

(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
2022年11月17日(17.11.2022)

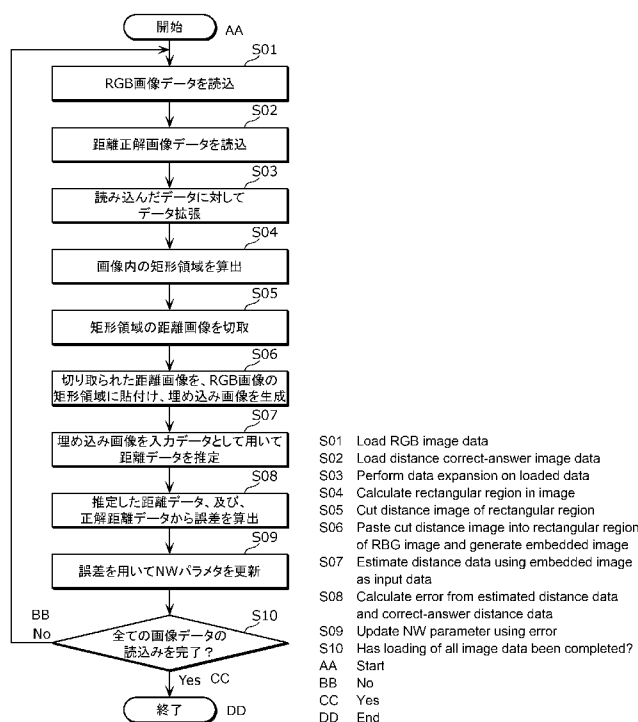


(10) 国際公開番号
WO 2022/239689 A1

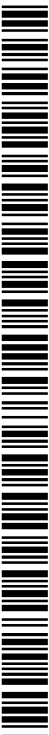
- (51) 国際特許分類: *G06T 7/00* (2017.01) *G06T 7/50* (2017.01) イート 450, ウエスト 190 ストリート 2050 California (US).
- (21) 国際出願番号: PCT/JP2022/019477 (72) 発明者: 石井 育規 (ISHII, Yasunori); 〒5718501 大阪府門真市大字門真1006番地 パナソニックホールディングス株式会社内 Osaka (JP). 遠間 正真 (TOMA, Tadamas). 小山 達也 (KOYAMA, Tatsuya).
- (22) 国際出願日: 2022年5月2日(02.05.2022)
- (25) 国際出願の言語: 日本語
- (26) 国際公開の言語: 日本語
- (30) 優先権データ: 63/188,013 2021年5月13日(13.05.2021) US (74) 代理人: 新居 広守, 外 (NII, Hiromori et al.); 〒5320011 大阪府大阪市淀川区西中島5丁目3番10号タナカ・イトーピア新大阪ビル6階新居国際特許事務所内 Osaka (JP).
- (71) 出願人: パナソニック インテレクチュアル プロパティ コーポレーション オブ アメリカ (PANASONIC INTELLECTUAL PROPERTY CORPORATION OF AMERICA) [US/US]; 90504 カリフォルニア州, トーランス, ス (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,

(54) Title: TRAINING METHOD, TRAINING DEVICE, AND PROGRAM

(54) 発明の名称: 学習方法、学習装置、及び、プログラム



(57) Abstract: This training method includes: acquiring an image and a distance image corresponding to the image (S01, S02); cutting a partial region of the acquired distance image (S05); pasting the distance image within the cut partial region into a prescribed region of the image that corresponds in position and size to the partial region, and thereby generating an embedded image (S06); and training a machine learning model (133) using teaching data in which the generated embedded image is employed as input data and the distance image is employed as correct-answer data (S07-S09).



WO 2022/239689 A1

CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO,
DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT,
HN, HR, HU, ID, IL, IN, IQ, 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, SD, SL, ST, SZ, TZ, UG, ZM, ZW), ユーラシア (AM, AZ, BY, KG, KZ, RU, TJ, TM), ヨーロッパ (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

添付公開書類 :

- 一 国際調査報告 (条約第21条(3))
-

(57) 要約 : 学習方法は、画像と、画像に対応する距離画像とを取得し (S 0 1、S 0 2)、取得された距離画像の一部の領域を切り取り (S 0 5)、切り取られた一部の領域の距離画像を、画像において一部の領域と位置及び大きさが対応する所定の領域に貼り付けることにより、埋め込み画像を生成し (S 0 6)、生成された埋め込み画像を入力データとし、距離画像を正解データとする教師データを用いて機械学習モデル (1 3 3) の学習を行う (S 0 7~S 0 9)。

明 細 書

発明の名称：学習方法、学習装置、及び、プログラム

技術分野

[0001] 本開示は、機械学習モデルを学習させる学習方法、学習装置、及び、プログラムに関する。

背景技術

[0002] 非特許文献1では、RGB画像を入力データとし、距離画像を正解データとする教師データを用いて機械学習モデルを学習する学習方法が開示されている。さらに、非特許文献1では、学習済みの機械学習モデルを用いて、RGB画像から距離画像を推定する際に、法線推定を行うことにより、従来よりも平面の推定精度を向上させることができることが開示されている。

先行技術文献

非特許文献

[0003] 非特許文献1: Jin Han Lee et al. "From Big to Small: Multi-Scale Local Planar Guidance for Monocular Depth Estimation". <https://doi.org/10.48550/arXiv.1907.10326>

発明の概要

発明が解決しようとする課題

[0004] しかしながら、非特許文献1に開示される学習方法で用いられるデータ拡張を行って教師データの数を増やしても、単眼Depth推定において多様なシーンに対するロバスト性が向上しにくいという課題がある。

[0005] 本開示は、上述の事情を鑑みてなされたもので、単眼Depth推定において多様なシーンに対するロバスト性を向上させることができる学習方法、及び、プログラムを提供することを目的とする。

課題を解決するための手段

[0006] 上記目的を達成するために、本開示の一形態に係る学習方法は、画像と、前記画像に対応する距離画像とを取得し、取得された前記距離画像のうちの一部の領域を切り取り、切り取られた前記一部の領域の距離画像を、前記画像において前記一部の領域と位置及び大きさが対応する所定の領域に貼り付けることにより、埋め込み画像を生成し、生成された前記埋め込み画像を入力データとし、前記距離画像を正解データとする教師データを用いて機械学習モデルの学習を行う。

[0007] なお、これらの全般的または具体的な態様は、装置、方法、集積回路、コンピュータプログラムまたはコンピュータで読み取り可能なCD-ROMなどの記録媒体で実現されてもよく、システム、方法、集積回路、コンピュータプログラム及び記録媒体の任意な組み合わせで実現されてもよい。

発明の効果

[0008] 本開示により、単眼Depth推定において多様なシーンに対するロバスト性を向上させることができる学習方法、及び、プログラムを提供できる。

図面の簡単な説明

[0009] [図1]図1は、実施の形態に係る学習装置を備える学習システムの機能構成を示すブロック図である。

[図2]図2は、埋め込み画像の生成方法を説明するための図である。

[図3]図3は、埋め込み画像の一例を示す図である。

[図4]図4は、埋め込み画像の他の例を示す図である。

[図5]図5は、機械学習モデルの例を模式的に示す図である。

[図6]図6は、実施の形態に係る学習装置の動作の一例を示すフローチャートである。

[図7]図7は、実施の形態に係る学習装置を備える推定システムの機能構成の一例を示すブロック図である。

[図8]図8は、推定装置の動作の一例を示すフローチャートである。

[図9]図9は、実験例1の結果を示す図である。

[図10]図10は、実験例2の結果を示す図である。

発明を実施するための形態

[0010] 以下で説明する実施の形態は、いずれも本開示の一具体例を示すものである。以下の実施の形態で示される数値、形状、構成要素、ステップ、ステップの順序などは、一例であり、本開示を限定する主旨ではない。また、以下の実施の形態における構成要素のうち、独立請求項に記載されていない構成要素については、任意の構成要素として説明される。また全ての実施の形態において、各々の内容を組み合わせることもできる。

[0011] (実施の形態)

以下、本実施の形態に係る学習装置及び学習方法について説明する。

[0012] [1. 構成]

本実施の形態に係る学習装置は、例えば、メモリ及びプロセッサ（マイクロプロセッサ）を含むコンピュータを備え、プロセッサがメモリに格納された制御プログラムを実行することにより、各種機能を実現し、機械学習モデルを学習させる。

[0013] 図1は、実施の形態に係る学習装置を備える学習システムの機能構成の一例を示す図である。学習システム200は、例えば、RGBカメラ10と、測距センサ20と、学習装置100とを備える。本実施の形態では、画像は、RGBの3チャンネルからなるRGB画像である例を説明するが、この例に限られない。画像は、例えば、モノクロ画像、赤外線画像、又は、3次元点群の座標データであってもよい。

[0014] [RGBカメラ10、測距センサ20]

RGBカメラ10は、RGB画像を撮影し、測距センサ20は、RGBカメラ10が撮影するRGB画像に対応する距離画像を撮影する。距離画像の各画素は、RGB画像の各画素に写る対象までの距離が格納されている。RGBカメラと距離を取得するセンサとの位置関係が予めキャリブレーションされていれば、距離画像とRGB画像との視点をそろえることができる。これにより、RGB画像と距離画像とは、物体の構造関係が互いに類似する画像となる。例えば、距離画像及びRGB画像は、ほぼ同じサイズで、かつ、

同じ物体が映っており、ほぼ同じ構造を有する。ほぼ同じ構造を有するとは、RGB画像及び距離画像からそれぞれエッジを算出したときに、距離が変化する箇所のエッジはほぼ同じ位置（つまり、完全一致ではないがほぼ同じ位置）に存在することをいう。2次元の情報しかないRGB画像であっても、距離が変化するエッジの位置がわかれば、シーンの3次元変化が生じる位置を認識することができる。距離画像及びRGB画像がほぼ同じ構造を有するとき、RGB画像及び距離画像でシーンの3次元変化する位置はほぼ同じ画素である。距離画像は、機械学習モデル133の学習用の教師データの正解データ（以下、正解距離画像データともいう）として使用される。RGBカメラ及び測距センサ20は、例えば、1つのセンサ装置に備えられてもよく、上下又は左右に並んで配置されてもよい。RGBカメラ10は、例えば、単眼カメラである。測距センサ20は、例えば、ステレオカメラ又はToF (Time of Flight) カメラである。なお、距離画像は、画像でなくてもよい。例えば、距離画像は、RGB画像とは異なるデータ型でもあってもよく、測距センサで取得した距離データを行列に置き換えたものであってもよい。そのため、測距センサは距離データの行列を含むデータを取得できる手段であれば特に限定されない。例えば、測距センサ20は、測距センサ20は、LiDAR (Light Detection and Ranging) であってもよい。なお、距離データは測距センサからの距離情報又は3次元空間中の任意の位置を原点とした3次元座標を格納した値であってもよい。距離情報は、実際の距離の値であっても、特定の距離を基準とした相対距離であってもよい。

[0015] [学習装置100]

図1に示されるように、学習装置100は、例えば、通信部110と、情報処理部120と、記憶部130とを備える。情報処理部120は、例えば、RGB画像取得部121と、距離画像取得部122と、データ拡張処理部123と、埋め込み画像生成部124と、学習部125とを備える。なお、学習装置100において、通信部110及びデータ拡張処理部123を備え

ることは必須ではない。

[0016] [通信部 110]

通信部 110 は、学習装置 100 が RGB カメラ 10 及び測距センサ 20 と通信を行うための通信回路（通信モジュール）である。通信部 110 は、局所通信ネットワークを介して通信を行うための通信回路（通信モジュール）を備えるが、広域通信ネットワークを介して通信を行うための通信回路（通信モジュール）を備えてもよい。通信部 110 は、例えば、無線通信を行う無線通信であるが、有線通信を行う有線通信回路であってもよい。なお、通信部 110 が行う通信の通信規格については特に限定されない。

[0017] [情報処理部 120]

情報処理部 120 は、学習装置 100 に関する各種情報処理を行う。より具体的には、例えば、情報処理部 120 は、通信部 110 により受信された RGB 画像データ及び距離画像データを記憶部 130 内の画像データベース 131 に格納する。また、例えば、情報処理部 120 は、画像データベース 131 に格納された RGB 画像データ及び当該 RGB 画像に対応する距離画像データを読み込んで、機械学習モデルの学習用データの入力画像を生成し、生成した入力画像と正解距離画像との組を用いて機械学習モデルを学習させる。

[0018] 情報処理部 12 は、具体的には、RGB 画像取得部 121 と、距離画像取得部 122 と、データ拡張処理部 123 と、埋め込み画像生成部 124 と、学習部 125 とを備える。RGB 画像取得部 121、距離画像取得部 122、データ拡張処理部 123、埋め込み画像生成部 124、及び、学習部 125 の機能は、情報処理部 120 を構成するプロセッサ又はマイクロコンピュータが記憶部 130 に記憶されたコンピュータプログラムを実行することによって実現される。

[0019] [RGB 画像取得部 121]

RGB 画像取得部 121 は、記憶部 130 内の画像データベース 131 に格納された RGB 画像データを読み込み、データ拡張処理部 123 及び埋め

込み画像生成部 124 に出力する。

[0020] [距離画像取得部 122]

距離画像取得部 122 は、記憶部 130 内の画像データベース 131 に格納された距離画像データを読み込み、データ拡張処理部 123 及び埋め込み画像生成部 124 に出力する。より具体的には、例えば、距離画像取得部 122 は、RGB 画像取得部 121 により画像データベース 131 から読み込まれた RGB 画像データに対応する距離画像データを画像データベース 131 から読み込む。この距離画像データは、RGB 画像とほぼ同じサイズで、かつ、同じ物体が映っており、ほぼ同じ構造を有する。この距離画像は、教師データの正解データ（正解距離画像データ）として使用される。

[0021] [データ拡張処理部 123]

データ拡張処理部 123 は、取得した RGB 画像データ及び距離画像データに対してデータ拡張処理を行い、M 個（M は 2 以上の整数）の RGB 画像データ及び当該 RGB 画像データに対応する距離画像データを得る。データ拡張処理部 123 は、M 個（M は 2 以上の整数）の RGB 画像データ及び距離画像データを埋め込み画像生成部 124 に出力する。

[0022] ここで、データ拡張処理は、画像データに対して変換処理を行うことで画像データを水増しする手法である。例えば、データ拡張処理部 123 は、取得した RGB 画像データ及び距離画像データに対して、例えば回転処理、ズーム処理、並進処理、又は、色変換処理などのデータの変換処理を行う。このような変換処理を行うことにより、データ拡張処理部 123 は、当該 RGB 画像データ及び距離画像データのデータセットを M 個の RGB 画像データ及び距離画像データのデータセットにデータを拡張する（言い換えると、データを水増しする）。

[0023] [埋め込み画像生成部 124]

埋め込み画像生成部 124 は、取得した M 個の RGB 画像データ及び距離画像データのデータセットのそれぞれについて、距離画像の一部の領域を切り取り、RGB 画像において当該一部の領域と位置及び大きさが対応する所

定の領域に貼り付けることにより、埋め込み画像を生成する。一部の領域は、RGB画像に映る物体の輪郭を示すエッジ部分を含む。所定の領域は、例えば、RGB画像の25%以上75%以下の面積であり、より好ましくは、RGB画像の30%以上70%以下の面積であり、さらに好ましくは、RGB画像の40%以上60%以下の面積であってもよい。特に、所定の領域は、RGB画像の50%の面積であることが好ましい。埋め込み画像生成部124は、生成した埋め込み画像を機械学習モデル133の学習用の入力データとし、距離画像データを出力データ（正解データ）とする教師データを生成する。なお、埋め込み画像生成部124の前、又は、埋め込み画像生成部124の後、すなわち埋め込み画像生成部124と学習部125との間に、画像サイズの調整、標準化などの前処理を行うデータ前処理部を備えてもよい。

[0024] 図2は、埋め込み画像の生成方法を説明するための図である。図2に示されるように、埋め込み画像生成部124は、RGB画像の所定の領域を距離画像データに置き換える、距離画像内の矩形領域の位置（例えば、左上角の座標 (x_1, y_1) ）及び大きさ（例えば、高さ $h \times$ 幅 w ）と、当該矩形領域に対応するRGB画像内の所定の矩形領域の位置（例えば、左上角の座標 (x_1, y_1) ）及び大きさ（例えば、高さ $h \times$ 幅 w ）とを算出する。埋め込み画像生成部124は、算出した矩形領域の距離画像を切り取り、RGB画像の矩形領域に貼り付けて、埋め込み画像を生成する。

[0025] 上記の方法で生成された埋め込み画像の例を示す。図3は、埋め込み画像の一例を示す図である。図4は、埋め込み画像の他の例を示す図である。

[0026] 図3では、埋め込み画像は、RGB画像の30%の面積を正解距離画像で置き換えられている。図4では、埋め込み画像は、RGB画像の70%の面積を正解距離画像で置き換えられている。図3及び図4に示されるように、埋め込み画像生成部124は、RGB画像内の所定の矩形領域の位置及び大きさと、当該所定の矩形領域に対応する、距離画像内の矩形領域の位置及び大きさとを算出し、算出した正解距離画像内の矩形領域のデータを切り取り

、RGB画像の所定の矩形領域に貼り付けて、埋め込み画像を生成する。より具体的には、埋め込み画像生成部124は、矩形領域の左上角の座標はランダムに決定し、左上角から矩形領域の幅と高さとを、RGB画像の面積に対して最大何%にするかの最大値（上記の25%以上75%以下）を決定し、その最大値の範囲内で矩形領域の大きさを決定する。このとき、矩形領域は、距離画像に映る物体のエッジ部分を含むように決定される。例えば、複数の物体が映る距離画像では、複数の物体の輪郭を示すエッジ部分を含むように矩形領域が決定されるとよい。これにより、距離画像において複数の物体の距離が異なる部分のエッジのみが情報として残るため、余計な情報を取らずに、距離に関する情報のみを、機械学習モデル133に効率よく学習させることができる。

[0027] [学習部125]

学習部125は、教師データを用いて機械学習モデル133を学習させる。教師データは、埋め込み画像生成部124で生成された埋め込み画像を入力データとし、距離画像を出力データ（いわゆる、正解データ）とする一組のデータセットである。

[0028] 学習部125は、埋め込み画像を機械学習モデル133に入力して出力される距離画像データと、正解データ（正解距離画像データ）との誤差を算出し、誤差を用いて機械学習モデル133の重みなどのネットワーク（NW）パラメタを更新する。学習部125は、更新したネットワークパラメタを学習パラメタデータベース132に格納する。

[0029] なお、パラメタの更新方法は、特に限定されないが、例えば、勾配降下法などが挙げられる。また、誤差もL2誤差などが挙げられるが、特に限定されない。

[0030] [記憶部]

記憶部130は、情報処理部120が各種情報処理を実行するための専用のアプリケーションプログラムなどが記憶される記憶装置である。例えば、記憶部130には、画像データベース131と、学習パラメタデータベース

132と、機械学習モデル133とが格納されている。記憶部130は、例えば、HDD (Hard Disk Drive) によって実現されるが、半導体メモリによって実現されてもよい。

[0031] 画像データベース131は、RGBカメラ10及び測距センサ20から受信されたRGB画像データ及び距離画像データを格納する。学習パラメータデータベース132は、学習部125により更新されたネットワークパラメータを格納する。

[0032] 機械学習モデル133は、学習装置100が学習させる機械学習モデルである。図5は、機械学習モデルの構造の一例を模式的に示す図である。

[0033] 機械学習モデル133は、学習装置100が学習させる機械学習モデルである。機械学習モデル133は、RGB画像を入力として、距離画像を出力する。例えば、機械学習モデル133は、図5の(a)に示されるように、エンコーダネットワークモデルと、出力層とで構成される。

[0034] エンコーダネットワークモデルは、入力されたRGB画像データの特徴表現を抽出する。なお、エンコーダネットワークモデルは、例えば複数の畳み込み層で構成されるCNN (Convolution Neural Networks) モデルであるが、これに限らない。エンコーダネットワークモデルは、ResNet (Residual Network) で構成されていてもよいし、MobileNetで構成されていてもよいし、Transformerで構成されていてもよい。

[0035] 出力層は、エンコーダネットワークモデルの最終層から出力される低次元の特徴表現をアップサンプリングして入力画像と同じサイズ of 出力画像を生成する。より具体的には、出力層は、エンコーダネットワークモデルの最終層から出力された距離データの行列 (1×幅×高さ) をアップサンプリングして、学習モデル133 (エンコーダネットワークモデル) に入力された入力データと同じ大きさの行列に変換して出力する。出力層は、図5の(b)に示されるように、デコーダネットワークモデルであってもよい。

[0036] なお、エンコーダネットワークモデルと最終層 (例えば、デコーダネット

ワークモデル)との間に、Skip Connection又はSPP (Spatial Pyramid Pooling)などが存在してもよい。

[0037] [2. 動作]

続いて、実施の形態に係る学習装置100の動作について説明する。図6は、実施の形態に係る学習装置の動作の一例を示すフローチャートである。

[0038] 図6に示されるように、まず、学習装置100は、記憶部130の画像データベース131に格納されたRGB画像データを読み込む(S01)。次に、学習装置100は、画像データベース131に格納された距離画像(いわゆる、正解距離画像)データを読み込む(S02)。なお、ステップS02で読み込んだ距離画像データは、ステップS01で読み込んだRGB画像データに対応する画像データであり、RGB画像を用いて距離データを推定したときの正解距離データである。

[0039] 次に、学習装置100は、ステップS01及びステップS02で読み込んだデータに対してデータ拡張を行い(S03)、M個(Mは2以上の整数)のRGB画像データ及び当該RGB画像データに対応する正解距離画像データを取得する。

[0040] 次に、学習装置100は、RGB画像及び正解距離画像内の矩形領域を算出する(S04)。より具体的には、学習装置100は、RGB画像内の所定の領域を正解距離画像に置き換える、正解距離画像内の矩形領域の位置(例えば、左上角の座標)及び大きさ(高さ×幅)と、当該矩形領域に対応するRGB画像内の矩形領域の位置とを算出する。

[0041] 次に、学習装置100は、正解距離画像から矩形領域の距離画像を切り取り(S05)、切り取られた距離画像を、RGB画像の矩形領域に貼り付け、埋め込み画像を生成する(S06)。

[0042] 次に、学習装置100は、ステップS06で生成された埋め込み画像を教師データの入力データとして用いて距離データを推定する(S07)。より具体的には、学習装置100は、機械学習モデル133に埋め込み画像を入力して距離データを推論させる。

[0043] 次に、学習装置100は、ステップS07で推定した距離データ及び正解距離データから誤差を算出し（S08）、誤差を用いてネットワーク（NW）パラメタを更新する（S09）。

[0044] 次に、学習装置100は、全ての画像データの読み込みを完了したか否かを判定し（S10）、読込を完了していないと判定した場合（S10でNo）、ステップS01に戻り、読込を完了したと判定した場合（S10でYes）、動作を終了する。

[0045] [3. 効果等]

以上説明したように、本実施の形態に係る学習方法は、画像と、画像に対応する距離画像とを取得し（図6のS01、S02）、取得された距離画像の一部の領域を切り取り（S05）、切り取られた一部の領域の距離画像を、画像において一部の領域と位置及び大きさが対応する所定の領域に貼り付けることにより、埋め込み画像を生成し（S06）、生成された埋め込み画像を入力データとし、距離画像を正解データとする教師データを用いて機械学習モデル133の学習を行う（S07、S08、S09）。

[0046] これにより、本実施の形態に係る学習方法によれば、画像に貼り付ける距離画像は、物体の色情報及びテクスチャ情報がないため、埋め込み画像を使用することで、色及びテクスチャの変動に対してロバスト性を向上させる学習を行うことができる。したがって、本実施の形態に係る学習方法によれば、単眼Depth推定において多様なシーンに対するロバスト性を向上させることができる。

[0047] 例えば、本実施の形態に係る学習方法では、所定の領域は、画像の25%以上75%以下の面積である。

[0048] これにより、本実施の形態に係る学習方法によれば、所定の領域を上記の割合に調整して画像に距離画像を貼り付けることにより、単眼Depth推定において多様なシーンに対するロバスト性をより向上させることができる。

[0049] 例えば、本実施の形態に係る学習方法では、一部の領域は、距離画像に映

る物体の輪郭を示すエッジ部分を含む。

- [0050] これにより、本実施の形態に係る学習方法によれば、距離画像において物体の輪郭を示すエッジ部分を含む一部の領域を画像に貼り付けることにより、機械学習モデル133に、距離が違うエッジから距離に関する情報を学習させることができる。したがって、本実施の形態に係る学習方法によれば、余計な情報を取らずに、距離に関する情報のみを、機械学習モデル133に効率よく学習させることができる。
- [0051] 例えば、本実施の形態に係る学習方法では、画像と距離画像との関係性を機械学習モデルに学習させる。
- [0052] これにより、本実施の形態に係る学習方法によれば、画像から抽出された特徴量に基づいて距離画像を推定することができるように機械学習モデル133を学習させることができる。
- [0053] 例えば、本実施の形態に係る学習方法では、機械学習モデル133は、エンコーダネットワークモデルと、エンコーダネットワークモデルから出力される低次元の特徴表現を画像と同じ大きさの出力画像にアップサンプリングする出力層とで構成される。
- [0054] これにより、実施の形態に係る学習方法によれば、エンコーダネットワークモデルで抽出されて出力される低次元の特徴表現を入力データと同じサイズの出力データに変換して出力することができる。
- [0055] 例えば、本実施の形態に係る学習方法では、機械学習モデルは、エンコーダネットワークモデルと、デコーダネットワークモデルとで構成される。
- [0056] これにより、実施の形態に係る学習方法によれば、エンコーダネットワークモデルで抽出されて出力される低次元の特徴表現を段階的にアップサンプリングすることにより、入力データと同じサイズの出力データに変換して出力することができる。
- [0057] また、本実施の形態に係る学習装置は、画像と、画像に対応する距離画像とを取得し、取得された距離画像のうちの一部の領域を切り取り、切り取られた一部の領域の距離画像を、画像において一部の領域と位置及び大きさが

対応する所定の領域に貼り付けることにより、埋め込み画像を生成する埋め込み画像生成部と、生成された埋め込み画像を入力データとし、距離画像を正解データとする教師データを用いて機械学習モデルの学習を行う学習部と、を備える。

[0058] これにより、実施の形態に係る学習装置は、画像に貼り付ける距離画像は、物体の色情報及びテクスチャ情報がないため、埋め込み画像を使用することで、色及びテクスチャの変動に対してロバスト性を向上させる学習を行うことができる。したがって、本実施の形態に係る学習装置は、単眼Depth推定において多様なシーンに対するロバスト性を向上させることができる。

[0059] また、本実施の形態に係るプログラムは、上記学習方法をコンピュータに実行させるためのプログラムである。

[0060] これにより、本実施の形態に係るプログラムによれば、上記学習方法と同様の効果を奏することができる。

[0061] [4. 適用例]

続いて、実施の形態に係る学習装置100の適用例について説明する。実施の形態に係る学習装置を備える推定システムの機能構成の一例を示すブロック図である。

[0062] 図7に示されるように、推定システム400は、例えば、学習装置100と、推定装置300とを備える。図7の例では、推定装置300は、学習装置100と別体で設けられているが、例えば、推定装置300は、学習装置100を備えてもよい。

[0063] 推定装置300は、RGB画像を用いて距離データを推定する。推定装置300は、例えば、車又は移動ロボットなどの移動体に適用されてもよく、建物内の監視システムなどに適用されてもよい。

[0064] 図示していないが、図7の例では、推定装置300は、学習パラメータデータベースと、学習装置100の機械学習モデル133と同じ機械学習モデルとを備えている。推定装置300は、学習装置100によりネットワークパ

ラメタが更新されると、更新されたネットワークパラメタを受信して学習パラメタデータベースに保存する。

[0065] 図8は、推定装置300の動作の一例を示すフローチャートである。図8に示されるように、推定装置300は、記憶部（図7で不図示）に格納されたRGB画像データを読み込む（S11）。

[0066] 次に、推定装置300は、RGB画像を用いて距離データを推定する（S12）。より具体的には、推定装置300は、機械学習モデル（不図示）にRGB画像を入力して距離データを推論させる。

[0067] 次に、推定装置300は、全ての画像データの読み込みを完了したか否かを判定し（S13）、全ての画像の読み込みを完了していないと判定した場合（S13でNo）、ステップS11に戻る。一方、推定部33は、全ての画像の読み込みを完了したと判定した場合（S13でYes）、動作を終了する。

[0068] [5. 実験例]

続いて、実験例を用いて本開示の学習方法を具体的に説明する。以下の実験例では、本開示の学習方法を用いて学習された機械学習モデル、及び、従来の学習方法を用いて学習された機械学習モデルの推定精度を評価した。学習済みの機械学習モデルには、RGB画像を入力した。

[0069] なお、従来の学習方法は、RGB画像を入力データとし、距離画像を正解の出力データとする教師データを用いて学習を行う方法である。

[0070] [実験例1]

実験例1では、単眼Depth推定アルゴリズムとして、非特許文献1に記載のFrom Big to Small (Bts) アルゴリズムを使用した。実験例1では、このBtsアルゴリズムに従来の学習方法（以下、従来手法）と本開示の学習方法（以下、本願手法）とを適用した。本開示の学習方法では、教師画像の入力データとして、埋込率（%）の異なる埋め込み画像を使用した。埋込率は、RGB画像に貼り付けられた正解距離画像の割合を示す。

[0071] B t s アルゴリズムには、埋め込み画像の生成に使用された R G B 画像を入力し、出力される距離画像と正解距離画像との誤差を算出した。誤差の算出には、二乗平均平方根 (r m s)、絶対相対誤差 (A b s _ r e l)、 $\log 10$ 、 $\log _ r m s$ が用いられた。結果を図 9 に示す。図 9 は、実験例 1 の結果を示す図である。

[0072] 図 9 に示されるように、B t s アルゴリズムの学習に本願手法を適用すると、いずれの埋込率においても、従来手法よりも推定精度が向上することが確認された。特に、埋込率 50% では、r m s、 $\log 10$ 及び $\log _ r m s$ の値が最も小さくなった。これにより、埋込率 50% の埋め込み画像を教師データの入力データとして使用すると、単眼 D e p t h 推定における推定精度が最も高くなることが確認された。

[0073] したがって、実験例 1 の結果から、本願手法によれば、単眼 D e p t h 推定において多様なシーンに対するロバスト性を向上させることができることが確認された。

[0074] [実験例 2]

実験例 2 では、単眼 D e p t h 推定アルゴリズムとして、L a p l a c i a n d e p t h (l a p D e p t h) アルゴリズムを使用したこと以外、実験例 1 と同様に行った。結果を図 10 に示す。図 10 は、実験例 2 の結果を示す図である。

[0075] 図 10 に示されるように、l a p D e p t h アルゴリズムの学習に本願手法を適用すると、いずれの埋込率においても、従来手法よりも推定精度が向上することが確認された。特に、埋込率 50% では、r m s、a b s _ r e l、 $\log 10$ 及び $\log _ r m s$ の値が最も小さくなった。これにより、埋込率 50% の埋め込み画像を教師データの入力データとして使用すると、単眼 D e p t h 推定における推定精度が最も高くなることが確認された。

[0076] したがって、実験例 2 の結果から、本願手法によれば、単眼 D e p t h 推定において多様なシーンに対するロバスト性を向上させることができることが確認された。

[0077] (その他の実施の形態)

以上、本開示に係る学習方法について、各実施の形態に基づいて説明したが、本開示は、これらの実施の形態に限定されるものではない。本開示の主旨を逸脱しない限り、当業者が思いつく各種変形を各実施の形態に施したもののや、各実施の形態における一部の構成要素を組み合わせて構築される別の形態も、本開示の範囲内に含まれる。

[0078] また、以下に示す形態も、本開示の一つ又は複数の態様の範囲内に含まれてもよい。

[0079] (1) 上記の学習方法を実現する学習装置を構成する構成要素の一部は、マイクロプロセッサ、ROM、RAM、ハードディスクユニット、ディスプレイユニット、キーボード、マウスなどから構成されるコンピュータシステムであってもよい。前記RAM又はハードディスクユニットには、コンピュータプログラムが記憶されている。前記マイクロプロセッサが、前記コンピュータプログラムにしたがって動作することにより、その機能を達成する。ここでコンピュータプログラムは、所定の機能を達成するために、コンピュータに対する指令を示す命令コードが複数個組み合わせられて構成されたものである。

[0080] (2) 上記の学習方法を実現する学習装置を構成する構成要素の一部は、1個のシステムLSI (Large Scale Integration : 大規模集積回路) から構成されているとしてもよい。システムLSIは、複数の構成部を1個のチップ上に集積して製造された超多機能LSIであり、具体的には、マイクロプロセッサ、ROM、RAMなどを含んで構成されるコンピュータシステムである。前記RAMには、コンピュータプログラムが記憶されている。前記マイクロプロセッサが、前記コンピュータプログラムにしたがって動作することにより、システムLSIは、その機能を達成する。

[0081] (3) 上記の学習方法を実現する学習装置を構成する構成要素の一部は、各装置に脱着可能なICカード又は単体のモジュールから構成されていると

してもよい。前記ICカード又は前記モジュールは、マイクロプロセッサ、ROM、RAMなどから構成されるコンピュータシステムである。前記ICカード又は前記モジュールは、上記の超多機能LSIを含むとしてもよい。マイクロプロセッサが、コンピュータプログラムにしたがって動作することにより、前記ICカード又は前記モジュールは、その機能を達成する。このICカード又はこのモジュールは、耐タンパ性を有するとしてもよい。

[0082] (4) また、上記の学習方法を実現する学習装置を構成する構成要素の一部は、前記コンピュータプログラム又は前記デジタル信号をコンピュータで読み取り可能な記録媒体、例えば、フレキシブルディスク、ハードディスク、CD-ROM、MO、DVD、DVD-ROM、DVD-RAM、BD (Blu-ray (登録商標) Disc)、半導体メモリなどに記録したものであるとしてもよい。また、これらの記録媒体に記録されている前記デジタル信号であるとしてもよい。

[0083] 例えば、上記の学習方法を実現するコンピュータプログラムは、画像と、前記画像に対応する距離画像とを取得し、取得された前記距離画像のうちの一部の領域を切り取り、切り取られた前記一部の領域の距離画像を、前記画像において前記一部の領域と位置及び大きさが対応する所定の領域に貼り付けることにより、埋め込み画像を生成し、生成された前記埋め込み画像を入力データとし、前記距離画像を正解データとする教師データを用いて機械学習モデルの学習を行う、ことをコンピュータに実行させる。

[0084] また、上記の学習方法を実現する学習装置を構成する構成要素の一部は、前記コンピュータプログラム又は前記デジタル信号を、電気通信回線、無線又は有線通信回線、インターネットを代表とするネットワーク、データ放送等を経由して伝送するものとしてもよい。

[0085] (5) 本開示は、上記に示す方法であるとしてもよい。また、これらの方法をコンピュータにより実現するコンピュータプログラムであるとしてもよいし、前記コンピュータプログラムからなるデジタル信号であるとしてもよい。

- [0086] (6) また、本開示は、マイクロプロセッサとメモリを備えたコンピュータシステムであって、前記メモリは、上記コンピュータプログラムを記憶しており、前記マイクロプロセッサは、前記コンピュータプログラムにしたがって動作するとしてもよい。
- [0087] (7) また、前記プログラム又は前記デジタル信号を前記記録媒体に記録して移送することにより、又は前記プログラム又は前記デジタル信号を、前記ネットワーク等を経由して移送することにより、独立した他のコンピュータシステムにより実施するとしてもよい。
- [0088] (8) また、上記の学習方法を実現する学習装置を構成する構成要素の一部をクラウドまたはサーバ装置でおこなってもよい。
- [0089] (9) 上記実施の形態及び上記変形例をそれぞれ組み合わせるとしてもよい。

産業上の利用可能性

- [0090] 本開示は、様々な単眼Depth推定アルゴリズムの学習に適用可能な教師あり対照学習の学習方法、及び、プログラムなどに利用できる。

符号の説明

- [0091] 10 RGBカメラ
20 測距センサ
100 学習装置
110 通信部
120 情報処理部
121 RGB画像取得部
122 距離画像取得部
123 データ拡張処理部
124 埋め込み画像生成部
125 学習部
130 記憶部
131 画像データベース

- 1 3 2 学習パラメータデータベース
- 1 3 3 機械学習モデル
- 2 0 0 学習システム
- 3 0 0 推定装置
- 4 0 0 推定システム

請求の範囲

- [請求項1] 画像と、前記画像に対応する距離画像とを取得し、
取得された前記距離画像の一部の領域を切り取り、
切り取られた前記一部の領域の距離画像を、前記画像において前記一部の領域と位置及び大きさが対応する所定の領域に貼り付けることにより、埋め込み画像を生成し、
生成された前記埋め込み画像を入力データとし、前記距離画像を正解データとする教師データを用いて機械学習モデルの学習を行う、
学習方法。
- [請求項2] 前記所定の領域は、前記画像の25%以上75%以下の面積である、
、
請求項1に記載の学習方法。
- [請求項3] 前記一部の領域は、前記画像に映る物体の輪郭を示すエッジ部分を含む、
請求項2に記載の学習方法。
- [請求項4] 前記画像と前記距離画像との関係性を前記機械学習モデルに学習させる、
請求項1～3のいずれか1項に記載の学習方法。
- [請求項5] 前記機械学習モデルは、エンコーダネットワークモデルと、前記エンコーダネットワークモデルから出力される低次元の特徴表現を前記画像と同じ大きさの出力画像にアップサンプリングする出力層とで構成される、
請求項1～3のいずれか1項に記載の学習方法。
- [請求項6] 前記機械学習モデルは、エンコーダネットワークモデルと、デコーダネットワークモデルとで構成される、
請求項1～3のいずれか1項に記載の学習方法。
- [請求項7] 画像と、前記画像に対応する距離画像とを取得し、取得された前記距離画像のうちの一部の領域を切り取り、切り取られた前記一部の領

域の距離画像を、前記画像において前記一部の領域と位置及び大きさが対応する所定の領域に貼り付けることにより、埋め込み画像を生成する埋め込み画像生成部と、

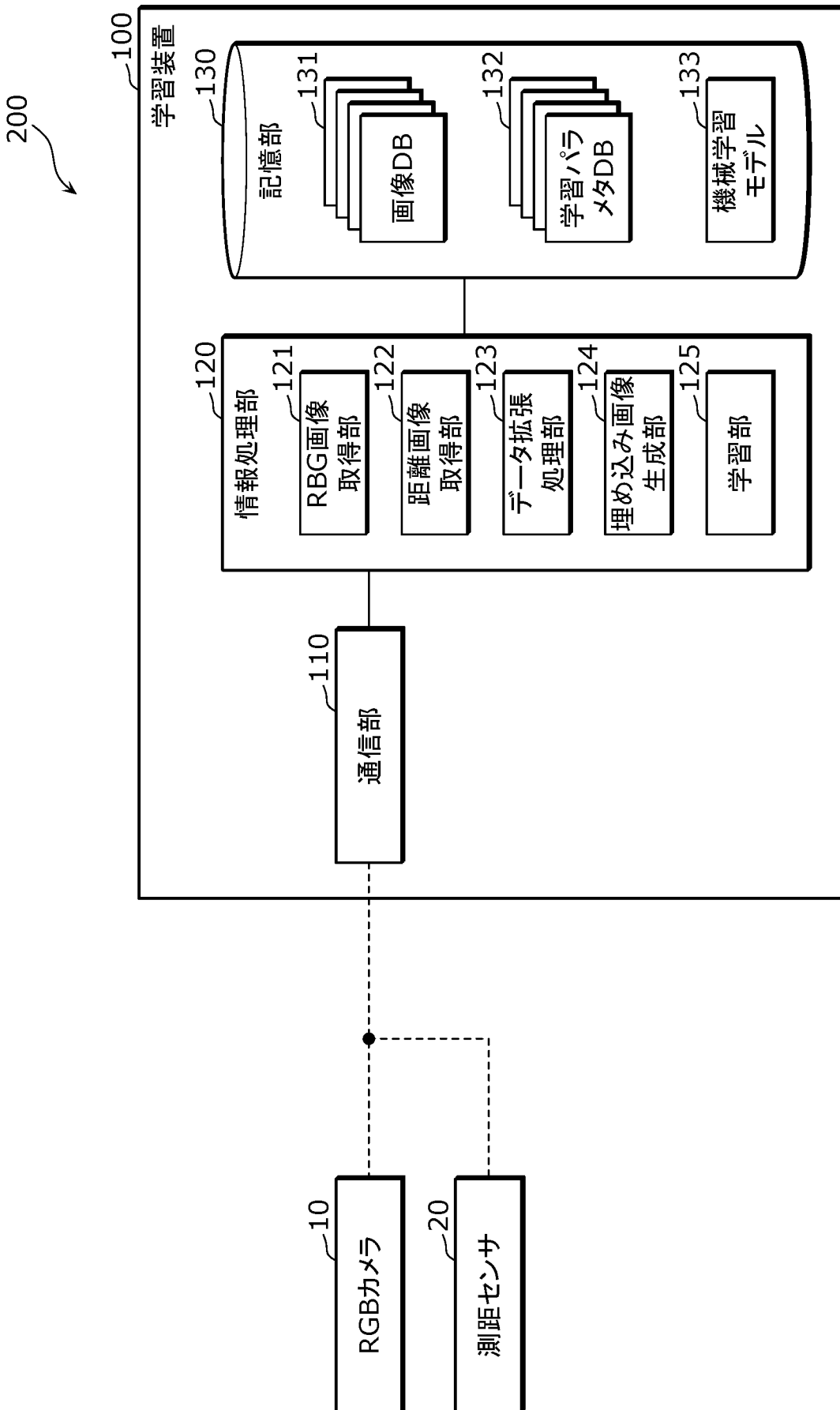
生成された前記埋め込み画像を入力データとし、前記距離画像を正解データとする教師データを用いて機械学習モデルの学習を行う学習部と、

を備える、

学習装置。

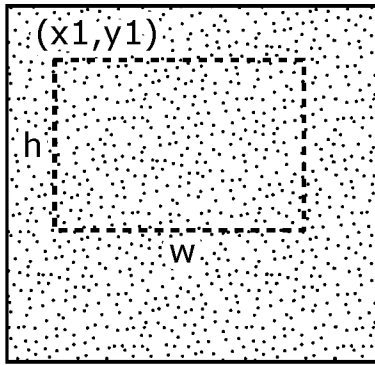
[請求項8] 請求項1に記載の学習方法をコンピュータに実行させるためのプログラム。

[図1]

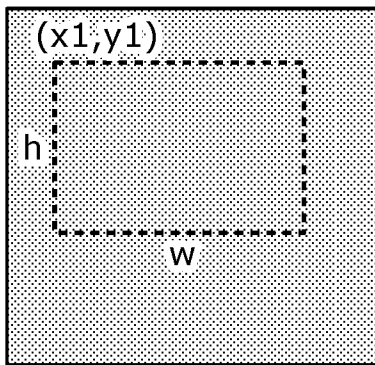


[図2]

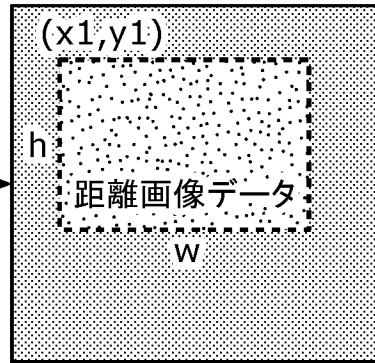
(a) 距離画像データ(教師データ)



(b) RGB画像データ

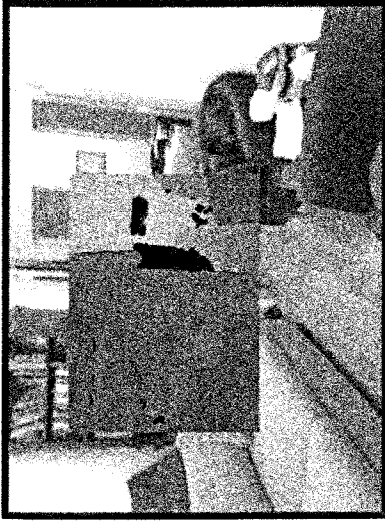


(c) 埋め込み画像

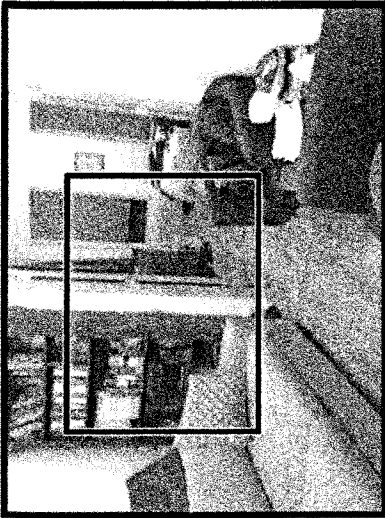


[図3]

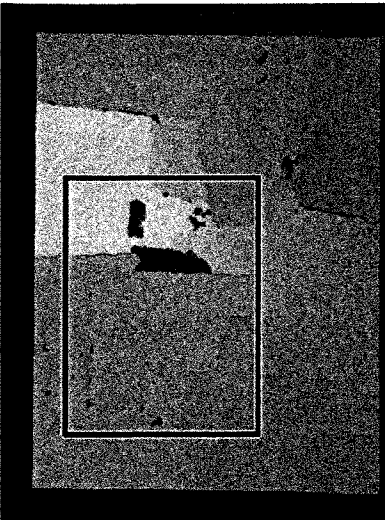
(c)埋め込み画像



(b)RGB画像データ



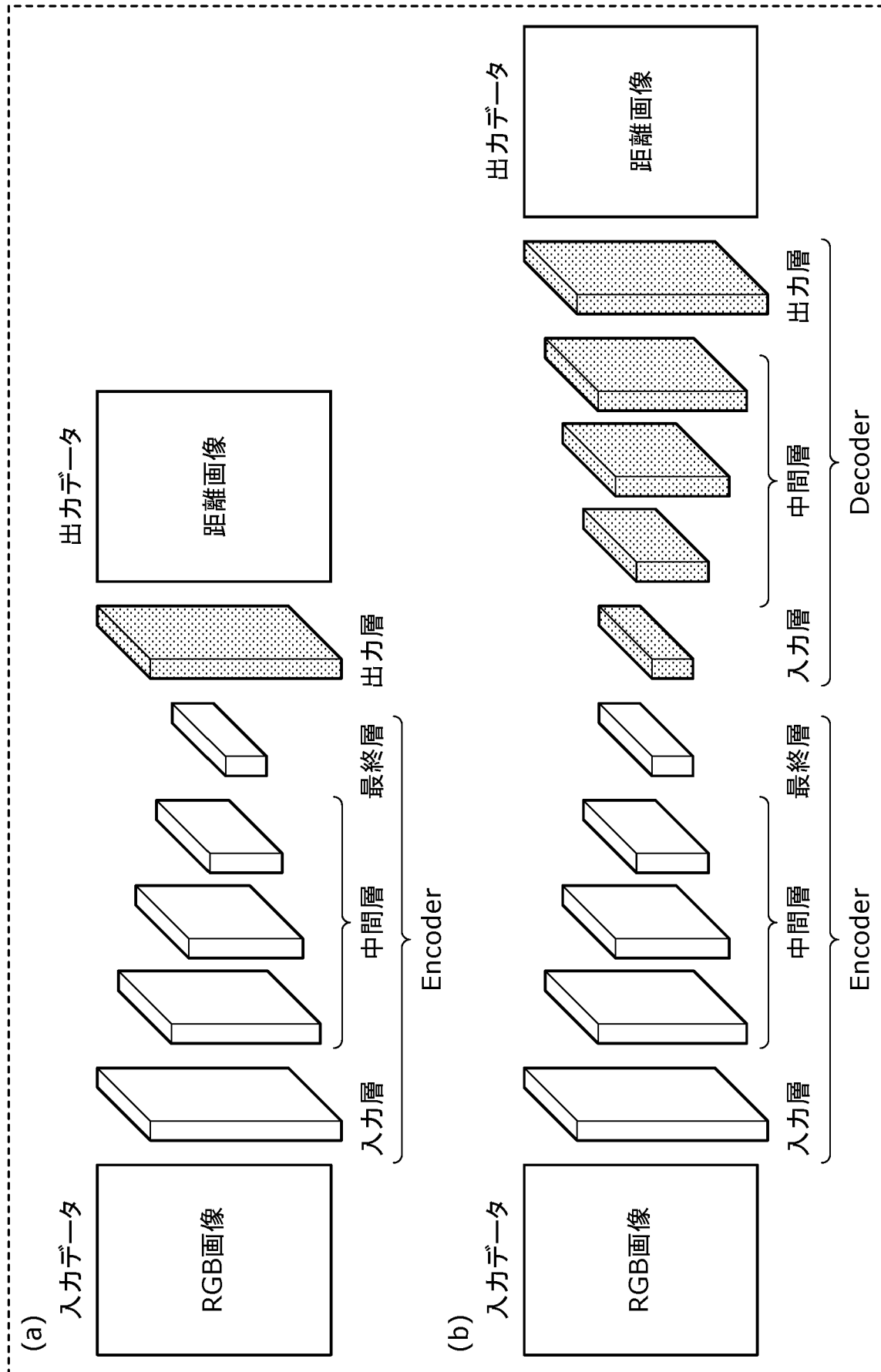
(a)距離画像データ



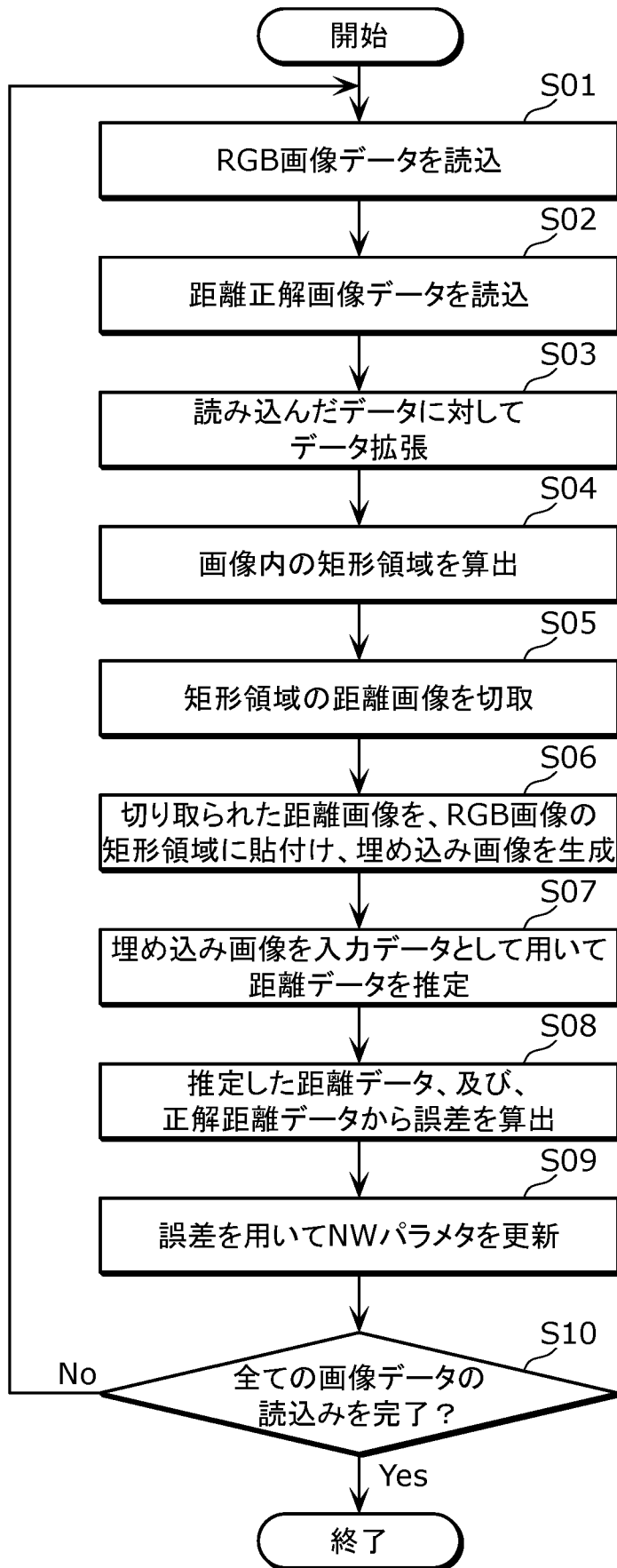
[図4]



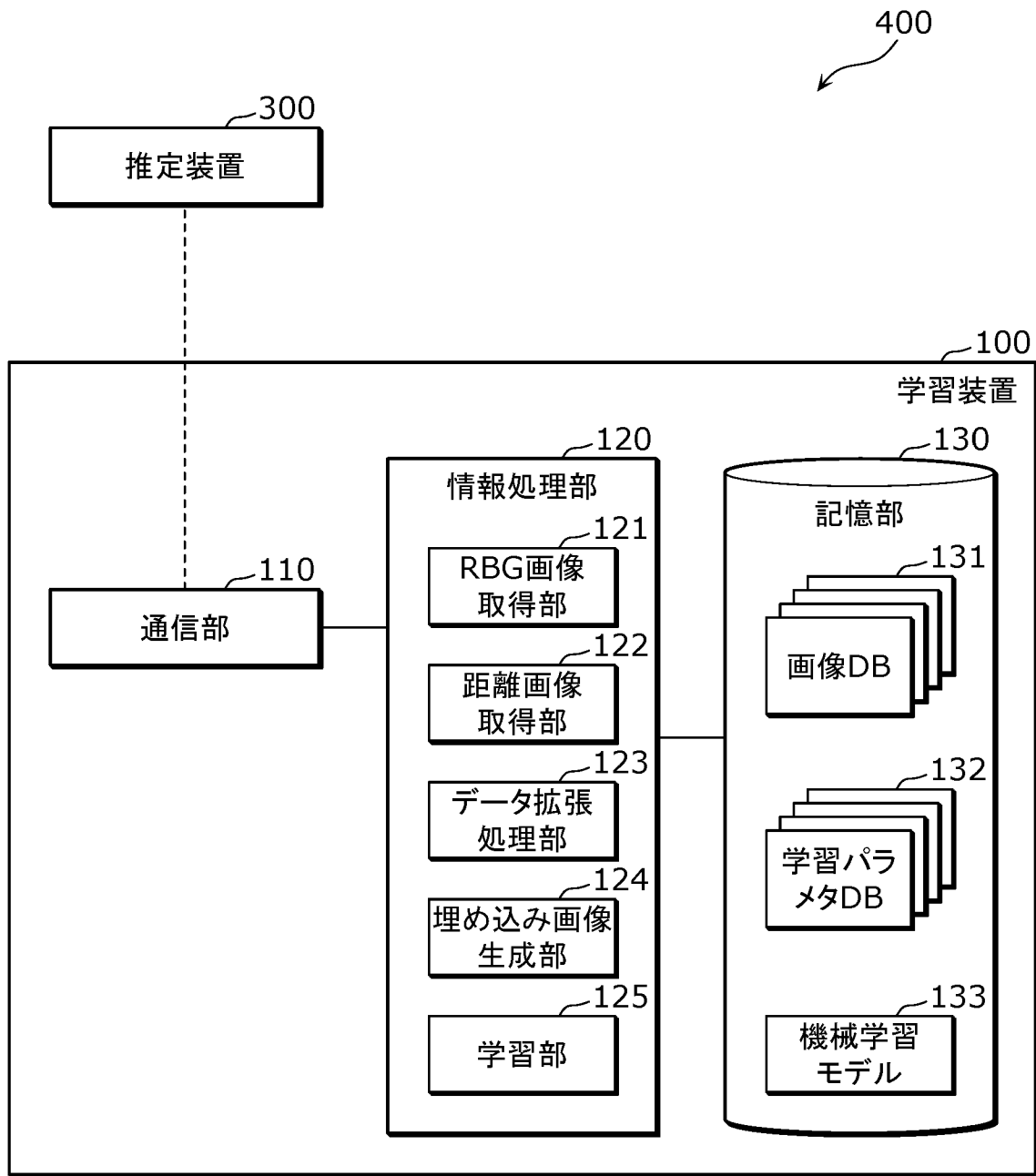
[図5]



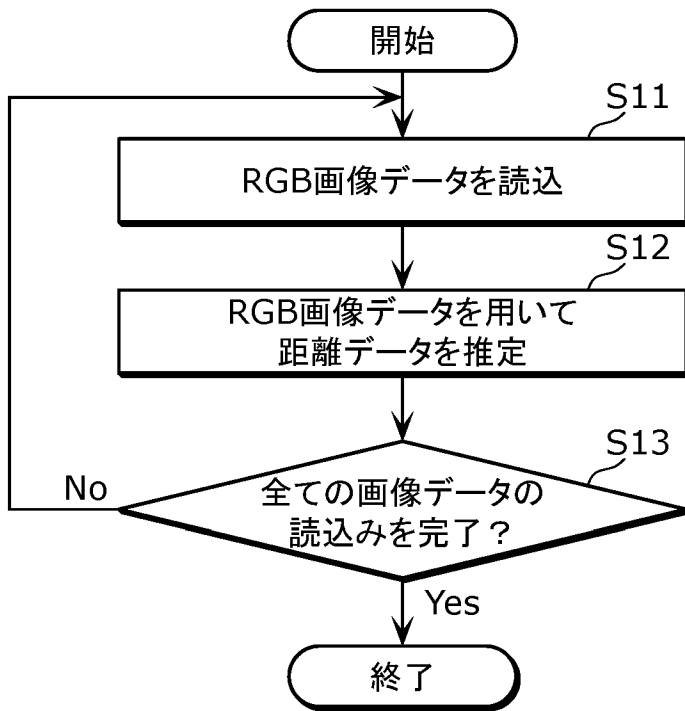
[図6]



[図7]



[図8]



[図9]

(Bts)

学習法	埋込率(%)	rms	abs_rel	log10	log_rms
従来手法	—	0.406	0.1122	0.048	0.145
本願手法	25	0.398	0.1083	0.047	0.141
	50	0.391	0.1077	0.046	0.14
	75	0.392	0.1074	0.047	0.14
	100	0.392	0.1127	0.047	0.142

[図10]

(lapDepth)

学習法	埋込率(%)	rms	abs_rel	log10	log_rms
従来手法	—	0.39	0.11	0.047	0.139
本願手法	25	0.38	0.106	0.045	0.135
	50	0.375	0.104	0.044	0.132
	75	0.379	0.106	0.045	0.135
	100	0.376	0.104	0.045	0.132

INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2022/019477

A. CLASSIFICATION OF SUBJECT MATTER		
<i>G06T 7/00</i> (2017.01)i; <i>G06T 7/50</i> (2017.01)i FI: G06T7/00 350B; G06T7/50		
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) G06T7/00; G06T7/50		
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.
A	JP 2019-125116 A (CANON KABUSHIKI KAISHA) 25 July 2019 (2019-07-25) paragraph [0028], fig. 3	1-8
A	JP 2020-154605 A (FUJI XEROX CO., LTD.) 24 September 2020 (2020-09-24) paragraphs [0065]-[0071], fig. 9, 10	1-8
<input type="checkbox"/> Further documents are listed in the continuation of Box C. <input checked="" type="checkbox"/> 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 13 July 2022		Date of mailing of the international search report 26 July 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/JP2022/019477

Patent document cited in search report	Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
JP 2019-125116	A 25 July 2019	(Family: none)	
JP 2020-154605	A 24 September 2020	(Family: none)	

A. 発明の属する分野の分類（国際特許分類（IPC）） G06T 7/00(2017.01)i; G06T 7/50(2017.01)i FI: G06T7/00 350B; G06T7/50		
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）） G06T7/00; G06T7/50 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
A	JP 2019-125116 A (キヤノン株式会社) 25.07.2019 (2019-07-25) 段落[0028], 図3	1-8
A	JP 2020-154605 A (富士ゼロックス株式会社) 24.09.2020 (2020-09-24) 段落[0065]-段落[0071], 図9-10	1-8
<input type="checkbox"/> C欄の続きにも文献が列挙されている。 <input checked="" type="checkbox"/> 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、一般的な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	13.07.2022	国際調査報告の発送日 26.07.2022
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） 宮島 潤 5H 8420 電話番号 03-3581-1101 内線 3531	

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/JP2022/019477

引用文献	公表日	パテントファミリー文献	公表日
JP 2019-125116 A	25.07.2019	(ファミリーなし)	
JP 2020-154605 A	24.09.2020	(ファミリーなし)	