 6 は、上記の訓練データだけでなく、追加の訓練データによっても学習される。追加の訓練データは、入力画像に基づいて生成される画像を含み、また入力画像に基づいて訓練データを追加するか否かは推定モデル 2 6 の出力に基づいて判定される。

[0042] 図 4 は、対象オブジェクト 5 1 のキーポイントの一例を示す図である。対象オブジェクト 5 1 のキーポイントの 3 次元位置は、対象オブジェクト 5 1 の 3 次元形状モデル（より具体的には 3 次元形状モデルに含まれる頂点の情報）から、例えば公知の Farthest Point アルゴリズムにより決定される。図 4 には説明の容易のため、3 つのキーポイント K 1 ~ K 3 が記載されているが、実際のキーポイントの数はより多くてよい。例えば本実施形態では対象オブジェクト 5 1 の実際のキーポイントの数は 8 である。

[0043] 学習済の推定モデル 2 6 は、対象領域 5 5 が入力された際に、対象領域 5 5 における対象オブジェクト 5 1 のキーポイントの 2 次元位置を示す情報を出力する。対象領域 5 5 におけるキーポイントの 2 次元位置と入力画像における対象領域 5 5 の位置とから、入力画像におけるキーポイントの 2 次元位置が求められる。キーポイントの位置を示すデータは、各点とその点とキーポイントとの位置関係（例えば方向）を示す位置画像であってよい。

[0044] 図 5 は、対象領域 5 5 における位置画像の一例を模式的に示す図である。位置画像は、キーポイントの種類ごとに生成されてよい。位置画像は、各点におけるその点とキーポイントとの相対的な方向を示す。図 5 に示される位置画像では、各点の値に応じたパターンが記載され、各点の値は、その点の座標とキーポイントの座標との方向を示している。図 5 はあくまで模式的な図であり、各点の実際の値は連続的に変化する。図では明示されていないが、位置画像は、各点におけるその点を基準としたキーポイントの相対的な方向を示す Vector Field 画像である。

[0045] キーポイント決定部 2 7 は、推定モデル 2 6 の出力に基づいて、対象領域 5 5 および入力画像におけるキーポイントの 2 次元位置を決定する。より具

体的には、例えば、キーポイント決定部 27 は、推定モデル 26 から出力される位置画像に基づいて、対象領域 55 におけるキーポイントの 2 次元位置の候補を算出し、算出された 2 次元位置の候補から入力画像におけるキーポイントの 2 次元位置を決定する。キーポイント決定部 27 は、例えば、位置画像のうちの任意の 2 点の組み合わせのそれぞれからキーポイントの候補点を算出し、複数の候補点に対して位置画像の各点が示す方向と合致しているかを示すスコアを生成する。キーポイント決定部 27 はそのスコアが最も大きい候補点をキーポイントの位置と推定してよい。またキーポイント決定部 27 は、キーポイントごとに上記の処理を繰り返す。

[0046] 姿勢算出部 28 は、入力画像におけるキーポイントの 2 次元位置を示す情報と対象オブジェクト 51 の 3 次元形状モデルにおけるキーポイントの 3 次元位置を示す情報とに基づいて、対象オブジェクト 51 の姿勢を推定し、推定された姿勢を示す姿勢データを出力する。対象オブジェクト 51 の姿勢は、公知のアルゴリズムによって推定される。例えば、姿勢推定についての Perspective-n-Point (PNP) 問題の解法 (例えば EPnP) により推定されてよい。また、姿勢算出部 28 は対象オブジェクト 51 の姿勢だけでなく入力画像における対象オブジェクト 51 の位置も推定し、姿勢データにその位置を示す情報が含まれてもよい。

[0047] 推定モデル 26、キーポイント決定部 27、姿勢算出部 28 の詳細は、PVNet: Pixel-Wise Voting Network for 6DoF Pose Estimation の論文に記載されたものであってよい。

[0048] 撮影画像取得部 33、識別訓練データ生成部 34、識別学習部 35、形状モデル取得部 36、推定訓練データ生成部 37、推定学習部 38、信頼度取得部 39 は、識別モデル 24 および推定モデルの学習に関する構成である。本実施形態では、まず、対象オブジェクト 51 が撮影された画像に基づいて、識別モデル 24 および推定モデル 26 が、例えばそれぞれ数秒、数分といった短時間で学習され、学習済の識別モデル 24、推定モデル 26 に基づく対象領域取得部 21 および姿勢推定部 25 の動作の後に、識別モデル 24、

推定モデル 26 についての再度の学習が行われる。

- [0049] 撮影画像取得部 33 は姿勢推定部 25 に含まれる推定モデル 26 および／または対象領域取得部 21 に含まれる識別モデル 24 を学習させるために、撮影部 20 により対象オブジェクト 51 が撮影された撮影画像を取得する。撮影部 20 は、予めキャリブレーションによってカメラ内部パラメータが取得されているものとする。このパラメータは、P n P 問題を解く際に用いられる。
- [0050] 識別訓練データ生成部 34 は、対象オブジェクト 51 を含む画像に基づく正例訓練データと、対象オブジェクト 51 を含まない画像に基づく負例訓練データとを生成する。対象オブジェクト 51 を含む画像は、撮影画像取得部 33 により取得されてよい。
- [0051] 識別学習部 35 は、識別訓練データ生成部 34 により生成された訓練データに基づいて、対象領域取得部 21 に含まれる識別モデル 24 を学習させる。
- [0052] 形状モデル取得部 36 は、撮影画像取得部 33 により取得された対象オブジェクト 51 についての複数の撮影画像のそれぞれについて局所的な特徴を示す複数の特徴ベクトルを抽出し、複数の撮影画像から抽出された互いに対応する複数の特徴ベクトルと撮影画像においてその特徴ベクトルが抽出された位置とからその特徴ベクトルが抽出された点の 3 次元位置を求め、その 3 次元位置に基づいて対象オブジェクト 51 の 3 次元形状モデルを取得する。この方法は、いわゆる S f M や Visual SLAM を実現するソフトウェアでも用いられる公知の方法であるので、詳細の説明は省略する。
- [0053] 推定訓練データ生成部 37 は、推定モデル 26 を学習させるための訓練データを生成する。より具体的には、推定訓練データ生成部 37 は、初期の訓練データとして、対象オブジェクト 51 の 3 次元形状モデルから、レンダリングされた訓練画像と、キーポイントの位置を示す正解データとを含む訓練データを生成する。
- [0054] 推定学習部 38 は、推定訓練データ生成部 37 により生成された訓練デー

タにより、姿勢推定部 25 に含まれる推定モデル 26 を学習させる。

[0055] 信頼度取得部 39 は、入力データが入力された際の機械学習モデルの出力に基づいて、その入力データに対する機械学習モデルの出力の信頼度を取得する。機械学習モデルの出力に基づいて信頼度を取得するとは、例えば、機械学習モデルである識別モデル 24 の出力、より具体的にはその出力を受けた後段の処理の結果に基づいて信頼度を算出することであり、推定モデル 26 が出力する位置画像に基づいて信頼度を算出することである。

[0056] 次に、姿勢の推定に関する処理について説明する。図 6 は、主に対象領域取得部 21 および姿勢推定部 25 の処理の一例を示すフロー図である。図 6 に示される処理は、定期的に繰り返し実行されてよい。

[0057] はじめに、対象領域取得部 21 に含まれる領域抽出部 22 は、撮影部 20 により撮影された入力画像を取得する（S101）。領域抽出部 22 は、撮影部 20 から直接的に入力画像を受信することにより入力画像を取得してもよいし、撮影部 20 から受信され記憶部 12 に格納された入力画像を取得してもよい。

[0058] 領域抽出部 22 は、入力画像から、何らかの物体が写っている 1 または複数の候補領域 56 を抽出する（S102）。領域抽出部 22 は、予め学習された RPN（Regional Proposal Network）を含んでよい。RPN は、対象オブジェクト 51 が撮影された画像と関連しない訓練データによって学習されてよい。この処理によって、計算の無駄が低減され、環境に対する一定のロバストネスが確保される。

[0059] ここで、領域抽出部 22 は、さらに、抽出された候補領域 56 の画像に対して、例えば、背景の除去処理（マスク処理）やサイズ調整などの加工処理を実行してよい。また加工された候補領域 56 の画像が以降の処理に用いられてよい。この処理によって、背景や照明条件によるドメインギャップを縮小させ、少ない訓練データで識別モデル 24 を学習させることが可能になる。

[0060] 対象領域取得部 21 は、候補領域 56 のそれぞれが対象オブジェクト 51

の画像を含むか判定する（S103）。この処理は、特徴抽出部23が候補領域56の画像から特徴量を抽出する処理と、識別モデル24がその特徴量から候補領域56が対象オブジェクト51を含むか否かを示す情報を出力する処理とを含む。

[0061] 特徴抽出部23は、候補領域56の画像からその画像に応じた特徴量を出力する。特徴抽出部23は、学習済のCNN（Convolutional Neural Network）を含む。このCNNは、画像の入力に応じて、当該画像に対応する特徴量を示す特徴量データ（入力特徴量データ）を出力する。特徴抽出部23は、RPNにより抽出された候補領域56の画像から特徴量を抽出してもよいし、例えばFaster R-CNNのように、RPNの処理において抽出された特徴量を取得してもよい。

[0062] 識別モデル24は、SVM（Support Vector Machine）などであり、一種の機械学習モデルである。識別モデル24は、候補領域56の画像に対応する特徴量を示す入力特徴量データの入力に応じて、候補領域56に写るオブジェクトが識別モデル24における正クラスに属するものである確率を示す識別スコアを出力する。識別モデル24は、正例についての複数の正例訓練データと負例についての複数の負例訓練データとにより学習されている。正例訓練データは対象オブジェクト51が撮影された画像を含む学習画像から生成され、負例訓練データは対象オブジェクト51と異なるオブジェクトの画像であって、予め準備された画像から生成される。負例訓練データは、撮影部20により撮影された、その撮影部20の環境を撮影することにより生成されてもよい。

[0063] 本実施形態では、このCNNを用いて、正規化処理が実行された画像に対応する特徴量を示す特徴量データの生成が行われる。なお、特徴抽出部23は、画像の特徴を示す特徴量を算出する他の公知のアルゴリズムにより、画像の入力に応じて、当該画像に対応する特徴量を示す特徴量データを出力してもよい。

[0064] 対象領域取得部21は、例えば識別スコアが閾値より大きい場合に、その

候補領域 5 6 が対象オブジェクト 5 1 の画像を含むと判定する。

[0065] 候補領域 5 6 のそれぞれが対象オブジェクト 5 1 の画像を含むか判定されると、対象領域取得部 2 1 はその判定結果に基づいて対象領域 5 5 を決定する (S 1 0 4)。より具体的には、対象領域取得部 2 1 は、対象オブジェクト 5 1 を含むと判定された候補領域 5 6 に基づいて、対象オブジェクト 5 1 の近傍の領域を含む矩形の領域を対象領域 5 5 として取得する。対象領域取得部 2 1 は、対象オブジェクト 5 1 の近傍の領域を含む正方形の領域を対象領域 5 5 として取得してよいし、単に候補領域 5 6 を対象領域 5 5 として取得してもよい。なお、対象領域取得部 2 1 は、常に S 1 0 2, S 1 0 3 の処理により対象領域 5 5 を取得しなくてもよい。例えば、対象領域取得部 2 1 は、対象領域 5 5 を取得した後に取得された入力画像に対して、公知の時系列の追尾処理を行うことにより、対象領域 5 5 を取得してもよい。

[0066] 姿勢推定部 2 5 は、学習済の推定モデル 2 6 に、対象領域 5 5 の画像を入力する (S 1 0 5)。ここで入力される対象領域 5 5 の画像は、推定モデル 2 6 の入力画像のサイズにあわせてサイズが調整 (拡大または縮小) された画像であってよい。サイズを調整 (正規化) することにより、推定モデル 2 6 の学習の効率が向上する。なお、姿勢推定部 2 5 は、対象領域 5 5 の画像の背景をマスクし、その背景がマスクされた対象領域 5 5 の画像を推定モデル 2 6 に入力してよい。

[0067] 姿勢推定部 2 5 に含まれるキーポイント決定部 2 7 は、推定モデル 2 6 の出力に基づいて、対象領域 5 5 および入力画像におけるキーポイントの 2 次元位置を決定する (S 1 0 6)。推定モデル 2 6 の出力が位置画像である場合には、キーポイント決定部 2 7 は位置画像の各点からキーポイントの位置の候補を算出し、その候補に基づいてキーポイントの位置を決定する。推定モデル 2 6 の出力が対象領域 5 5 におけるキーポイントの位置である場合には、その位置から入力画像におけるキーポイントの位置を算出してよい。なお、S 1 0 5、S 1 0 6 の処理はキーポイントの種類ごとに行われる。

[0068] 姿勢推定部 2 5 に含まれる姿勢算出部 2 8 は、決定されたキーポイントの

2次元位置に基づいて、対象オブジェクト51の推定される姿勢を算出する（S107）。姿勢算出部28は姿勢とともに対象オブジェクト51の位置を算出してよい。姿勢および位置は、前述のPNP問題の解法により算出されてよい。

[0069] 図7は、検出された対象オブジェクト51の姿勢を説明する図である。図7では、説明の容易のため、対象オブジェクト51のローカル座標系を示すローカル座標軸59により対象オブジェクト51の姿勢を表している。ローカル座標軸59の原点の位置が対象オブジェクト51の位置を示し、ローカル座標軸59の線の向きが姿勢を示している。

[0070] ここで、信頼度取得部39は、対象領域55に対する推定モデル26の出力について信頼度を算出する（S108）。そして、その信頼度があらかじめ定められた条件を満たす場合に、識別訓練データ生成部34および推定訓練データ生成部37は、その対象領域に基づいて、それぞれ識別モデル24および推定モデル26に対する追加の訓練データを生成する（S109）。S109の処理は、機械学習モデルの学習後（推論時）に、その機械学習モデルに入力されるデータに基づいて追加の訓練データを生成するものである。S108およびS109の処理の詳細については後述する。

[0071] 推定された対象オブジェクト51の姿勢および位置は、様々に利用されてよい。例えば、コントローラによって入力される操作情報の代わりにゲームなどのアプリケーションソフトウェアに入力されてよい。そしてアプリケーションソフトウェアの実行コードを実行するプロセッサ11は、その姿勢（および位置）に基づいて、画像のデータを生成し、表示部18にその画像を出力させてよい。またプロセッサ11は、情報処理装置10または情報処理装置10に接続される音声出力装置に、その姿勢（および位置）に基づく音を出力させてよい。またプロセッサ11は、例えばロボットのようなAIエージェントにオブジェクトの位置姿勢を通知することにより、AIエージェントの動作を制御し、例えば物体の把持などを行わせてもよい。

[0072] 次に、識別モデル24および推定モデル26の学習の概要について説明す

る。図8は、識別モデル24および推定モデル26の学習を概略的に説明するフロー図である。

[0073] はじめに、識別訓練データ生成部34は識別モデル24の初期の訓練データを取得し、推定訓練データ生成部37は、推定モデル26の初期の訓練データを取得する（S201）。

[0074] ステップS201の処理についてさらに詳細に説明する。図9は、初期の訓練データを生成する処理の一例を示すフロー図である。

[0075] 撮影画像取得部33は、対象オブジェクト51が撮影された複数の撮影画像を取得する（S301）。

[0076] 図10は、対象オブジェクト51の撮影を説明する図である。対象オブジェクト51は、例えば手53によって保持されており、撮影部20により撮影される。本実施形態では、対象オブジェクト51を様々な方向から撮影することが望ましい。そのため、撮影部20は動画撮影のように定期的に画像を撮影しつつ、対象オブジェクト51の撮影方向を変化させる。例えば手53によって対象オブジェクト51の姿勢を変化させることで対象オブジェクト51の撮影方向を変化させてよい。またARマーカー上に対象オブジェクト51を配置し、撮影部20を動かすことにより撮影方向を変化させてもよい。後述の処理で用いられる撮影画像の取得間隔は、動画の撮影間隔より広くてもよい。

[0077] 撮影画像が取得されると、撮影画像取得部33は、それらの撮影画像から手53の画像をマスクする（S302）。手53の画像のマスクは公知の方法により行われてよい。例えば、撮影画像取得部33は、撮影画像に含まれる肌の色の領域を検出することにより手53の画像をマスクしてよい。

[0078] そして形状モデル取得部36は、複数の撮影画像から、対象オブジェクト51の3次元形状モデルと、撮影画像のそれぞれにおける姿勢とを算出する（S303）。この処理は、いわゆるSfMやVisual SLAMを実現するソフトウェアでも用いられる前述の公知の方法より行われてよい。形状モデル取得部36は、この方法によるカメラの撮影方向の算出ロジックに基づいて対象

オブジェクト51の姿勢を算出してよい。

[0079] 対象オブジェクト51の3次元形状モデルが算出されると、形状モデル取得部36は、その3次元形状モデルの姿勢の推定に用いる複数のキーポイントの3次元位置を決定する(S304)。形状モデル取得部36は、例えば、公知のFarthest Pointアルゴリズムにより複数のキーポイントの3次元位置を決定してよい。

[0080] キーポイントの3次元位置が算出されると、推定訓練データ生成部37は、推定モデル26向けに、複数の訓練画像と、複数の位置画像とを含む訓練データを生成する(S305)。より具体的には、推定訓練データ生成部37は、3次元形状モデルからレンダリングされた複数の訓練画像を生成し、その複数の訓練画像におけるキーポイントの位置を示す位置画像を生成する。複数の訓練画像は、互いに異なる複数の方向からみた対象オブジェクト51のレンダリング画像であり、位置画像は訓練画像とキーポイントとの組み合わせごとに生成される。

[0081] 推定訓練データ生成部37はレンダリングされた訓練画像にキーポイントの位置を仮想的に投影し、その投影されたキーポイントの位置と画像内の各点との相対位置に基づいて位置画像を生成する。推定モデル26の学習に用いる訓練データは訓練画像と位置画像とを含む。

[0082] 初期の訓練データに含まれる訓練画像は、レンダリングされた画像である。これは、短時間で多様な撮影方向から撮影された撮影画像を取得することが難しい一方、3次元形状モデルを用いれば容易に多様な撮影方向からみた画像を生成できるからである。なお、初期の訓練データに実写の訓練画像が含まれてもよい。

[0083] 識別訓練データ生成部34は、撮影画像取得部33により取得された複数の撮影画像、より具体的には対象オブジェクト51を含む画像から、正例訓練データを生成し、例えば記憶部12に格納された対象オブジェクトを含まない画像から負例訓練データを取得する(S306)。正例訓練データおよび負例訓練データが、識別モデル24の訓練データである。

- [0084] 識別訓練データ生成部34は、識別モデル24に入力される画像に応じた加工、例えば対象オブジェクト51を含む領域の切り出し、サイズの正規化、背景のマスク、特徴量の抽出をすることにより、撮影画像から正例訓練データを生成してよい。識別訓練データ生成部34は、予め記憶部12に格納される負例サンプル画像を特徴抽出部23に入力し、出力される特徴量データを取得することにより複数の負例訓練データを生成する。特徴量は、識別モデル24に含まれる特徴抽出部23と同じ処理により抽出される。負例サンプル画像は、例えば、予め撮影部20によって撮影された画像、Webから収集された画像、他の物体についての正例の画像であってよい。負例訓練データは予め生成され記憶部12に格納されていてもよい。
- [0085] なお識別モデル24はこれまでに説明したものには限られず、画像から直接的に対象オブジェクト51が存在するか判定するものであってもよい。
- [0086] 識別モデル24および推定モデル26の初期の訓練データが取得されると、識別学習部35は識別モデル24を識別モデル向けの初期の訓練データにより学習させ、推定学習部38は推定モデル26を識別モデル向けの初期の訓練データにより学習させる（S202）。識別モデル24は、例えばSVMであり、識別学習部35はそのSVMを正例訓練データおよび負例訓練データにより学習させてよい。
- [0087] 識別モデル24および推定モデル26が学習されると、S203からS207において、情報処理システムは、それらのモデルを用いていわゆる推論の処理を実行しつつ、信頼度に応じて識別モデル24および推定モデル26のそれぞれについての追加の訓練データ（追加訓練データ）を取得する。
- [0088] S203においては、情報処理システムは、撮影された画像を入力画像として対象領域取得部21に入力し、対象領域取得部21および姿勢推定部25は対象領域55の抽出および対象領域55に含まれる対象オブジェクト51の姿勢の推定の処理を実行する。S203の処理は、図6のS101からS107までの処理に相当する。
- [0089] 次に、信頼度取得部39は、姿勢推定部25に含まれる推定モデル26の

出力に基づいて、その出力の信頼度を算出する（S 2 0 4）。この処理は図 6 の S 1 0 8 の処理に相当する。

[0090] より具体的には、信頼度取得部 3 9 は例えば以下の手順で信頼度を算出する。信頼度取得部 3 9 は推定モデル 2 6 が出力する位置画像から、それぞれ 2 つの点を含む複数のグループを選択する。信頼度取得部 3 9 は、そのグループのそれぞれについて、グループに含まれる各点が示すキーポイントの方向に基づいて、キーポイントの候補位置を算出する。候補位置は、ある点からその点が示す方向に伸ばした直線と、もう一つの点からその点が示す方向に伸ばした直線との交点に相当する。グループのそれぞれについて信頼度が算出されると、信頼度取得部 3 9 は、候補位置のばらつきを示す値を信頼度として算出する。信頼度取得部 3 9 は、例えば候補位置の重心からの距離の平均値を信頼度としてとってもよいし、候補位置の任意の方向の標準偏差を信頼度として算出してもよい。

[0091] 信頼度取得部 3 9 は、候補位置のばらつきを示す値以外から信頼度を算出してもよい。例えば、信頼度取得部 3 9 は、例えば対象領域の画像のような入力画像が推定モデル 2 6 に入力された際のその推定モデル 2 6 の出力と、その入力画像が所定の加工処理により加工された加工画像が推定モデル 2 6 に入力された際の出力との相違を示す情報に基づいて、信頼度を算出してもよい。

[0092] より具体的には、はじめに、信頼度取得部 3 9 は、対象領域の画像に所定の加工（Augmentation）を実行する。この加工は例えば明度の変更やノイズの付加のうちいずれかであってよい。次に信頼度取得部 3 9 は、加工された画像を推定モデル 2 6 に入力し、その出力である位置画像を取得する。そして信頼度取得部 3 9 は、当初の対象領域の画像に対して出力された位置画像（当初の出力）と、加工された画像により出力された位置画像との違いを示す値を信頼度として算出する。この値は、当初の出力と加工された画像に対する出力との各点における値の違いの統計量であってもよいし、当初の出力により算出されたキーポイントの位置と加工された画像に対する出力により

算出されたキーポイントの位置との距離であってもよい。また当初の対象領域の画像に対する出力の代わりに、対象領域の画像に所定の加工と異なる加工がされた画像を推定モデル 26 に入力した際の出力が用いられてもよい。なお、ここで行われる加工 (Augmentation) は、後述の推定訓練データ生成部 37 による加工と手法が異なってよい。手法が異なることにより、追加訓練データにより学習された推定モデルに 26 を用いて信頼度を算出する際に生じる信頼度の精度が抑制される。

[0093] 信頼度取得部 39 は、候補位置のばらつきを示す値から算出される信頼度 (の要素) と、当初の出力と加工された画像に対する出力との違いを示す値とを組み合わせる最終的な信頼度を出力してよい。信頼度取得部 39 は例えば前者と後者とを重みづけ加算した値を信頼度として出力してよい。

[0094] 信頼度が算出されると、信頼度取得部 39 は、算出された信頼度が訓練データを追加するための追加条件を満たすか判定する (S205)。追加条件は、例えば、信頼度として算出されたばらつきの値が閾値より小さいことであってよい。

[0095] 信頼度が追加条件を満たす場合には (S205のY)、識別訓練データ生成部 34 および推定訓練データ生成部 37 は、それぞれ識別モデル 24 および推定モデル 26 の訓練データに追加される追加訓練データを生成する (S206)。S205 および S206 は図 6 の S109 の処理に相当する。

[0096] より具体的には、識別訓練データ生成部 34 はその位置画像の元となった対象領域 55 に相当する画像 (例えば対応する候補領域 56 の画像) を正例画像として決定し、その正例画像のデータを識別モデルの訓練データに追加する。識別訓練データ生成部 34 は正例画像として決定された画像に、識別モデル 24 に入力される画像に応じた加工、例えば特徴量の抽出をすることにより、撮影画像から正例訓練データを生成してよい。

[0097] また推定訓練データ生成部 37 は、その位置画像の元となった対象領域の画像に基づいて第 1 の追加画像、第 2 の追加画像のセットを生成し、そのセットを推定モデルの追加訓練データに追加する。

- [0098] より具体的には、推定訓練データ生成部 37 は、対象領域の画像に第 1 の加工 (Augmentation) を実行し、加工された画像を第 1 の追加画像として取得する。また推定訓練データ生成部 37 は、対象領域の画像に第 2 の加工 (Augmentation) を実行し、加工された画像を第 2 の追加画像として取得する。第 1 の加工および第 2 の加工は互いに異なり、それぞれ、例えば明度の変更やノイズの付加のうちいずれかが行われてよい。また、第 1 の加工および第 2 の加工のうち一方は、実質的な加工を行わないものであってもよい。第 1 の追加画像および第 2 の追加画像のセットを用いた推定モデルの学習の手法 (Consistency loss) については後述する。
- [0099] なお、推定訓練データ生成部 37 は、その位置画像の元となった対象領域の画像と、その画像について姿勢推定部 25 により算出された姿勢を示す正解データとのセットを追加訓練データに追加してもよい。こちらについては初期の学習と同じ手法により推定モデル 26 が学習されてよい。
- [0100] 本実施形態では、学習済の機械学習モデルを用いて推論をする際の入力データの一部を訓練データに追加している。一方、通常は推論の際の入力データを訓練データに追加することはない。例えば入力データに対する出力が誤っている場合に、その入力データの追加により訓練データの質の低下を招く恐れがあるからである。本実施形態では、機械学習モデルの出力について信頼度を算出し、その信頼度を用いて訓練データに追加するか否かをフィルタリングすることにより、その追加されるデータの質を確保し、訓練データの生成の手間を減らしつつ機械学習モデルの精度を向上させることを可能にしている。
- [0101] ここで、本実施形態で算出される信頼度は、識別モデル 24 および推定モデル 26 の信頼度と考えることができる。識別モデル 24 の出力の信頼度という観点では、識別モデル 24 の後段にある推定モデル 26 が出力する位置画像がキーポイントを正確に求められる状態であることを示す信頼度を求めているといえる。このように後段の機械学習モデルを含む処理が適切に行えることを信頼度の指標とすることで、簡易かつ効果的に信頼度を算出すること

ができる。また推定モデル 26 の出力の信頼度という観点では、出力となる位置画像からキーポイントを求める後段の処理が適切に行えることを信頼度の指標としているといえる。

[0102] 追加訓練データが取得されると、再学習を開始する条件を満たさない限り（S 207 の N）、S 203 以降の処理が繰り返し実行される。再学習を開始する条件は、取得された追加訓練データの数が増加することによって、閾値に達することであってもよい、いわゆる繰り返しの推定の処理の終了の操作が入力されることであってもよい。

[0103] 再学習を開始する条件が満たされると（S 207 の Y）、識別学習部 35 および推定学習部 38 は、それぞれ識別モデル 24 および推定モデル 26 を再学習させる（S 208）。

[0104] ここで、再学習は、追加訓練データを含む訓練データを用いて機械学習モデルを学習させることを指している。学習の対象となる機械学習モデル（識別モデル 24 および推定モデル 26）は、推論を実行している機械学習モデルである識別モデル 24 および推定モデル 26 と異なるインスタンスであってもよい、推論を実行している機械学習モデルと同一のインスタンスであってもよい。前者の場合には、学習が終了した後に推論に用いる識別モデル 24 および推定モデル 26 のインスタンスが切り替えられてよい。またインスタンスの切替の代わりに、推論に用いる識別モデル 24 および推定モデル 26 のインスタンスに対して新たに学習された機械学習モデルのパラメータがコピーされてもよい。

[0105] 識別モデル 24 については、識別学習部 35 は初期の訓練データに追加訓練データを追加し、その追加後の訓練データにより識別モデル 24 を学習してよい。また識別モデル 24 の学習に用いられる訓練データは、初期の訓練データおよび追加訓練データのすべてであってもよい、それらの一部であってもよい。識別モデル 24 の学習に用いられる一部の訓練データは、例えばその数がサンプル総数の最大値以下になるように選択されたものであってもよい、何らかの手法で品質が低いと判定されたサンプルが除外された

ものであってもよい。

[0106] 一方、第1の追加画像および第2の追加画像についての追加訓練データを有する推定モデル26の再学習の手法は異なる。図11は、推定モデルの再学習の処理の一例を示すフロー図であり、再学習においては、図11に示される処理が複数回繰り返し実行される。

[0107] まず、推定学習部38は、推定モデル26向けの初期の訓練データにより、推定モデル26を学習させる（S501）。この学習は、ステップS202における推定モデル26の学習と同様の手法、より具体的には、推定学習部38は、推定モデル26が出力する位置画像と正解データとの相違（L1ロス）を教師信号として、推定モデル26のパラメータを調整する。

[0108] 次に、推定学習部38は、推定モデル26向けの追加訓練データに含まれる未取得のセットのうち1つを取得する（S502）。推定学習部38はそのセットに含まれる第1の追加画像を推定モデル26に入力し、推定モデル26の出力（第1の出力）を取得する（S503）。また推定学習部38はそのセットに含まれる第2の追加画像を推定モデル26に入力し、推定モデル26の出力（第2の出力）を取得する（S504）。

[0109] 推定学習部38は、第1の出力と第2の出力との相違（Consistency loss）を示す情報を算出し（S505）、その相違を示す情報に基づいて推定モデル26のパラメータを調整する（S506）。相違を示す情報は、第1の出力および第2の出力の各点における値の違いの統計量（例えば平均）であってよい。

[0110] ここで、追加訓練データについては、第1の出力と第2の出力との相違に応じた学習が行われるため、主にこの手法で学習すると、例えば入力に関わらず同じ位置画像を出力するように推定モデル26のパラメータが収束する恐れがある。その事態を避けるため、初期の訓練データを含むすべての訓練データの数に対する追加訓練データの数の割合を所定の値（例えば20%）以内に抑えることが望ましい。

[0111] 追加訓練データを用いて、同一の画像をベースとする2つの画像が一致す

るか否かで再学習させることにより、正解のラベルのない訓練データも用いて学習させることが可能になり、精度を向上させることができる。

[0112] 本実施形態では、対象領域取得部 21 の処理により、推定モデル 26 に入力する画像を、撮影された画像のうち対象オブジェクト 51 が存在する領域の画像であって、対象オブジェクト 51 が中央に存在する蓋然性が十分に高い画像に限定している。また姿勢推定部 25 の推定モデル 26 は 3 次元形状モデルにより生成された訓練データにより学習され、一方で、対象領域取得部 21 の識別モデル 24 は、対象オブジェクト 51 が撮影された画像に基づいて学習されている。

[0113] 推定モデル 26 に入力される画像を適切に限定することにより、推定モデル 26 の出力の精度が向上し、推定される対象オブジェクト 51 の姿勢の精度が向上する。さらに、識別モデル 24 を、3 次元形状モデルに基づく画像ではなく撮影画像に基づいて学習させることにより、対象領域 55 をより正確に選択することが可能になり、ひいては推定モデル 26 の精度を向上させることができる。

[0114] 本実施形態では、姿勢推定部 25 の推定モデル 26 を学習するための 3 次元形状モデルを生成するための撮影画像を、識別モデル 24 を学習する際にも用いている。これにより、対象オブジェクト 51 の撮影にかかる手間を低減し、推定モデル 26 および識別モデル 24 の学習にかかる時間を低減する。

[0115] 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。

[0116] 例えば、識別モデル 24 は、任意のカーネルの SVM であってもよい。また、識別モデル 24 は、K 近傍法、ロジスティック回帰、アダブースト等のブースティング方法などの方法を用いた識別器であってもよい。また、識別モデル 24 が、ニューラルネットワーク、ナイーブベイズ分類器、ランダムフォレスト、決定木などによって実装されてもよい。

[0117] 推定モデル 26 の出力は、キーポイントの位置を示すヒートマップのような位置画像であってもよい。この場合、例えば、信頼度取得部 39 は、推定

モデル 2 6 が出力する位置画像が有するピークの数信頼度として求めてよい。このピークの数閾値より小さい場合に、入力データが訓練データに追加されてよい。

[0118] 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ず、必要に応じて改変されてよい。

請求の範囲

- [請求項1] 訓練データにより学習された機械学習モデルと、
入力データが入力された際の前記機械学習モデルの出力に基づいて、前記入力データに対する当該出力の信頼度を出力する信頼度出力手段と、
前記信頼度が所定の条件を満たす場合に、前記入力データに基づいて新たな訓練データを生成する生成手段と、
前記新たな訓練データにより機械学習モデルを学習させる学習制御手段と、
を含む情報処理システム。
- [請求項2] 請求項1に記載の情報処理システムにおいて、
前記機械学習モデルの出力に基づいて推定結果を出力する学習済の推定モデルを含み、
前記信頼度出力手段は、前記推定モデルの出力に基づいて前記入力データに対する前記機械学習モデルの出力の信頼度を出力する、
情報処理システム。
- [請求項3] 請求項2に記載の情報処理システムにおいて、
前記入力データは、対象オブジェクトが撮影された画像を含み、
前記推定モデルは、前記機械学習モデルの出力に基づいて、前記対象オブジェクトの姿勢推定のためのキーポイントを示す画像を出力し、
前記信頼度出力手段は、前記画像に基づいて前記信頼度を出力する、
情報処理システム。
- [請求項4] 請求項3に記載の情報処理システムにおいて、
前記推定モデルは、各点がキーポイントとの位置関係を示す画像を出力し、
前記信頼度出力手段は、複数のキーポイントの位置の候補であって

それぞれ前記推定モデルが出力した画像に含まれる互いに異なる点から生成されるキーポイントの位置の候補のばらつきに基づいて、前記信頼度を出力する、

情報処理システム。

[請求項5]

請求項3または4に記載の情報処理システムにおいて、

前記信頼度出力手段は、入力画像が前記推定モデルに入力された際の当該推定モデルの出力と、前記入力画像が所定の加工処理により加工された加工画像が前記推定モデルに入力された際の出力との相違を示す情報に基づいて、前記信頼度を出力する、

情報処理システム。

[請求項6]

請求項3から5のいずれか1項に記載の情報処理システムにおいて

、

前記機械学習モデルは前記入力データが前記対象オブジェクトを含むか否かを示す情報を出力する、

情報処理システム。

[請求項7]

請求項2から6のいずれかに記載の情報処理システムにおいて、

前記推定モデルは推定訓練データにより学習され、

前記生成手段は、前記信頼度が前記所定の条件を満たす場合に、前記入力データに基づいて新たな推定訓練データを生成し、

前記学習制御手段は、前記新たな訓練データにより機械学習モデルを学習させる、

情報処理システム。

[請求項8]

請求項1に記載の情報処理システムにおいて、

前記入力データは、対象オブジェクトが撮影された画像を含み、

前記機械学習モデルは、前記入力データに基づいて、対象オブジェクトの姿勢推定のためのキーポイントを示す画像を出力し、

前記信頼度出力手段は、前記画像に基づいて前記信頼度を出力する

、

情報処理システム。

[請求項9]

請求項8に記載の情報処理システムにおいて、

前記訓練データは3次元形状モデルからレンダリングされた複数の学習画像とそれぞれが前記学習画像に対する正解データである正解画像とを含む、

情報処理システム。

[請求項10]

請求項9に記載の情報処理システムにおいて、

前記生成手段は、前記入力データが第1の加工処理により加工された第1の追加画像と、前記入力データが前記第1の加工処理と異なる第2の加工処理により加工された第2の追加画像とを含む新たな訓練データを生成し、

前記学習制御手段は、前記機械学習モデルに前記第1の追加画像を入力した際の出力と、前記機械学習モデルに前記第2の追加画像を追加した際の出力との違いに基づいて、機械学習モデルを学習させる、
情報処理システム。

[請求項11]

訓練データにより学習された機械学習モデルに入力データが入力された際の当該機械学習モデルの出力に基づいて、前記入力データに対する当該出力の信頼度を出力するステップと、

前記信頼度が所定の条件を満たす場合に、前記入力データに基づいて新たな訓練データを生成するステップと、

前記新たな訓練データにより機械学習モデルを学習させるステップと、

を含む情報処理方法。

[請求項12]

訓練データにより学習された機械学習モデルに入力データが入力された際の当該機械学習モデルの出力に基づいて、前記入力データに対する当該出力の信頼度を出力し、

前記信頼度が所定の条件を満たす場合に、前記入力データに基づいて新たな訓練データを生成し、

前記新たな訓練データにより機械学習モデルを学習させる、
処理をコンピュータに実行させるためのプログラム。

補正された請求の範囲
[2022年9月22日(22.09.2022)国際事務局受理]

- [請求項1] 訓練データにより学習された機械学習モデルと、
入力データが入力された際の前記機械学習モデルの出力に基づいて、前記入力データに対する当該出力の信頼度を出力する信頼度出力手段と、
前記信頼度が所定の条件を満たす場合に、前記入力データに基づいて新たな訓練データを生成する生成手段と、
前記新たな訓練データにより機械学習モデルを学習させる学習制御手段と、
を含む情報処理システム。
- [請求項2] 請求項1に記載の情報処理システムにおいて、
前記機械学習モデルの出力に基づいて推定結果を出力する学習済の推定モデルを含み、
前記信頼度出力手段は、前記推定モデルの出力に基づいて前記入力データに対する前記機械学習モデルの出力の信頼度を出力する、
情報処理システム。
- [請求項3] 請求項2に記載の情報処理システムにおいて、
前記入力データは、対象オブジェクトが撮影された画像を含み、
前記推定モデルは、前記機械学習モデルの出力に基づいて、前記対象オブジェクトの姿勢推定のためのキーポイントを示す画像を出力し、
前記信頼度出力手段は、前記画像に基づいて前記信頼度を出力する、
情報処理システム。
- [請求項4] 請求項3に記載の情報処理システムにおいて、
前記推定モデルは、各点がキーポイントとの位置関係を示す画像を出力し、
前記信頼度出力手段は、複数のキーポイントの位置の候補であって

それぞれ前記推定モデルが出力した画像に含まれる互いに異なる点から生成されるキーポイントの位置の候補のばらつきに基づいて、前記信頼度を出力する、

情報処理システム。

[請求項5] (補正後) 請求項2から4のいずれか1項に記載の情報処理システムにおいて、

前記信頼度出力手段は、入力画像が前記推定モデルに入力された際の当該推定モデルの出力と、前記入力画像が所定の加工処理により加工された加工画像が前記推定モデルに入力された際の出力との相違を示す情報に基づいて、前記信頼度を出力する、

情報処理システム。

[請求項6] (補正後) 請求項3または4に記載の情報処理システムにおいて、

前記機械学習モデルは前記入力データが前記対象オブジェクトを含むか否かを示す情報を出力する、

情報処理システム。

[請求項7] 請求項2から6のいずれかに記載の情報処理システムにおいて、

前記推定モデルは推定訓練データにより学習され、

前記生成手段は、前記信頼度が前記所定の条件を満たす場合に、前記入力データに基づいて新たな推定訓練データを生成し、

前記学習制御手段は、前記新たな訓練データにより機械学習モデルを学習させる、

情報処理システム。

[請求項8] 請求項1に記載の情報処理システムにおいて、

前記入力データは、対象オブジェクトが撮影された画像を含み、

前記機械学習モデルは、前記入力データに基づいて、対象オブジェクトの姿勢推定のためのキーポイントを示す画像を出力し、

前記信頼度出力手段は、前記画像に基づいて前記信頼度を出力する

、

情報処理システム。

[請求項9]

請求項8に記載の情報処理システムにおいて、

前記訓練データは3次元形状モデルからレンダリングされた複数の学習画像とそれぞれが前記学習画像に対する正解データである正解画像とを含む、

情報処理システム。

[請求項10]

(補正後) 請求項2または9に記載の情報処理システムにおいて、

前記生成手段は、前記入力データが第1の加工処理により加工された第1の追加画像と、前記入力データが前記第1の加工処理と異なる第2の加工処理により加工された第2の追加画像とを含む新たな訓練データを生成し、

前記学習制御手段は、前記機械学習モデルに前記第1の追加画像を入力した際の出力と、前記機械学習モデルに前記第2の追加画像を追加した際の出力との違いに基づいて、機械学習モデルを学習させる、
情報処理システム。

[請求項11]

訓練データにより学習された機械学習モデルに入力データが入力された際の当該機械学習モデルの出力に基づいて、前記入力データに対する当該出力の信頼度を出力するステップと、

前記信頼度が所定の条件を満たす場合に、前記入力データに基づいて新たな訓練データを生成するステップと、

前記新たな訓練データにより機械学習モデルを学習させるステップと、

を含む情報処理方法。

[請求項12]

訓練データにより学習された機械学習モデルに入力データが入力された際の当該機械学習モデルの出力に基づいて、前記入力データに対する当該出力の信頼度を出力し、

前記信頼度が所定の条件を満たす場合に、前記入力データに基づいて新たな訓練データを生成し、

前記新たな訓練データにより機械学習モデルを学習させる、
処理をコンピュータに実行させるためのプログラム。

条約第19条(1)に基づく説明書

1. 補正の内容

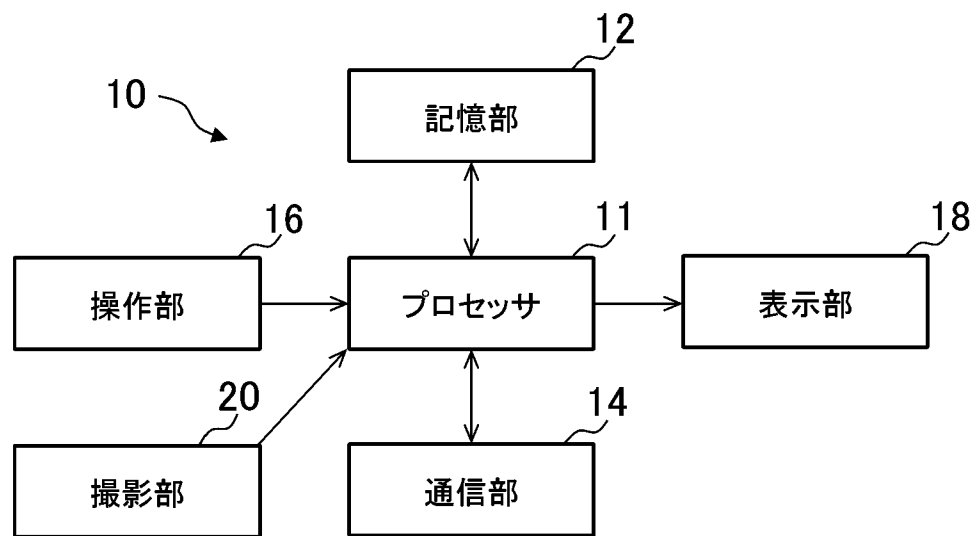
請求項5, 6, 10を補正した。

2. 説明

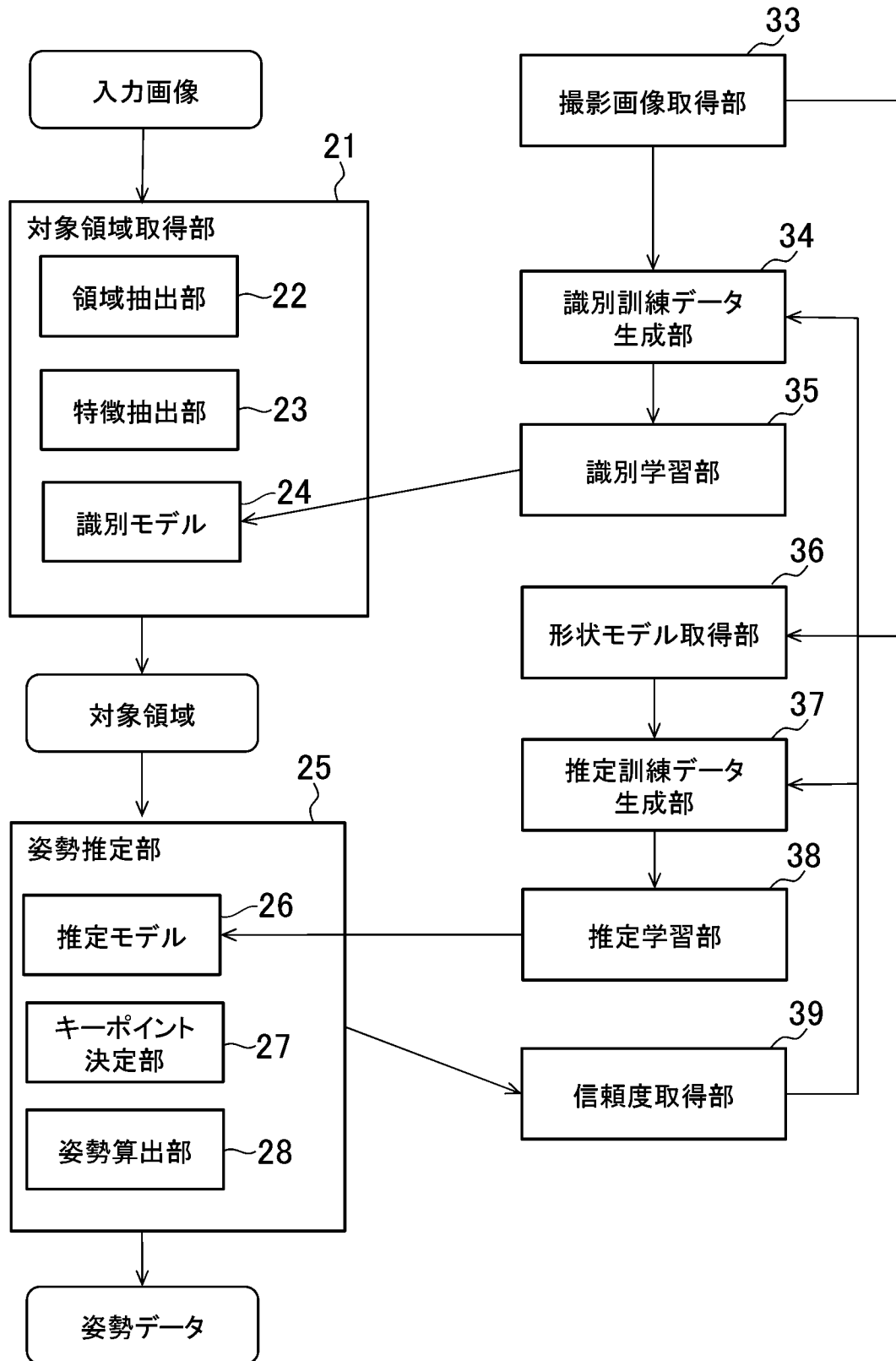
- (1) 請求項5を請求項3, 4に加えて請求項2にも従属させる。
- (2) 請求項6の従属先から請求項5を削除する。
- (3) 請求項10を請求項9に加えて請求項2に従属させる。

以上

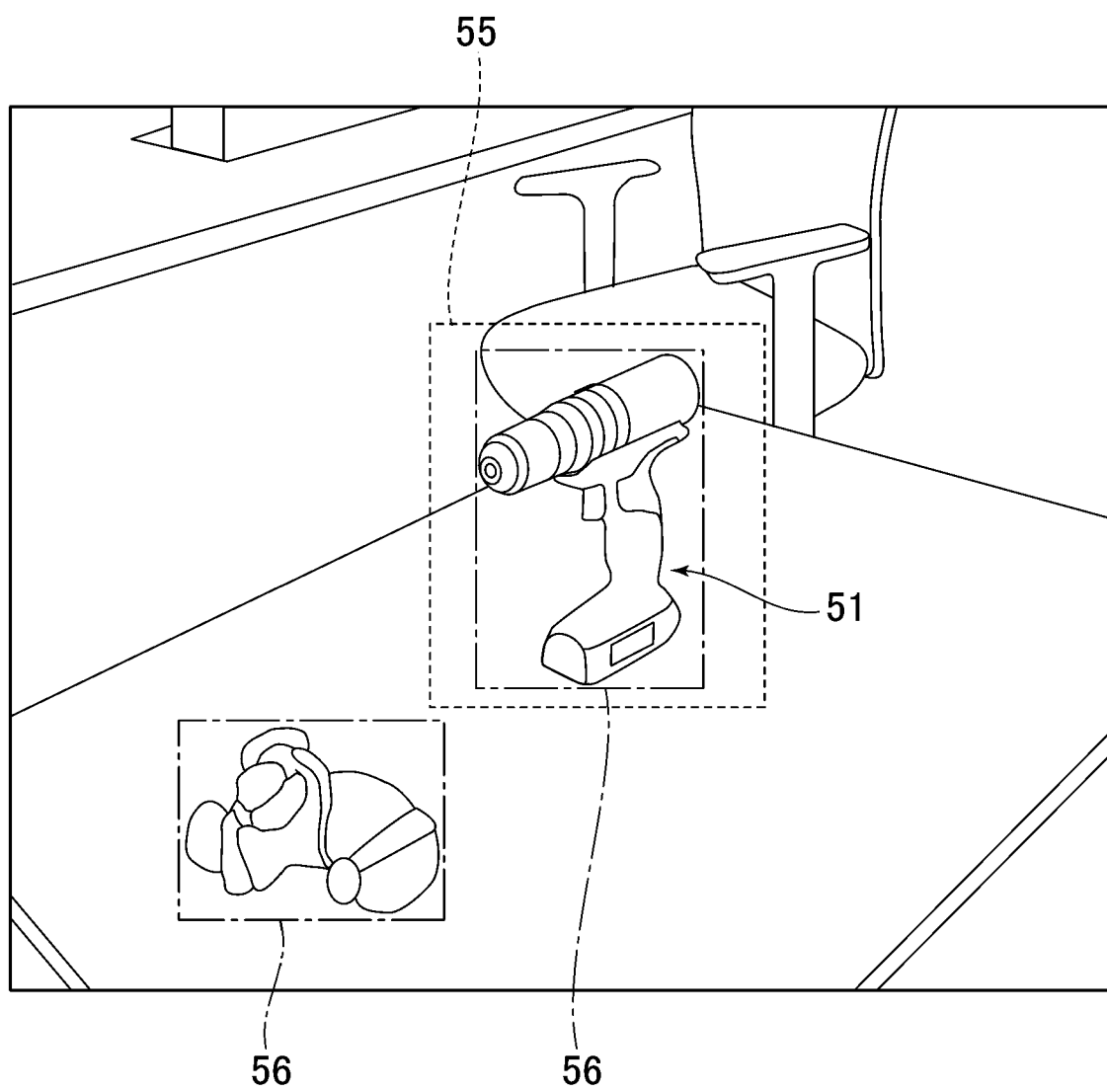
[図1]



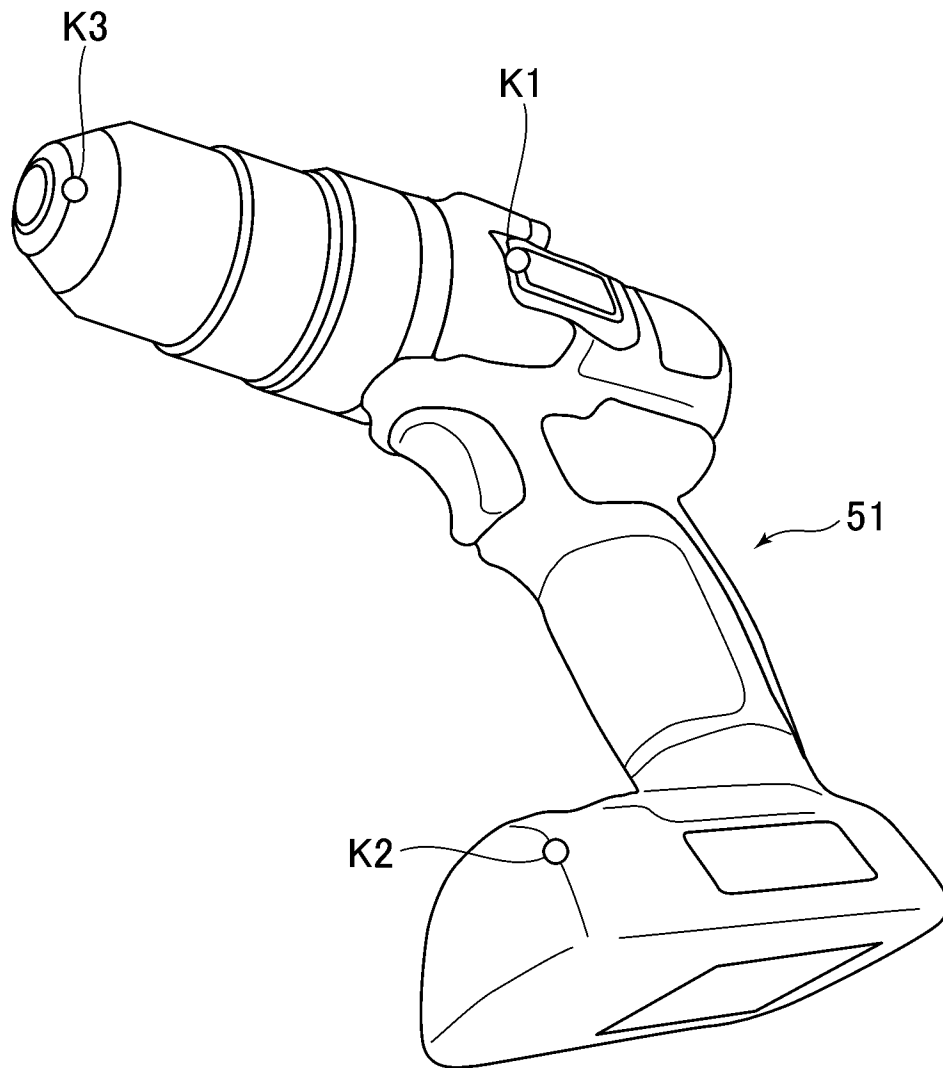
[図2]



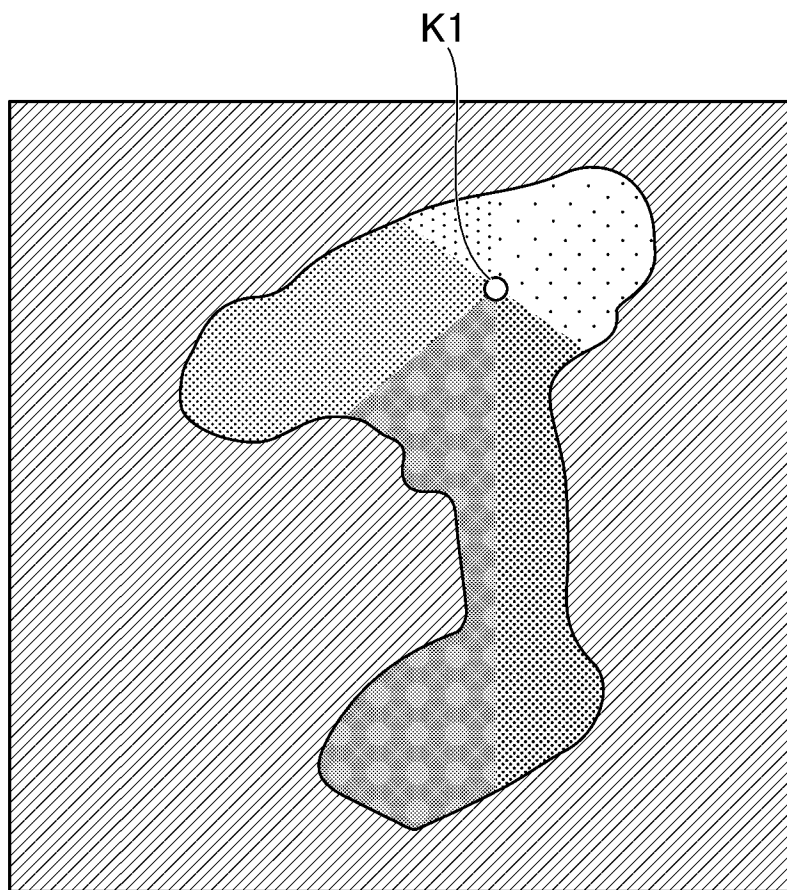
[図3]



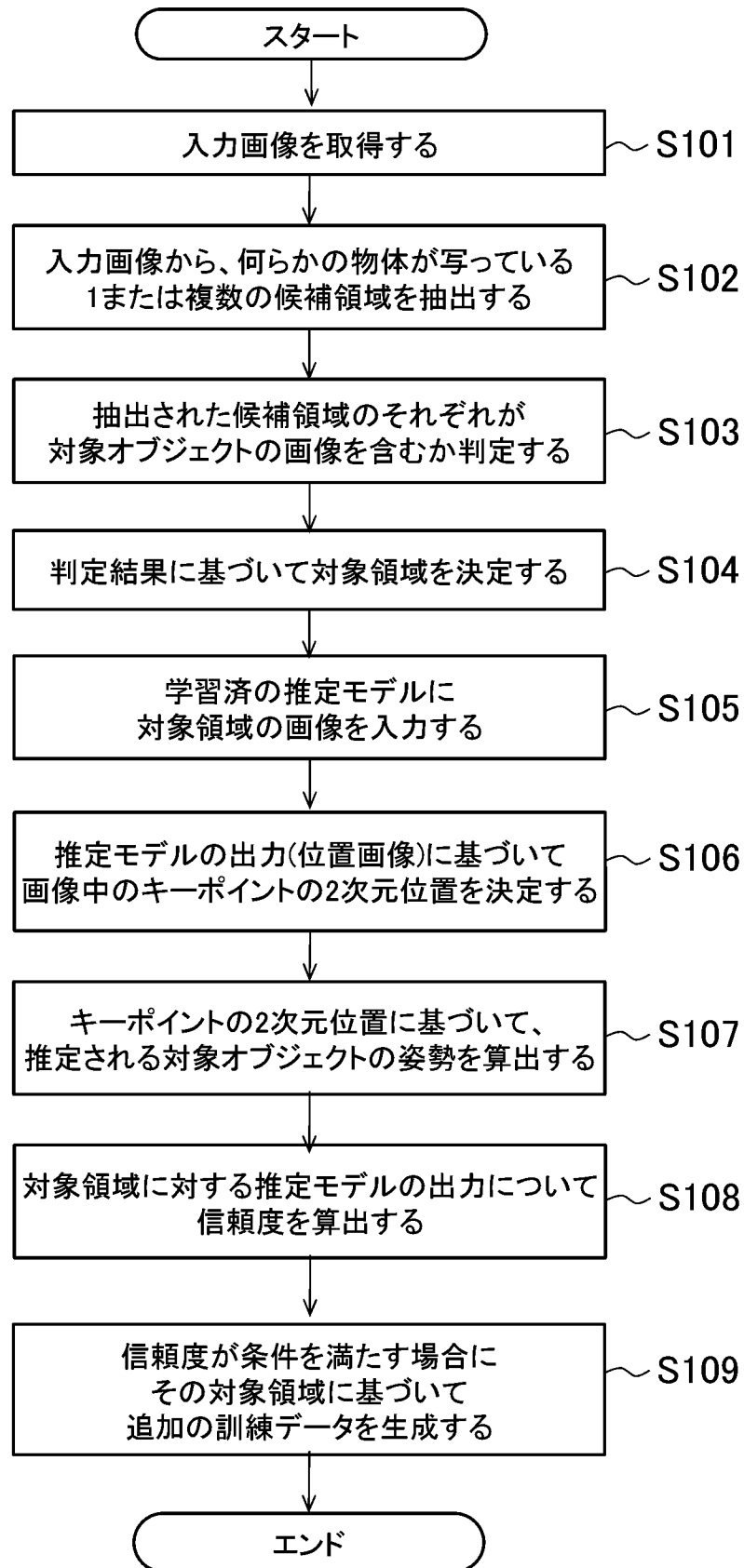
[図4]



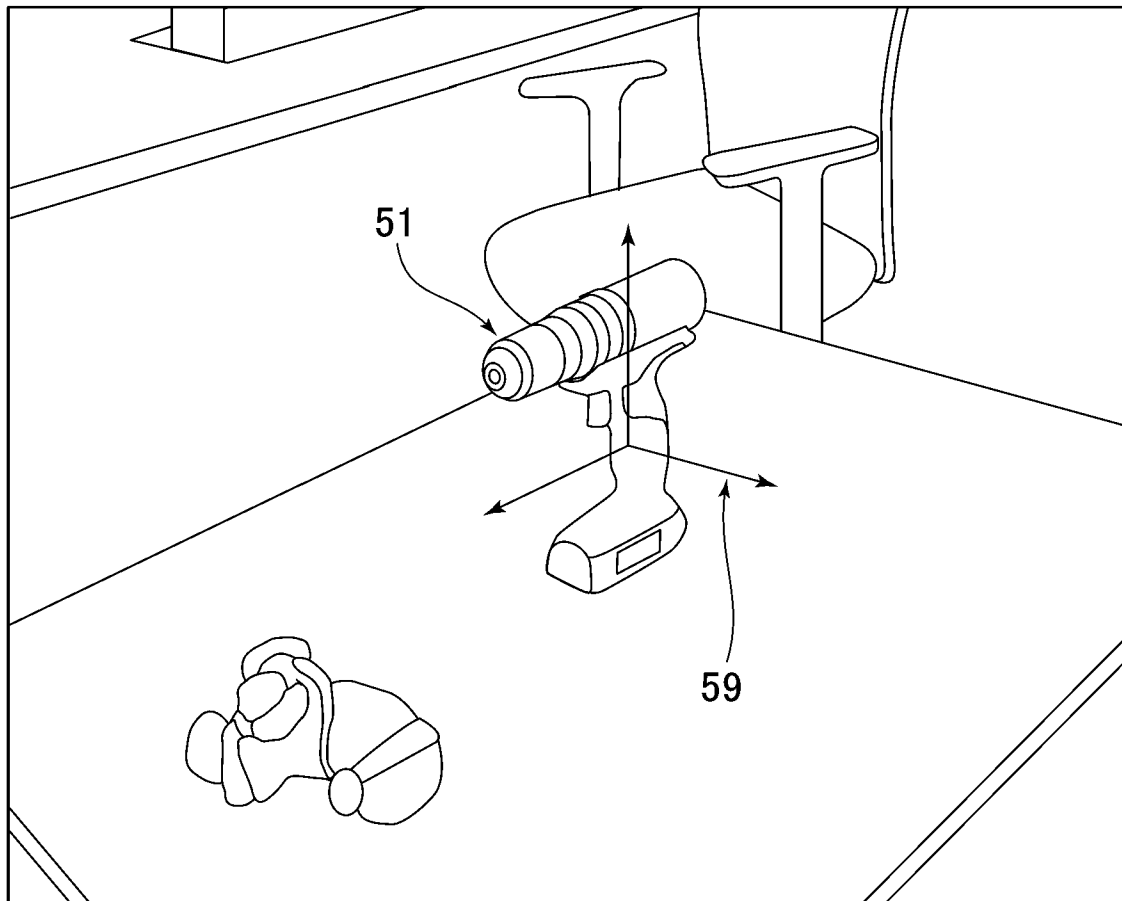
[図5]



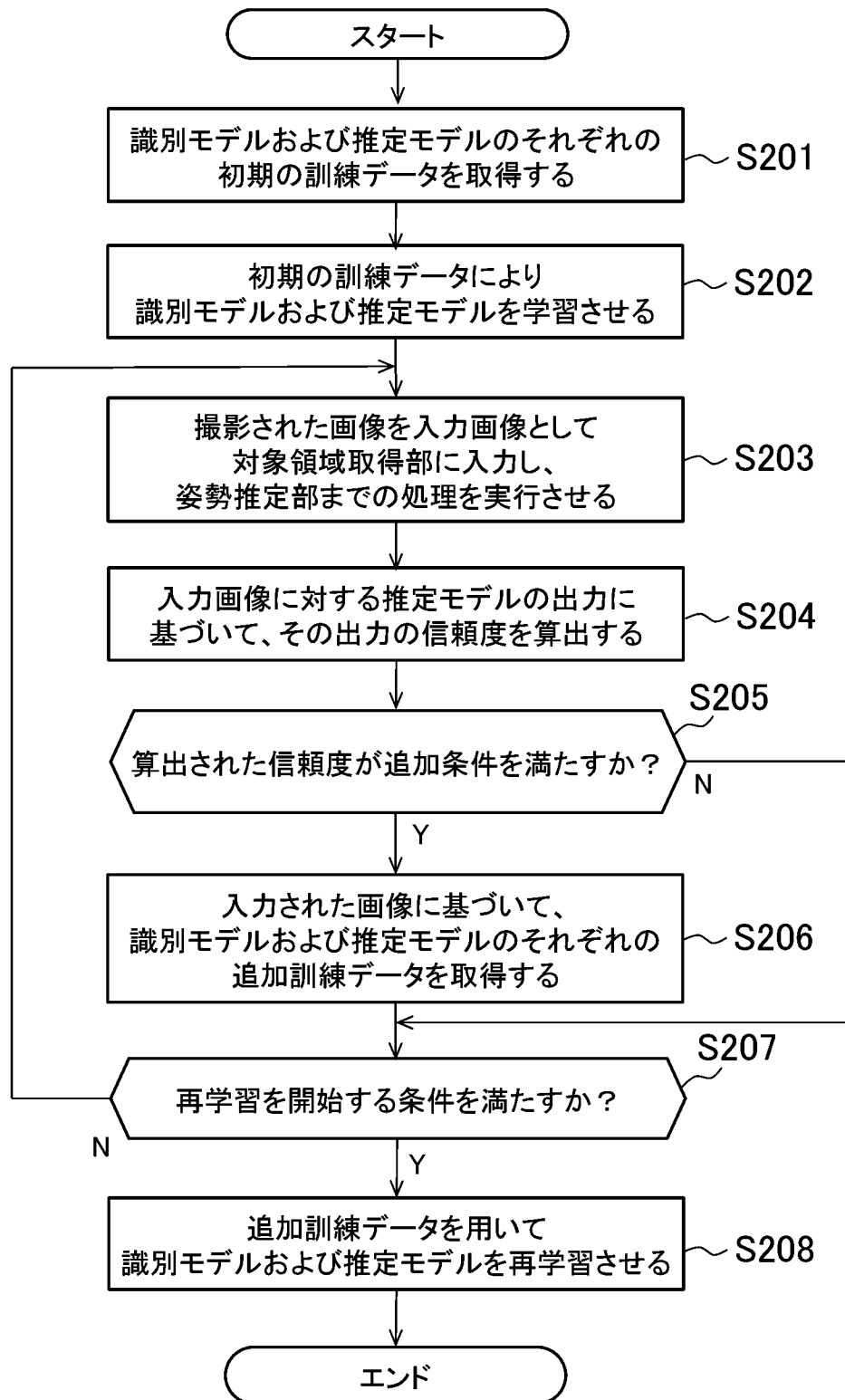
[図6]



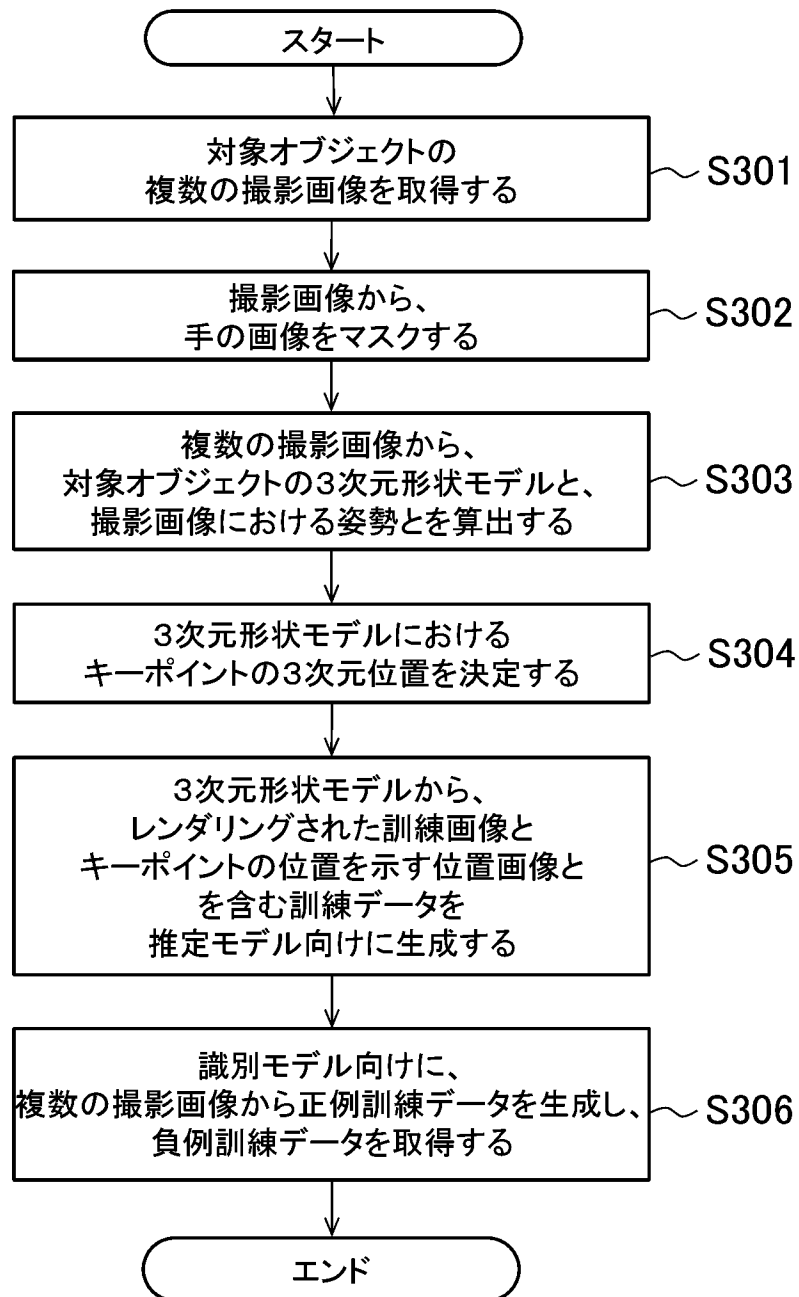
[図7]



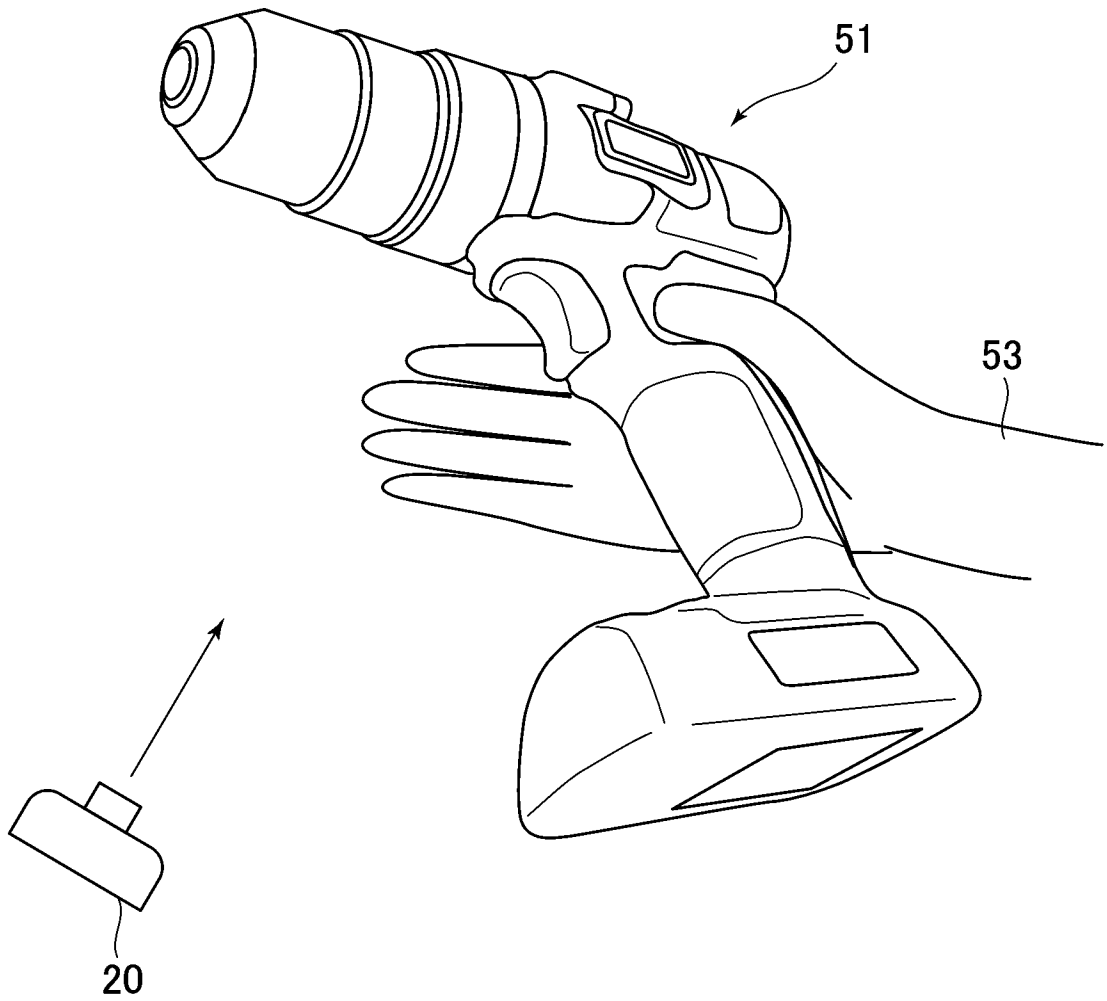
[図8]



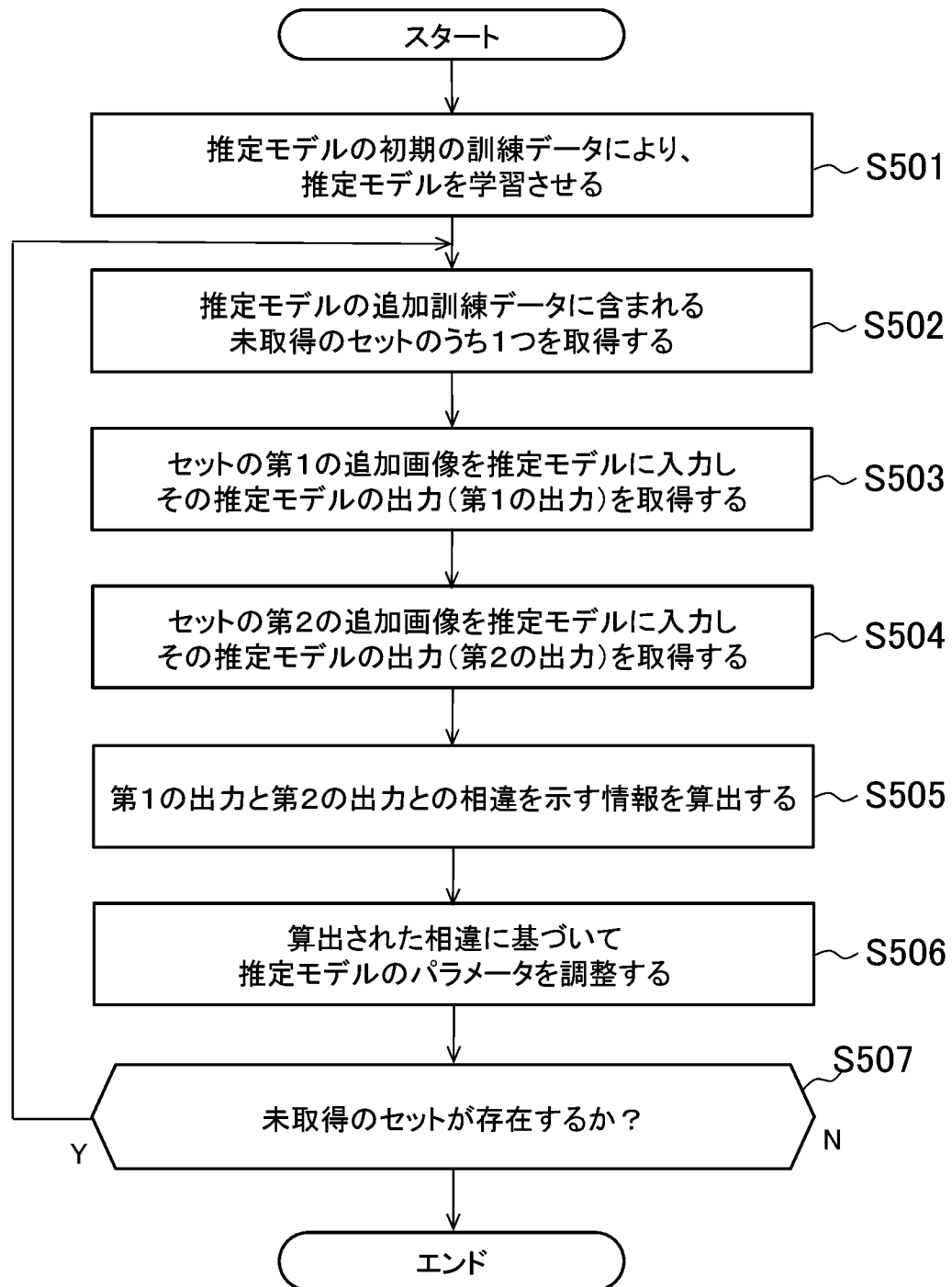
[図9]



[図10]



[図11]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2022/011645

A. CLASSIFICATION OF SUBJECT MATTER**G06N 20/00**(2019.01)i

FI: G06N20/00 130

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)

G06N3/00-99/00; G06T7/00

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 1971-2022
 Registered utility model specifications of Japan 1996-2022
 Published registered utility model applications of Japan 1994-2022

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	CN 112907583 A (SUZHOU KEDA TECHNOLOGY CO., LTD.) 04 June 2021 (2021-06-04)	1-2, 7, 11-12
A	abstract, paragraphs [0053]-[0060], fig. 2	3-6, 8-10
Y	JP 2020-160804 A (MITSUBISHI ELECTRIC CORP) 01 October 2020 (2020-10-01)	1-2, 7, 11-12
A	paragraphs [0042]-[0052]	3-6, 8-10
Y	JP 2021-137747 A (MEIDENSHA ELECTRIC MFG CO LTD) 16 September 2021 (2021-09-16)	1-2, 7, 11-12
A	paragraph [0054]	3-6, 8-10
A	US 2020/0380723 A1 (SEIKO EPSON CORPORATION) 03 December 2020 (2020-12-03)	1-12
A	US 2018/0137644 A1 (QUALCOMM INCORPORATED) 17 May 2018 (2018-05-17)	1-12
A	US 2020/0401793 A1 (INTEL CORPORATION) 24 December 2020 (2020-12-24)	1-12

☐ 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

25 May 2022

Date of mailing of the international search report

07 June 2022

Name and mailing address of the ISA/JP

Japan Patent Office (ISA/JP)
 3-4-3 Kasumigaseki, Chiyoda-ku, Tokyo 100-8915
 Japan

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/JP2022/011645

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
CN	112907583	A	04 June 2021	(Family: none)	
JP	2020-160804	A	01 October 2020	(Family: none)	
JP	2021-137747	A	16 September 2021	(Family: none)	
US	2020/0380723	A1	03 December 2020	(Family: none)	
US	2018/0137644	A1	17 May 2018	WO 2018/089163 A1	
US	2020/0401793	A1	24 December 2020	(Family: none)	

A. 発明の属する分野の分類（国際特許分類（IPC）） G06N 20/00(2019.01)i FI: G06N20/00 130		
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）） G06N3/00-99/00; G06T7/00		
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 1971-2022年 日本国実用新案登録公報 1996-2022年 日本国登録実用新案公報 1994-2022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	CN 112907583 A (SUZHOU KEDA TECHNOLOGY CO., LTD.) 04.06.2021 (2021-06-04) [要約], 段落[0053]-[0060]並びに[図2]	1-2, 7, 11-12 3-6, 8-10
Y A	JP 2020-160804 A (三菱電機株式会社) 01.10.2020 (2020-10-01) 段落[0042]-[0052]	1-2, 7, 11-12 3-6, 8-10
Y A	JP 2021-137747 A (株式会社明電舎) 16.09.2021 (2021-09-16) 段落[0054]	1-2, 7, 11-12 3-6, 8-10
A	US 2020/0380723 A1 (SEIKO EPSON CORPORATION) 03.12.2020 (2020-12-03)	1-12
A	US 2018/0137644 A1 (QUALCOMM INCORPORATED) 17.05.2018 (2018-05-17)	1-12
A	US 2020/0401793 A1 (INTEL CORPORATION) 24.12.2020 (2020-12-24)	1-12
☐ 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抵触するものではなく、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 25.05.2022		国際調査報告の発送日 07.06.2022
名称及びあて先 日本国特許庁(ISA/JP) 〒100-8915 日本国 東京都千代田区霞が関三丁目4番3号		権限のある職員（特許庁審査官） 吉倉 大智 5B 1599 電話番号 03-3581-1101 内線 3545

国際調査報告
 パテントファミリーに関する情報

国際出願番号

PCT/JP2022/011645

引用文献			公表日	パテントファミリー文献	公表日
CN	112907583	A	04.06.2021	(ファミリーなし)	
JP	2020-160804	A	01.10.2020	(ファミリーなし)	
JP	2021-137747	A	16.09.2021	(ファミリーなし)	
US	2020/0380723	A1	03.12.2020	(ファミリーなし)	
US	2018/0137644	A1	17.05.2018	WO 2018/089163 A1	
US	2020/0401793	A1	24.12.2020	(ファミリーなし)	