

(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6月2日(02.06.2022)



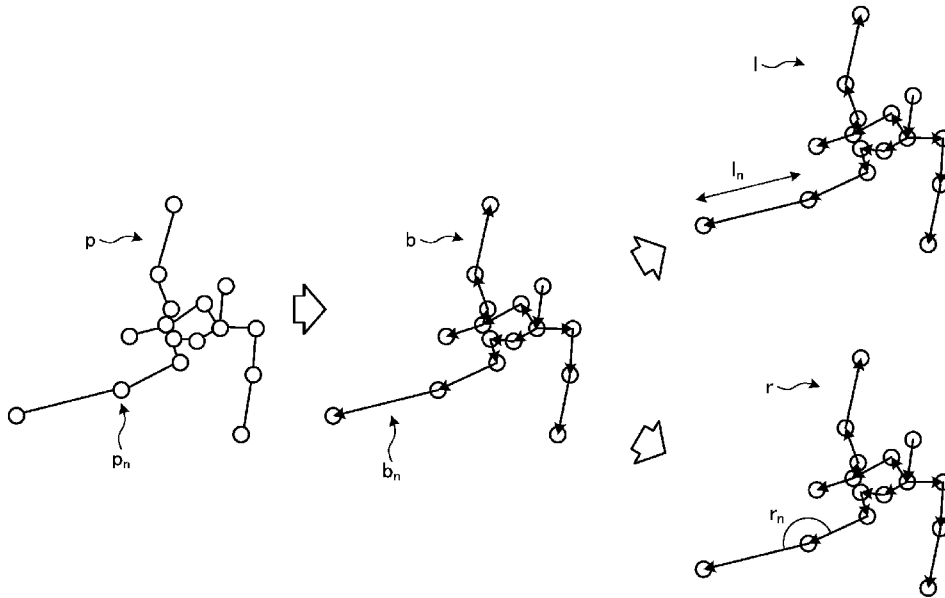
(10) 国際公開番号

WO 2022/113349 A1

- (51) 国際特許分類:
G06T 7/00 (2017.01)
- (21) 国際出願番号: PCT/JP2020/044524
- (22) 国際出願日: 2020年11月30日(30.11.2020)
- (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: 富士通株式会社 (**FUJITSU LIMITED**)
[JP/JP]; 〒2118588 神奈川県川崎市中原区上小田中4丁目1番1号 Kanagawa (JP).
- (72) 発明者: 鈴木 達也 (**SUZUKI, Tatsuya**);
〒2118588 神奈川県川崎市中原区上小田中4丁目1番1号 富士通株式会社内 Kanagawa (JP).
- (74) 代理人: 特許業務法人酒井国際特許事務所 (**SAKAI INTERNATIONAL PATENT OFFICE**); 〒1000013 東京都千代田区霞が関3丁目8番1号 虎の門三井ビルディング 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, 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, 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,

(54) **Title:** SKELETON RECOGNITION DEVICE, LEARNING METHOD, AND LEARNING PROGRAM

(54) 発明の名称: 骨格認識装置、学習方法および学習プログラム



(57) **Abstract:** This skeleton recognition device inputs training data to a learning model and calculates an output by forward propagation. The skeleton recognition device calculates a value of a loss function on the basis of: the difference in bone length based on skeleton information that is a right answer value of the training data and an output result from a recognition processing unit; and the difference in angle between bones based on the skeleton information and an output result from the recognition processing unit.



WO 2022/113349 A1

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: 骨格認識装置は、教師データを学習モデルに入力し順伝播による出力を計算する。骨格認識装置は、教師データの正解値である骨格情報と認識処理部の出力結果とに基づく骨の長さの差、および、骨格情報と前記認識処理部の出力結果とに基づく骨同士の角度の差に基づいて、損失関数の値を算出する。

明 細 書

発明の名称：骨格認識装置、学習方法および学習プログラム

技術分野

[0001] 本発明は、骨格認識装置等に関する。

背景技術

[0002] 3次元の人の動きを検出する骨格認識に関しては、複数台の3Dレーザセンサから3D骨格座標を±1cmの精度で抽出する3Dセンシング技術が確立され、様々なスポーツへの適用が期待されている。3Dレーザセンサを用いた方式を「レーザ方式」と表記する。

[0003] 上記のレーザ方式では、レーザを1秒間に約200万回照射し、レーザの走行時間 (Time of Flight: ToF) を基に、対象となる人を含めて、各照射点の深度データを求める。レーザ方式では、高精度な深度データを取得できるが、レーザスキャンやToF測定などの複雑な構成、処理を行うため、ハードウェアが複雑で高価になるという欠点がある。

[0004] レーザ方式に対して、CMOS (Complementary Metal Oxide Semiconductor) イメージャによって各ピクセルのRGBデータを取得する画像方式では、安価なRGBカメラを用いて3D骨格認識を行うことが可能となる。

[0005] たとえば、画像方式によって骨格認識を行う従来技術として、従来技術1がある。図11は、従来技術1を説明するための図である。従来技術1では、各入力画像1に2D backbone処理1aをかけることで、各関節特徴を表す2D features 2を取得する。従来技術1は、各2D features 2をカメラパラメータに従って3Dcubeに投影することで、aggregated volumes 3を取得する。

[0006] 従来技術1は、aggregated volumes 3を、V2V (ニューラルネットワーク、P3) 4に入力することで、各関節の尤度を表すprocessed volumes 5を取得する。processed volumes 5は、各関節の3Dでの尤度を表すheatmapに対応する。従来技術1は、processed volumes 5に対して、soft-argmax 6を実行することで、3D骨格情報7を取得する。

- [0007] 従来技術1で用いるV2V4の学習の一例について説明する。V2V4は、主に3Dのカーネルで構成されており、カーネルの重みとバイアスとを誤差逆伝播で学習する。カーネルの重みとバイアスとをまとめて、適宜、「パラメータ」と表記する。従来技術1では、soft-argmax6から出力される3D骨格情報および正解値から関節位置誤差の平均値 L_p をLoss値として算出し、このLoss値が最小化するように、パラメータを学習（調整）する。
- [0008] ここで、従来技術1では、オクルージョンの発生によって一部関節の誤差が大きくなると、V2V4のパラメータを適切に調整することが難しくなる。たとえば、一部関節の誤差が大きくても、他の関節の誤差が小さいと、平均化によって、一部関節の誤差が、Loss関数の出力に大きく影響を与えない。すなわち、一部関節の誤差が改善されるように、V2V4のパラメータが調整されなくなる。
- [0009] 図12は、一部関節の誤差が大きい骨格の例を示す図である。図12において3D骨格情報10aは、正解値に対応する3D骨格情報である。3D骨格情報10bは、soft-argmax6から出力される骨格認識結果に対応する。図12に示す例では、左足首関節の誤差が大きいが、他の関節の誤差が小さいため、平均するとLoss値に大きく影響を与えないため、左足首位置を改善するような学習効果を期待することができない。
- [0010] 上記の問題に対して、従来技術2がある。従来技術2では、1段階目において、距離画像を基にして、関節位置を推定しておき、2段階目において、推定した関節位置に対して、人体モデルを基に整合性を評価し、最も良い関節位置の組み合わせとなるように、推定した関節位置を補正する。従来技術2では、上記のように2段階で処理を行うことで、関節位置の推定精度を向上させている。

先行技術文献

特許文献

- [0011] 特許文献1：特開2015-167008公報
特許文献2：特開2004-13474号公報

非特許文献

[0012] 非特許文献1: K. Iskakov, et al., “Learnable Triangulation of Human Pose”, ICCV 2019, oral

非特許文献2: G. Moon, et al., “V2V-PoseNet: Voxel-to-Voxel Prediction Network for Accurate 3D Hand and Human Pose Estimation from a Single Depth Map”, CVPR 2018

発明の概要

発明が解決しようとする課題

[0013] 上述した従来技術2では、関節位置の推定精度を向上させているものの、補正した関節位置を出力するまでに、2段階の処理を実行することになり、骨格認識に処理時間を要してしまうという問題がある。

[0014] 1つの側面では、本発明は、処理時間をかけずに骨格認識結果を行うことができる骨格認識装置、学習方法および学習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

[0015] 第1の案では、骨格認識装置は、認識処理部と、損失関数算出部とを有する。認識処理部は、教師データを学習モデルに入力し順伝播による出力を計算する。損失関数算出部は、教師データの正解値である骨格情報と認識処理部の出力結果とに基づく骨の長さの差、および、骨格情報と前記認識処理部の出力結果とに基づく骨同士の角度の差に基づいて、損失関数の値を算出する。

発明の効果

[0016] 処理時間をかけずに骨格認識結果を行うことができる。

図面の簡単な説明

[0017] [図1]図1は、本実施例に係るシステムの一例を示す図である。

[図2]図2は、本実施例に係る骨格認識装置の構成を示す機能ブロック図である。

[図3]図3は、教師データのデータ構造の一例を示す図である。

[図4]図4は、人体モデルのデータ構造の一例を示す図である。

[図5]図5は、学習フェーズの処理を説明するための図である。

[図6]図6は、骨格認識装置の前処理フェーズの処理手順を示すフローチャートである。

[図7]図7は、骨格認識装置の学習フェーズの処理手順を示すフローチャートである。

[図8]図8は、骨格認識装置の推論フェーズの処理手順を示すフローチャートである。

[図9]図9は、本実施例の効果を説明するための図である。

[図10]図10は、実施例の骨格認識装置と同様の機能を実現するコンピュータのハードウェア構成の一例を示す図である。

[図11]図11は、従来技術1を説明するための図である。

[図12]図12は、一部関節の誤差が大きい骨格の例を示す図である。

発明を実施するための形態

[0018] 以下に、本願の開示する骨格認識装置、学習方法および学習プログラムの実施例を図面に基づいて詳細に説明する。なお、この実施例によりこの発明が限定されるものではない。

実施例

[0019] 図1は、本実施例に係るシステムの一例を示す図である。図1に示すように、このシステムは、カメラ20a、20b、20cと、骨格認識装置100とを有する。カメラ20a～20cと、骨格認識装置100とは、無線又は有線によって相互に接続される。

[0020] 図1のシステムでは、一例として、カメラ20a～20cを示すが、他のカメラを含んでもよい。以下の説明では、カメラ20a～20cをまとめて、カメラ20と表記する。

[0021] カメラ20は、人物15の画像を撮影する装置である。カメラ20は、撮影した画像のデータを、骨格認識装置100に送信する。以下の説明では、

人物15の画像のデータを、画像データと表記する。後述するように、画像データは、骨格認識装置100の推論フェーズにて使用される。

[0022] 骨格認識装置100は、教師データを用いた「前処理フェーズ」、「学習フェーズ」の処理を実行することで、骨格認識を行うための学習モデルの学習を実行する。また、骨格認識装置100は、「推論フェーズ」において、カメラ20から入力される画像データと、学習済みの学習モデルとを基にして、人物15の骨格認識結果を出力する。

[0023] 次に、本実施例に係る骨格認識装置の構成の一例について説明する。図2は、本実施例に係る骨格認識装置の構成を示す機能ブロック図である。図2に示すように、この骨格認識装置100は、通信部110、入力部120、表示部130、記憶部140、制御部150を有する。

[0024] 通信部110は、有線又は無線で外部装置等に接続され、カメラ20、外部装置等との間で情報の送受信を行う。たとえば、通信部110は、NIC (Network Interface Card) 等によって実現される。制御部150は、通信部110を介して、カメラ20から画像データを取得する。通信部110は、図示しないネットワークに接続されていてもよい。

[0025] 入力部120は、各種の情報を、骨格認識装置100に入力する入力装置である。入力部120は、キーボードやマウス、タッチパネル等に対応する。

[0026] 表示部130は、制御部150から出力される情報を表示する表示装置である。表示部130は、液晶ディスプレイ、有機EL (Electro Luminescence) ディスプレイ、タッチパネル等に対応する。

[0027] 記憶部140は、教師データ141、人体モデル142、学習モデル143、画像データテーブル144を有する。記憶部140は、たとえば、RAM (Random Access Memory)、フラッシュメモリ (Flash Memory) 等の半導体メモリ素子、または、ハードディスク、光ディスク等の記憶装置によって実現される。

[0028] 教師データ141は、学習モデル143の学習に用いられるデータである

。図3は、教師データのデータ構造の一例を示す図である。図3に示すように、教師データ141は、入力データと正解値との組を複数有する。入力データは、予め準備された複数の画像データから生成されるデータであり、たとえば「aggregated volumes」等に対応する。正解値は、入力データに対応する骨格認識結果の正解値であり、たとえば、各関節の3次元座標、関節を結ぶ骨の長さ、骨同士の角度等の情報が含まれる。

[0029] 人体モデル142は、予め定義される人体モデルであり、複数の関節が定義され、各関節が骨で接続される。図4は、人体モデルのデータ構造の一例を示す図である。図4に示すように、この人体モデル142は、関節A0, A1, A2, A3, A4, A5, A6, A7, A8, A9, A10, A11, A12, A13, A14, A15, A16, A17, A18, A19, A20, A21を有する。また、人体モデル142には、各関節を接続する骨が20本定義される。

[0030] 学習モデル143は、V2V等のニューラルネットワーク (NN: Neural Network) に対応する。たとえば、学習モデル143は、3Dのカーネルで構成されており、カーネルにはパラメータ (重み、バイアス等) が設定される。後述する学習フェーズの処理によって、パラメータが調整される。

[0031] 画像データテーブル144は、「推論フェーズ」において、カメラ20から出力される画像データを登録するテーブルである。たとえば、画像データテーブル144は、カメラ20の識別情報と、カメラ20が撮影した画像データとを対応付けて保持する。

[0032] 図2の説明に戻る。制御部150は、認識処理部151と、損失関数算出部152と、学習部153と、画像データ処理部154とを有する。制御部150は、例えば、CPU (Central Processing Unit) やMPU (Micro Processing Unit) により実現される。また、制御部150は、例えばASIC (Application specific Integrated Circuit) やFPGA (Field Programmable gate Array) 等の集積回路により実行されてもよい。

[0033] 制御部150は、(1) 前処理フェーズの処理、(2) 学習フェーズの処

理、(3) 推論フェーズの処理を順に実行する。ここでは、説明の便宜上、まず、「学習フェーズの処理」を説明した後に、「前処理フェーズの処理」の説明を行い、最後に「推論フェーズの処理」について説明する。

[0034] 制御部150が実行する「学習フェーズの処理」について説明する。図5は、学習フェーズの処理を説明するための図である。制御部150が前処理フェーズを実行する場合には、認識処理部151、損失関数算出部152、学習部153が次の処理を実行する。

[0035] 学習フェーズにおける認識処理部151の処理について説明する。認識処理部151は、教師データ141に含まれる入力データを、学習モデル143に入力し、順伝播による出力（骨格認識結果 $p = (p_1, p_2, \dots, p_N, p_p)$ ）を計算する。「 p_1, p_2, \dots, p_{N_p} 」は、各関節を示し、3次元座標が設定される。図5に示す例では、骨格認識結果 p において、ある関節が、 p_n で示される。

[0036] より具体的には、認識処理部151は、入力データを、学習モデル143に入力することで、各関節の3Dでの尤度を示すheatmapを計算し、heatmapの重心を求めるsoft-argmaxを実行することで、骨格認識結果 p を計算する。

[0037] 認識処理部151は、骨格認識結果を、損失関数算出部152に出力する。認識処理部151は、教師データ141に含まれる各入力データについて、上記処理を繰り返し実行することで、各入力データに対応する骨格認識結果を計算し、損失関数算出部152に出力する。

[0038] 学習フェーズにおける損失関数算出部152の処理について説明する。損失関数算出部152は、骨の長さの誤差を L_{oss} 値として算出する処理、骨同士の角度の誤差を L_{oss} 値として算出する処理、 L_{oss} 値を統合する処理を実行する。

[0039] 損失関数算出部152が実行する「骨の長さの誤差を L_{oss} 値として算出する処理」について説明する。損失関数算出部152は、骨格認識結果 p の各関節を、人体モデル142の各関節と対応付けることで、骨ベクトル $b = (b_1, b_2, \dots, b_{N_p})$ を算出する。図5に示す例では、骨ベクトル

bにおいて、ある骨ベクトルが、 b_n で示される。

[0040] 損失関数算出部152は、骨ベクトルbから骨の長さ $l = (l_1, l_2, \dots, l_{N_p})$ を算出する。図5に示す例では、骨の長さlにおいて、ある骨の長さが、 l_n で示される。

[0041] 損失関数算出部152は、骨の長さlと、教師データ141の正解値（骨の長さの正解値）との差から、骨の長さのLoss値を、式(1)に基づいて算出する。骨の長さのLoss値を「 L_l 」と表記する。式(1)において、 N_l は、骨の数（たとえば、 $N_l = 20$ ）を示す。 l_i^{pred} は、骨ベクトルbから推定された骨の長さであって、i番目の骨の長さである。 l_i^{gt} は、正解値の骨の長さであって、i番目の骨の長さである。

[0042] [数1]

$$L_l = \sum_{i=1}^{N_l} \frac{(l_i^{pred} - l_i^{gt})^2}{N_l} \quad \dots(1)$$

[0043] 損失関数算出部152が実行する「骨同士の角度の誤差をLoss値として算出する処理」について説明する。損失関数算出部152は、骨ベクトルbから骨同士の角度 $r = (r_1, r_2, \dots, r_{N_p})$ を算出する。図5に示す例では、骨同士の角度lにおいて、ある骨同士の角度が、 l_n で示される。

[0044] 損失関数算出部152は、骨同士の角度rと、教師データ141の正解値（骨同士の角度の正解値）との差から、骨同士の角度のLoss値を、式(2)に基づいて算出する。骨同士の角度のLoss値を「 L_r 」と表記する。式(2)において、 N_r は、骨同士の角度の数を示す。 r_i^{pred} は、骨ベクトルbから推定された骨同士の角度であって、i番目の骨同士の角度である。 r_i^{gt} は、正解値の骨同士の角度であって、i番目の骨同士の角度である。

[0045] [数2]

$$L_r = \sum_{i=1}^{N_r} \frac{(r_i^{pred} - r_i^{gt})^2}{N_r} \quad \dots(2)$$

[0046] 損失関数算出部152が実行する「Loss値を統合する処理」について

説明する。損失関数算出部 152 は、骨格認識結果 p と、教師データ 141 の正解値（各関節の 3 次元座標）とを基にして、関節座標 L_{oss} 値を算出する。関節座標 L_{oss} 値を「 L_p 」と表記する。式を省略するが、 L_p は、各関節の 3 次元座標の誤差を平均化した値となる。

[0047] 損失関数算出部 152 は、式 (3) を基にして、 L_{oss} 値を統合する。統合した L_{oss} 値を「 L 」と表記する。損失関数算出部 152 は、 L_{oss} 値 L を、学習部 153 に出力する。式 (3) に含まれる w_l 、 w_r は、 L_l 、 L_r がオーバーフィッティングしないように、「前処理フェーズ」で予め算出される重みである。

[0048] [数3]

$$L = L_p + w_l L_l + w_r L_r \quad \dots (3)$$

[0049] 損失関数算出部 152 は、認識処理部 151 から、骨格認識結果 p を取得する度に、上記処理を繰り返し実行する。

[0050] ところで、損失関数算出部 152 は、骨の長さ、骨同士の角度を計算する場合に、Kinematic Chain Space等の技術を用いてもよい。たとえば、Kinematic Chain Spaceに関しては「B. Wandt, et al., “A kinematic chain space for monocular motion capture”, (ECCV Work-shops, 2018)」に記載されている。以下の説明では、Kinematic Chain Spaceを「KCS」と表記する。KCSでは、骨ベクトル b をまとめた行列 B を算出する。行列 B を式 (4) に示す。

[0051] [数4]

$$B = (b_1, b_2, \dots, b_b) \quad \dots (4)$$

[0052] 損失関数算出部 152 は、行列 B の転置行列 B^T と、 B との積を示す特徴 Ψ を算出する。特徴 Ψ は、式 (5) によって示される。特徴 Ψ の対角成分が骨の長さに対応し、非対角成分が、骨同士の角度に相当する。

[0053]

[数5]

$$\Psi = B^T B = \begin{pmatrix} l_1^2 & \cdot & \cdot & \cdot \\ \cdot & l_2^2 & \cdot & \cdot \\ \cdot & \cdot & \ddots & \cdot \\ \cdot & \cdot & \cdot & l_b^2 \end{pmatrix} \quad \dots (5)$$

[0054] 損失関数算出部152は、特徴 Ψ と、教師データ141の正解値(GT)との差の平均値 L_{KCS} を式(6)によって算出する。式(6)において、 Ψ_{ij}^{pred} は、式(6)の*i*行*j*列の要素を示す、 Ψ_{ij}^{gt} は、 Ψ_{ij}^{pred} に対応する正解値の要素である。bは、行列Bの要素の数に対応する。

[0055] [数6]

$$L_{kcs} = \sum_{i=1}^b \sum_{j=1}^b \frac{(\Psi_{ij}^{pred} - \Psi_{ij}^{gt})^2}{b^2} \quad \dots (6)$$

[0056] 上記の説明では、骨の長さ、骨同士の角度を計算する場合に、Kinematic Chain Space等の技術を用いる場合について説明したが、これに限定されるものではない。損失関数算出部152は、骨の長さの計算を、ユークリッド距離を用いて計算してもよい。損失関数算出部152は、骨同士の角度の計算を、cos類似度を用いて計算してもよい。

[0057] 学習フェーズにおける学習部153の処理について説明する。学習部153は、Loss値Lが最小化するように、学習モデル143のパラメータを、誤差逆伝播によって学習(調整)する処理を繰り返し実行する。学習部153は、Loss値Lが収束した場合に、学習モデル143のパラメータの調整を終了する。

[0058] 学習フェーズでは、式(3)に示すように、人体を考慮した骨の長さの誤差に関する「 L_l 」、骨同士の角度の誤差に関する「 L_r 」を用いて、Loss値Lが算出されているため、一部関節が大きくずれた場合のLoss値Lへの寄与度が高くなる。これによって、一部関節の誤差が大きい場合に、かかる誤差を改善する学習を実行することができる。

[0059] 次に、制御部150が実行する「前処理フェーズ」について説明する。前

処理フェーズでは、学習フェーズで利用される w_l 、 w_r を算出する。学習フェーズにおいて、式 (3) を用いて、Loss 値を統合する場合に、 L_l 、 L_r がオーバーフィッティングしないように、関節座標 Loss 値「 L_p 」と、骨の長さの Loss 値「 L_l 」との相対値から、「 w_l 」を算出する。関節座標 Loss 値「 L_p 」と、骨の長さの Loss 値「 L_r 」との相対値から、「 w_r 」を算出する。

[0060] 制御部 150 が、前処理フェーズを実行する場合には、認識処理部 151、損失関数算出部 152、学習部 153 が次の処理を実行する。

[0061] 前処理フェーズにおける認識処理部 151 の処理について説明する。前処理フェーズにおける認識処理部 151 の処理は、学習フェーズにおける認識処理部 151 の処理と同様である。すなわち、認識処理部 151 は、教師データ 141 に含まれる入力データを、学習モデル 143 に入力し、順伝播による出力（骨格認識結果 $p = (p_1, p_2, \dots, p_{N_p})$ ）を計算する。認識処理部 151 は、骨格認識結果 p を、損失関数算出部 152 に出力する。

[0062] 前処理フェーズにおける損失関数算出部 152 の処理について説明する。損失関数算出部 152 は、学習フェーズにおける損失関数算出部 152 の処理と同様にして、骨の長さの Loss 値「 L_l 」、骨同士の間角の Loss 値「 L_r 」、関節座標 Loss 値「 L_p 」をそれぞれ算出する。

[0063] ここで、損失関数算出部 152 は、関節座標 Loss 値「 L_p 」のみを、学習部 153 に出力する。損失関数算出部 152 は、認識処理部 151 から、骨格認識結果 p を取得する度に、上記処理を繰り返し実行する。

[0064] 損失関数算出部 152 は、学習部 153 から、関節座標 Loss 値「 L_p 」が収束した旨の情報を受け付けた場合に、関節座標 Loss 値「 L_p 」が収束した時点における「 L_l 」、 L_r 、 L_p のオーダーをそれぞれ算出する。「 L_l 」のオーダーを「 o_{L_l} 」とする。「 L_r 」のオーダーを「 o_{L_r} 」とする。「 L_p 」のオーダーを「 o_{L_p} 」とする。

[0065] たとえば、損失関数算出部 152 は、 $L_p = 4.0 \times 10^{-3}$ となる場合、「 $o_{L_p} = 10^{-3}$ 」とする。損失関数算出部 152 は、「 o_{L_l} 」、「 o_{L_p} 」につ

いても同様に算出する。

[0066] 損失関数算出部 152 は、式 (7) に基づいて、重み w_l を算出する。式 (8) に基づいて、重み w_r を算出する。係る重み w_l 、 w_r は、学習フェーズで利用される。損失関数算出部 152 が、上記処理によって、重み w_l 、 w_r を決定することで、常に、 L_p のオーダーを、 L_l 、 L_r が上回ることはないので、 L_l 、 L_r がオーバーフィッティングしないようにしつつ、 L_l 、 L_r が学習に適切な影響を与えることができる。

[0067] [数7]

$$w_l = \frac{O_{L_p}}{O_{L_l}} \quad \dots (7)$$

[数8]

$$w_r = \frac{O_{L_p}}{O_{L_r}} \quad \dots (8)$$

[0068] 前処理フェーズにおける学習部 153 の処理について説明する。学習部 153 は、Loss 値 L_p が最小化するように、学習モデル 143 のパラメータを、誤差逆伝播によって学習 (調整) する処理を繰り返し実行する。学習部 153 は、Loss 値 L_p が収束した場合に、学習モデル 143 のパラメータの調整を終了する。

[0069] 学習部 153 は、Loss 値 L_p が収束した場合に、Loss 値 L_p が収束した旨を損失関数算出部 152 に出力し、学習を終了する。

[0070] 次に、制御部 150 が実行する「推論フェーズ」について説明する。制御部 150 が前処理フェーズを実行する場合には、認識処理部 151、画像データ処理部 154 が次の処理を実行する。

[0071] 推論フェーズにおける認識処理部 151 の処理について説明する。認識処理部 151 は、画像データ処理部 154 から入力データを取得する。認識処理部 151 は、取得した入力データを、学習済みの学習モデル 143 に入力し、順伝播による出力 (人物 15 の骨格認識結果) を計算する。認識処理部 151 は、骨格認識結果を、表示部 130 に出力して表示してもよいし、骨

格認識結果を用いて各種の処理を実行する外部装置に、骨格認識結果を出力してもよい。

[0072] 認識処理部151は、画像データ処理部154から入力データを取得する度に、上記処理を繰り返し実行する。

[0073] 推論フェーズにおける画像データ処理部154の処理について説明する。画像データ処理部154は、カメラ20から画像データを取得し、取得した画像データを、画像データテーブル144に格納する。画像データ処理部154は、画像データに対して、2D backbone処理をかけることで、各関節特徴を表す2D featuresを特定する。画像データ処理部154は、各2D featuresをカメラパラメータに従って3Dcubeに投影することで、aggregated volumesを計算し、計算したaggregated volumesを入力データとして、認識処理部151に出力する。

[0074] なお、上述した画像データ処理部154が、各画像データを基にして、入力データを計算する処理は一例あり、他の技術を用いて、複数の画像データを統合して、入力データを計算してもよい。画像データ処理部154は、時間毎に、カメラ20から画像データを取得し、上記処理を繰り返し実行する。

[0075] 次に、本実施例に係る骨格認識装置100の処理手順の一例について説明する。図6は、骨格認識装置の前処理フェーズの処理手順を示すフローチャートである。図6に示すように、骨格認識装置100の認識処理部151は、教師データ141の入力データを学習モデル143に入力する（ステップS101）。認識処理部151は、学習モデル143の順伝播による骨格認識結果を算出する（ステップS102）。

[0076] 骨格認識装置100の損失関数算出部152は、関節座標の誤差をLoss値として算出する（ステップS103）。損失関数算出部152は、骨の長さの誤差、骨同士の角度の誤差を非Loss値として算出する（ステップS104）。

[0077] 骨格認識装置100の学習部153は、Loss値を最小化するように、

誤差逆伝播による学習モデル143のパラメータを調整する（ステップS105）。

[0078] 損失関数算出部152は、Loss値と非Loss値とを基にして、重み w_l 、 w_r を算出する（ステップS106）。損失関数算出部152は、重み w_l 、 w_r を出力する（ステップS107）。

[0079] 図7は、骨格認識装置の学習フェーズの処理手順を示すフローチャートである。図7に示すように、骨格認識装置100の認識処理部151は、教師データ141の入力データを学習モデル143に入力する（ステップS201）。認識処理部151は、学習モデル143の順伝播による骨格認識結果を算出する（ステップS202）。

[0080] 骨格認識装置100の損失関数算出部152は、関節座標の誤差、骨の長さの誤差、骨同士の角度の誤差をLoss値として算出する（ステップS203）。損失関数算出部152は、重み w_l 、 w_r を使用してLoss値を統合する（ステップS204）。

[0081] 骨格認識装置100の学習部153は、Loss値を最小化するように、誤差逆伝播による学習モデル143のパラメータを調整する（ステップS205）。学習部153は、学習済みの学習モデル143を出力する（ステップS206）。

[0082] 図8は、骨格認識装置の推論フェーズの処理手順を示すフローチャートである。図8に示すように、骨格認識装置100の画像データ処理部154は、カメラ20から画像データを取得する（ステップS301）。

[0083] 画像データ処理部154は、複数の画像データから入力データを生成する（ステップS302）。骨格認識装置100の認識処理部151は、入力データを学習済みの学習モデル143に入力する（ステップS303）。

[0084] 認識処理部151は、学習モデル143の順伝播による骨格認識結果を算出する（ステップS304）。認識処理部151は、骨格認識結果を出力する（ステップS305）。

[0085] 次に、本実施例に係る骨格認識装置100の効果について説明する。骨格

認識装置 100 は、学習モデル 143 に教師データの入力データを入力して骨格認識結果を算出し、骨認識結果と、教師データの正解値との比較により、骨の長さの誤差、骨同士の角度の誤差を算出し、損失関数の Loss 値を算出する。係る損失関数の Loss 値が最小化するように、学習モデル 143 のパラメータを調整することで、一部関節が大きくずれた場合の Loss 値への寄与度を高くすることができ、一部関節の誤差が大きい場合に、かかる誤差を改善する学習を行うことができる。また、かかる学習を行うことで、学習済みの学習モデル 143 に、入力データを入力するという一段階の処理によって、人物 15 の骨格認識結果を算出可能であり、2 段階で処理を実行する従来技術 2 と比較して、処理速度を向上させることができる。

[0086] より具体的には、骨格認識装置 100 は、教師データの正解値との比較により、関節座標の誤差、関節座標の誤差、関節座標の誤差をそれぞれ算出し、損失関数の Loss 値を算出することによって、学習モデル 143 のパラメータを精度よく学習することができる。

[0087] 骨格認識装置 100 は、式 (7) に基づいて、重み w_l を算出し、式 (8) に基づいて、重み w_r を算出する。係る重み w_l 、 w_r は、学習フェーズで利用される。重み w_l 、 w_r を決定することで、常に、関節座標の誤差 L_p のオーダーを、関節座標の誤差 L_l 、関節座標の誤差 L_r が上回ることがないので、 L_l 、 L_r がオーバーフィッティングしないようにしつつ、 L_l 、 L_r が学習に適切な影響を与えることができる。

[0088] 図 9 は、本実施例の効果の説明するための図である。図 9 において、骨格認識結果 30a は、従来技術 1 による骨格認識結果であり、左膝の角度が異常となっている。骨格認識結果 30b は、本実施例に係る骨格認識装置 100 の骨格認識結果であり、左膝の角度が正常化されている。

[0089] 次に、上記実施例に示した骨格認識装置 100 と同様の機能を実現するコンピュータのハードウェア構成の一例について説明する。図 10 は、実施例の骨格認識装置と同様の機能を実現するコンピュータのハードウェア構成の一例を示す図である。

- [0090] 図10に示すように、コンピュータ200は、各種演算処理を実行するCPU201と、ユーザからのデータの入力を受け付ける入力装置202と、ディスプレイ203とを有する。また、コンピュータ200は、記憶媒体からプログラム等を読み取る読み取り装置204と、有線または無線ネットワークを介して、カメラ20、他の外部装置等との間でデータの授受を行う通信装置205とを有する。また、コンピュータ200は、各種情報を一時記憶するRAM206と、ハードディスク装置207とを有する。そして、各装置201～207は、バス208に接続される。
- [0091] ハードディスク装置207は、認識処理プログラム207a、損失関数算出プログラム207b、学習プログラム207c、学習データ処理プログラム207dを有する。また、CPU201は、各プログラム207a～207dを読み出してRAM206に展開する。
- [0092] 認識処理プログラム207aは、認識処理プロセス206aとして機能する。損失関数算出プログラム207bは、損失関数算出プロセス206bとして機能する。学習プログラム207cは、学習プロセス206cとして機能する。画像データ処理プログラム207dは、画像データ処理プロセス206dとして機能する。
- [0093] 認識処理プロセス206aの処理は、認識処理部151の処理に対応する。損失関数算出プロセス206bの処理は、損失関数算出部152の処理に対応する。学習プロセス206cの処理は、学習部153の処理に対応する。画像データ処理プロセス206dの処理は、画像データ処理部154の処理に対応する。
- [0094] なお、各プログラム207a～207dについては、必ずしも最初からハードディスク装置207に記憶させておかなくても良い。例えば、コンピュータ200に挿入されるフレキシブルディスク(FD)、CD-ROM、DVD、光磁気ディスク、ICカードなどの「可搬用の物理媒体」に各プログラムを記憶させておく。そして、コンピュータ200が各プログラム207a～207dを読み出して実行するようにしてもよい。

符号の説明

[0095]	1 0 0	骨格認識装置
	1 1 0	通信部
	1 2 0	入力部
	1 3 0	表示部
	1 4 0	記憶部
	1 4 1	教師データ
	1 4 2	人体モデル
	1 4 3	学習モデル
	1 4 4	画像データテーブル
	1 5 0	制御部
	1 5 1	認識処理部
	1 5 2	損失関数算出部
	1 5 3	学習部
	1 5 4	画像データ処理部

請求の範囲

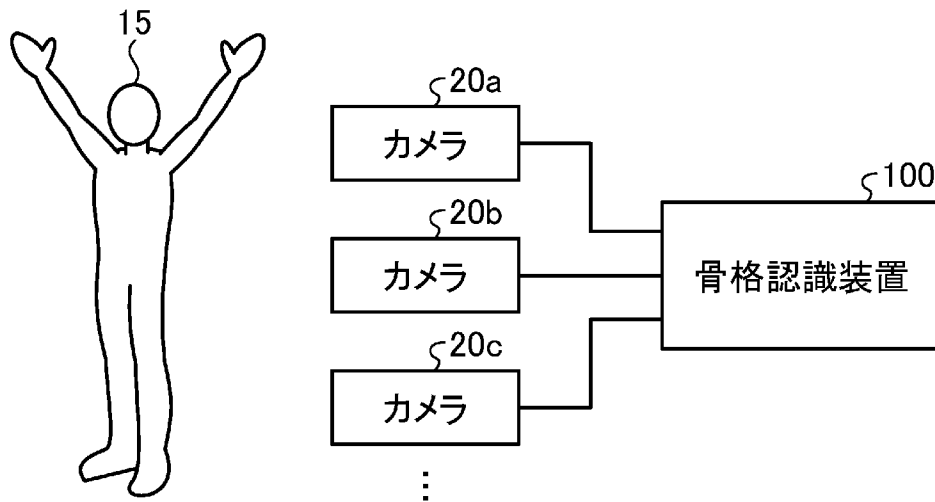
- [請求項1] 教師データを学習モデルに入力し順伝播による出力を計算する認識処理部と、
- 前記教師データの正解値である骨格情報と前記認識処理部の出力結果とに基づく骨の長さの差、および、前記骨格情報と前記認識処理部の出力結果とに基づく骨同士の角度の差に基づいて、損失関数の値を算出する損失関数算出部と
- を有することを特徴とする骨格認識装置。
- [請求項2] 前記損失関数算出部は、前記骨格情報と前記認識処理部の出力結果とに基づく関節位置の差を更に算出し、前記関節位置の差と、前記骨の長さの差と、前記骨同士の角度の差を基にして、前記損失関数の値を算出することを特徴とする請求項1に記載の骨格認識装置。
- [請求項3] 前記損失関数算出部は、前記関節位置の差および前記骨の長さの差を基にして第1の重みを算出し、前記関節位置の差および前記骨同士の角度の差を基にして第2の重みを算出し、前記関節位置の差の値と、前記骨の長さの差の値に前記第1の重みを乗算した値と、前記骨同士の角度の差の値に前記第2の重みを乗算した値とを統合することで、損失関数の値を算出することを特徴とする請求項2に記載の骨格認識装置。
- [請求項4] 前記損失関数の値を基にして、前記学習モデルのパラメータを調整する学習部を更に有することを特徴とする請求項3に記載の骨格認識装置。
- [請求項5] コンピュータが実行する学習方法であって、
- 教師データを学習モデルに入力し順伝播による出力結果を計算し、
- 前記教師データの正解値である骨格情報と前記出力結果とに基づく骨の長さの差、および、前記骨格情報と前記認識処理部の出力結果とに基づく骨同士の角度の差に基づいて、損失関数の値を算出する
- 処理を実行することを特徴とする学習方法。

- [請求項6] 前記損失関数の値を算出する処理は、前記骨格情報と前記認識処理部の出力結果とに基づく関節位置の差を更に算出し、前記関節位置の差と、前記骨の長さの差と、前記骨同士の角度の差を基にして、前記損失関数の値を算出することを特徴とする請求項5に記載の学習方法。
- [請求項7] 前記損失関数の値を算出する処理は、前記関節位置の差および前記骨の長さの差を基にして第1の重みを算出し、前記関節位置の差および前記骨同士の角度の差を基にして第2の重みを算出し、前記関節位置の差の値と、前記骨の長さの差の値に前記第1の重みを乗算した値と、前記骨同士の角度の差の値に前記第2の重みを乗算した値とを統合することで、損失関数の値を算出することを特徴とする請求項6に記載の学習方法。
- [請求項8] 前記損失関数の値を基にして、前記学習モデルのパラメータを調整する処理を更に実行することを特徴とする請求項7に記載の学習方法。
- [請求項9] コンピュータに、
教師データを学習モデルに入力し順伝播による出力結果を計算し、
前記教師データの正解値である骨格情報と前記出力結果とに基づく骨の長さの差、および、前記骨格情報と前記認識処理部の出力結果とに基づく骨同士の角度の差に基づいて、損失関数の値を算出する
処理を実行させることを特徴とする学習プログラム。
- [請求項10] 前記損失関数の値を算出する処理は、前記骨格情報と前記認識処理部の出力結果とに基づく関節位置の差を更に算出し、前記関節位置の差と、前記骨の長さの差と、前記骨同士の角度の差を基にして、前記損失関数の値を算出することを特徴とする請求項9に記載の学習プログラム。
- [請求項11] 前記損失関数の値を算出する処理は、前記関節位置の差および前記骨の長さの差を基にして第1の重みを算出し、前記関節位置の差およ

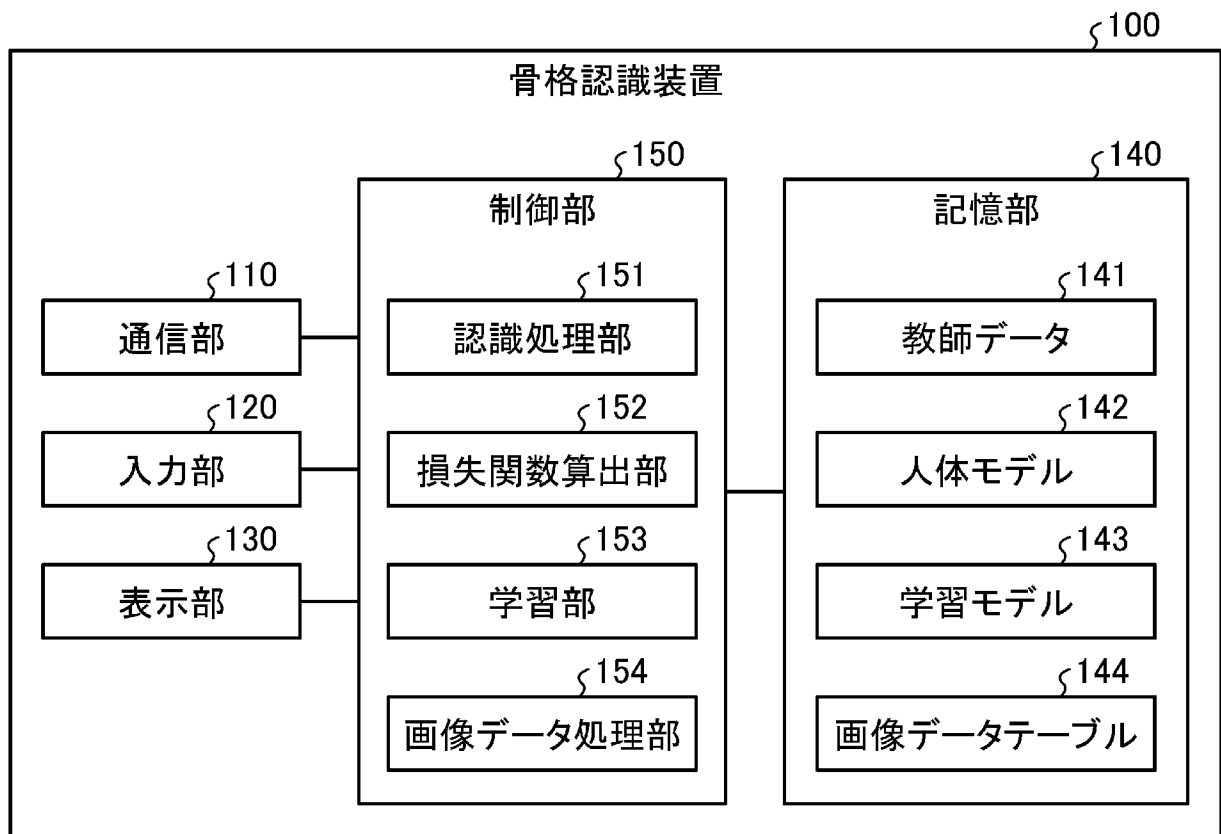
び前記骨同士の角度の差を基にして第2の重みを算出し、前記関節位置の差の値と、前記骨の長さの差の値に前記第1の重みを乗算した値と、前記骨同士の角度の差の値に前記第2の重みを乗算した値とを統合することで、損失関数の値を算出することを特徴とする請求項10に記載の学習プログラム。

[請求項12] 前記損失関数の値を基にして、前記学習モデルのパラメータを調整する処理を更に実行することを特徴とする請求項11に記載の学習プログラム。

[図1]



[図2]

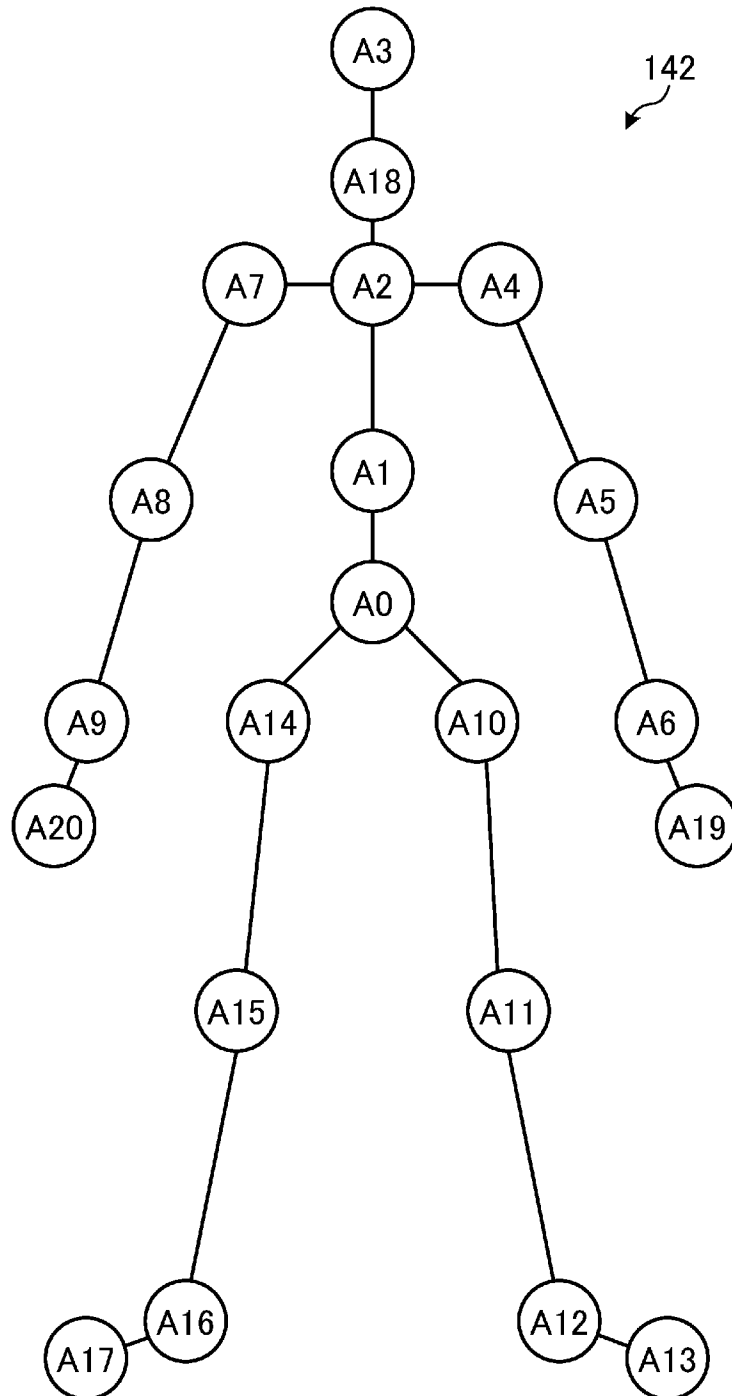


[図3]

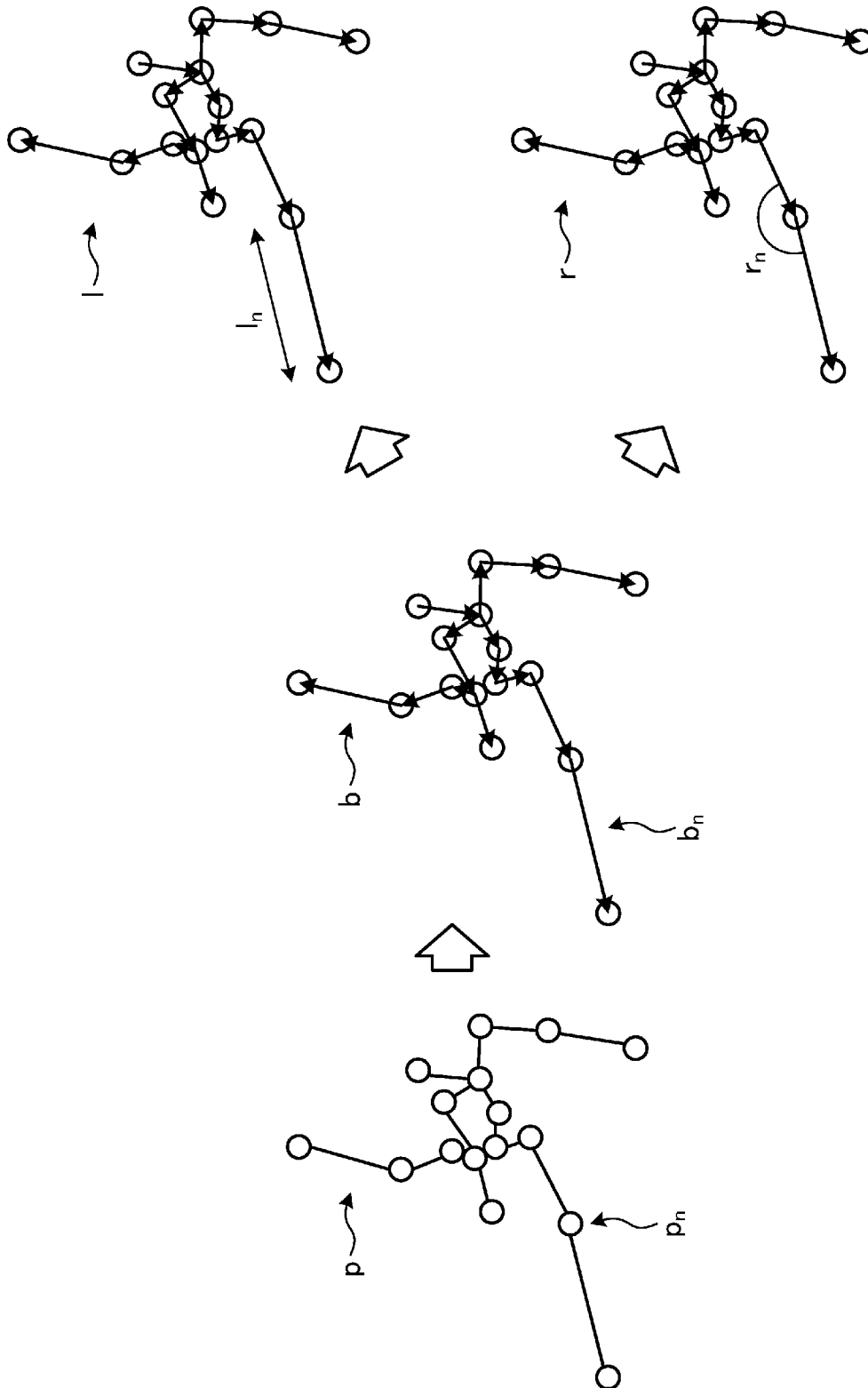
§141

入力データ	正解値
入力データ(1)	正解値(1)
入力データ(2)	正解値(2)
入力データ(3)	正解値(3)
...	...

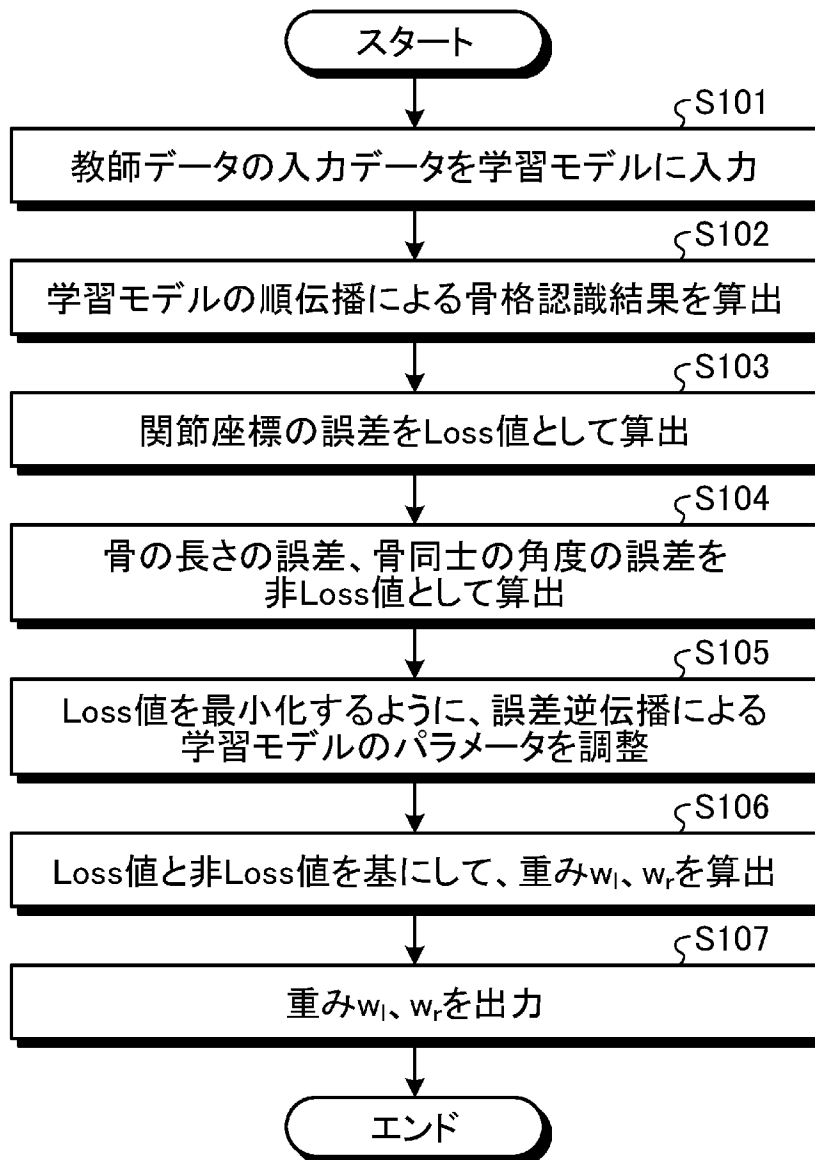
[図4]



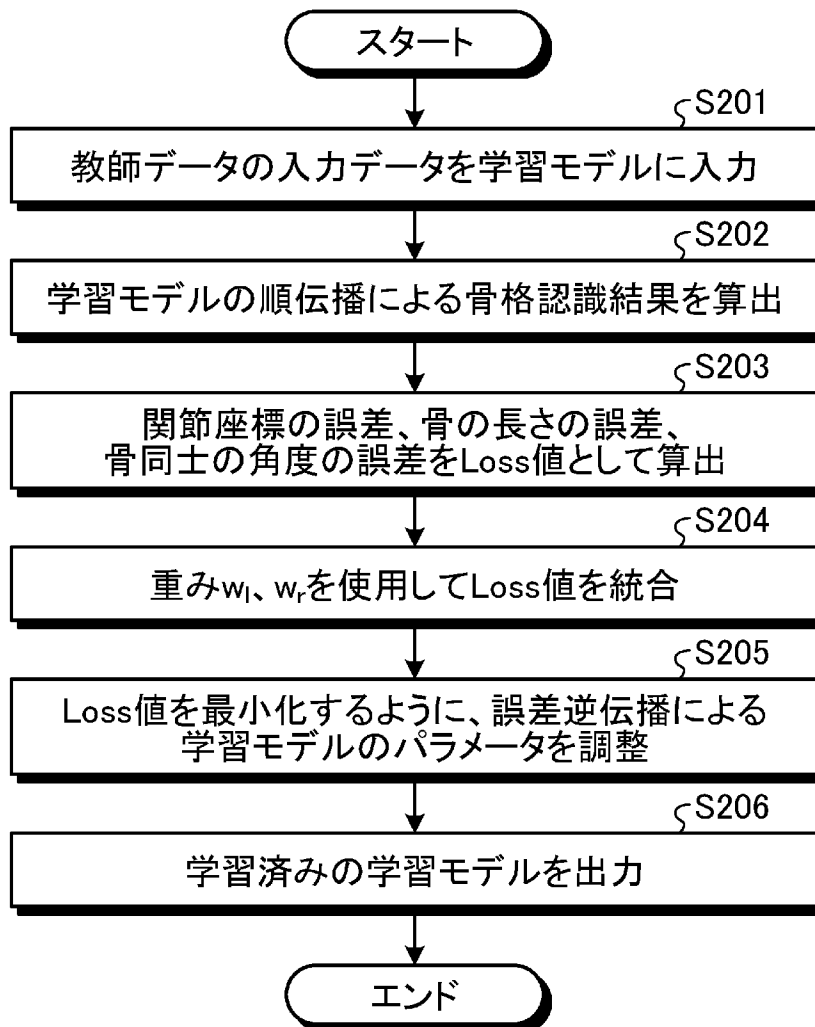
[図5]



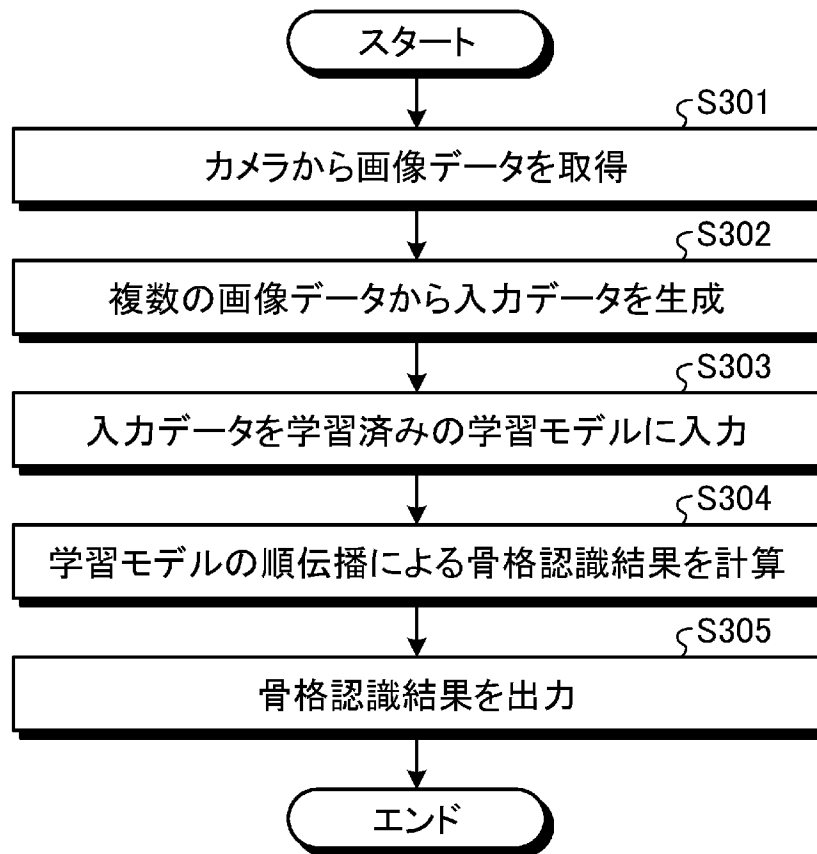
[図6]



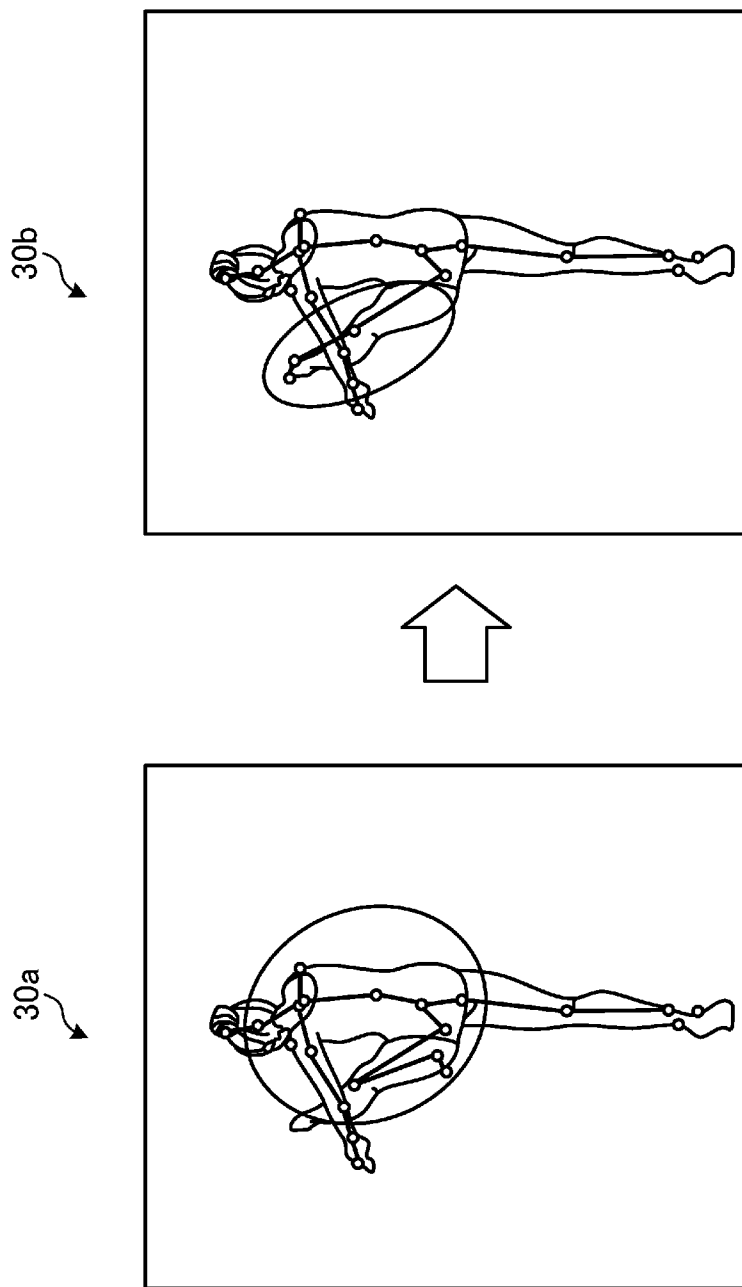
[図7]



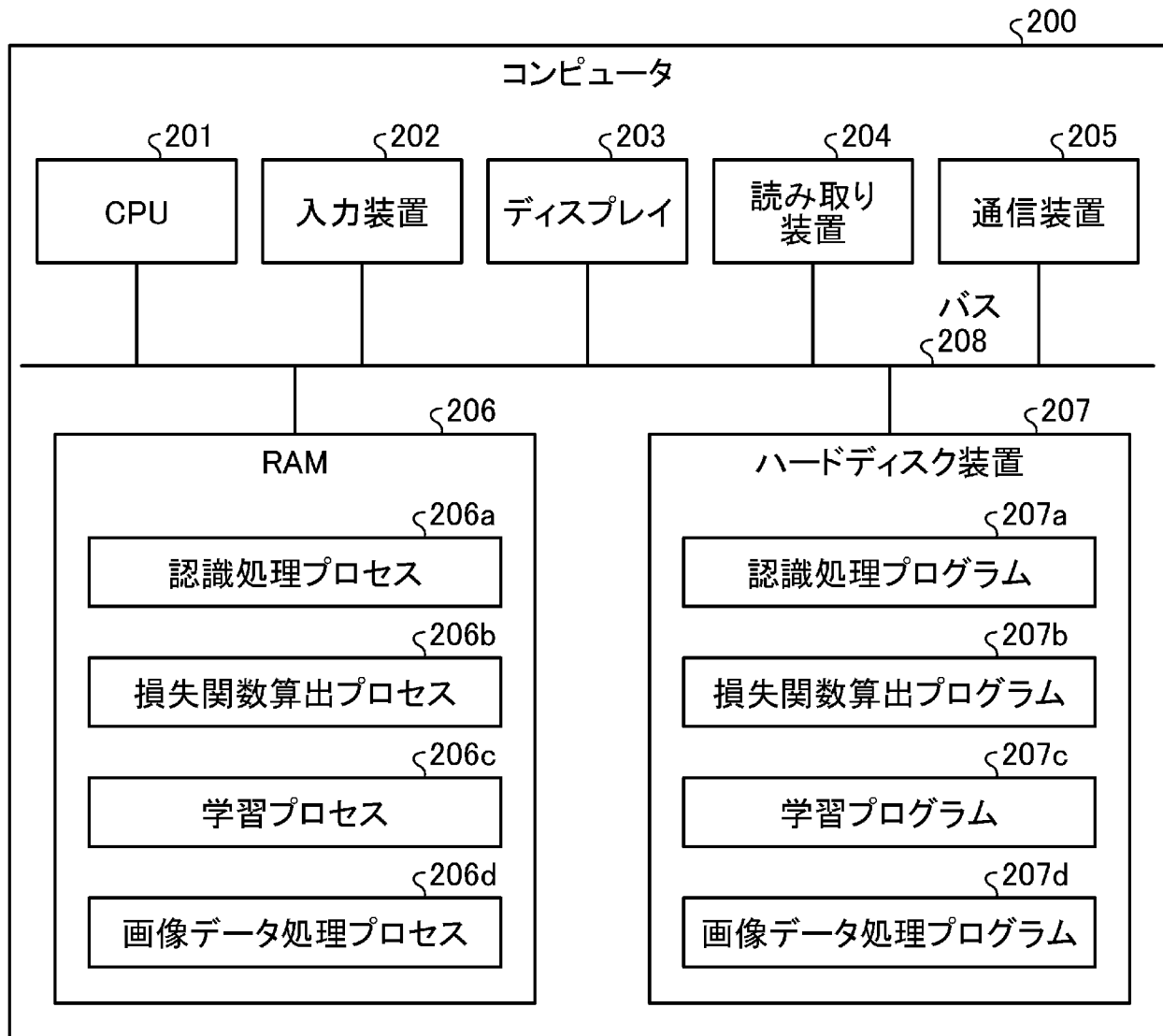
[図8]



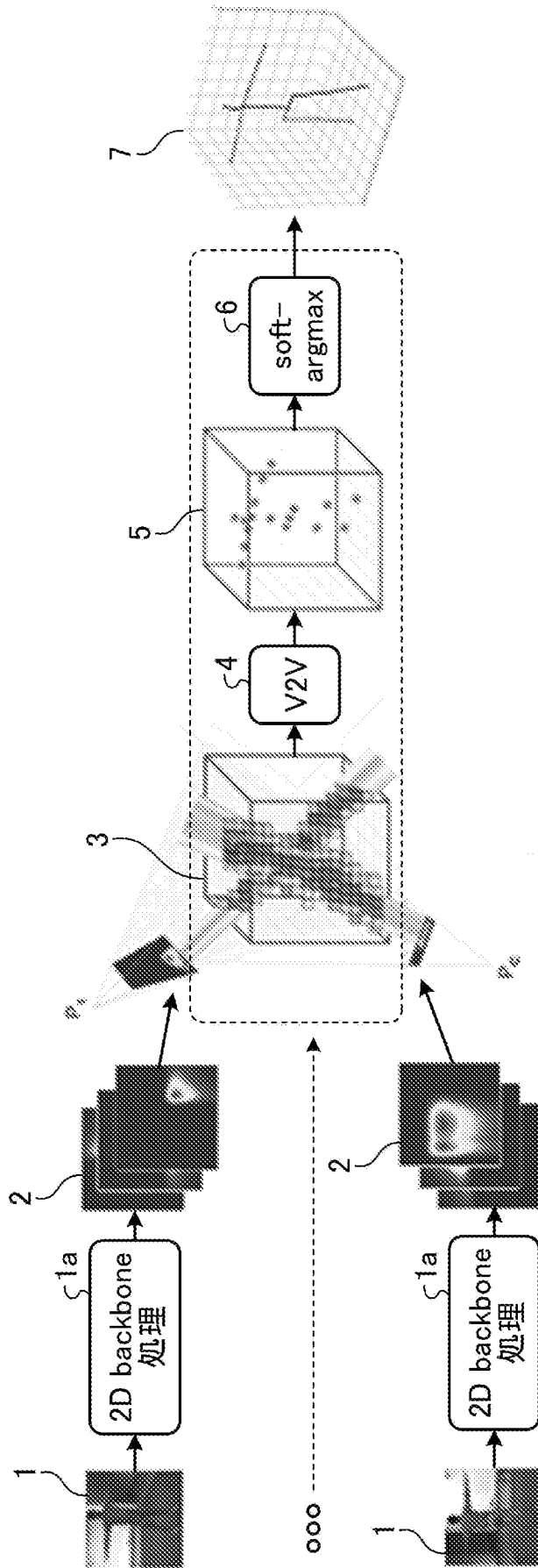
[図9]



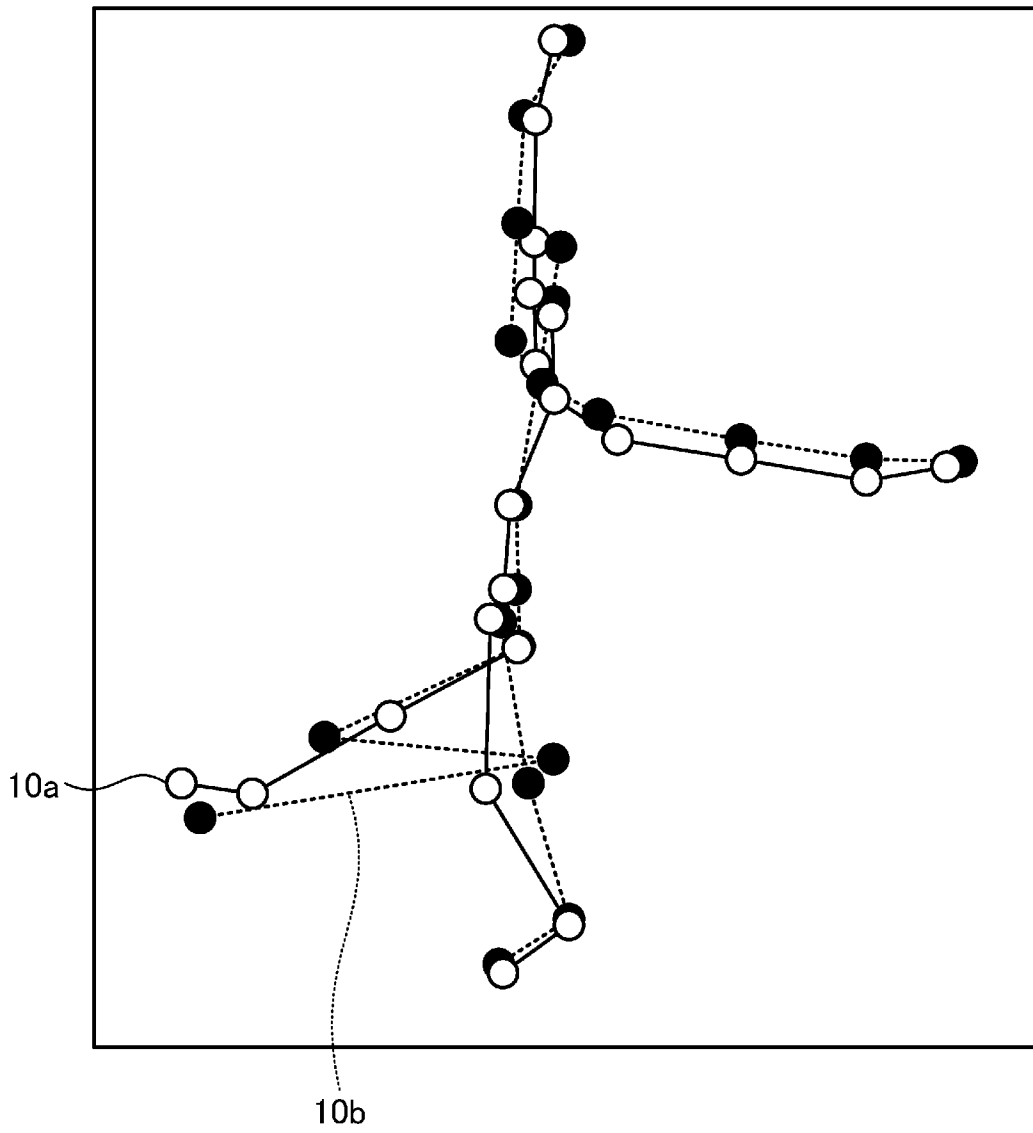
[図10]



[図11]



[図12]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2020/044524

A. CLASSIFICATION OF SUBJECT MATTER G06T 7/00 (2017.01) i FI: G06T7/00 350C 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		
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		1971-2020
Registered utility model specifications of Japan		1996-2020
Published registered utility model applications of Japan		1994-2020
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 2019-79487 A (NIPPON TELEGRAPH AND TELEPHONE CORP.) 23 May 2019 (2019-05-23) paragraphs [0015]-[0059]	1-12
A	JP 2019-191974 A (DNA KK.) 31 October 2019 (2019-10-31) entire text	1-12
A	ISKAKOV, Karim et al., "Learnable Triangulation of Human Pose", Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2019, Computer Vision Foundation, 2019, pp. 7718-7727, entire text	1-12
<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	"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	
"E" earlier application or patent but published on or after the international filing date	"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	
"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)	"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	
"O" document referring to an oral disclosure, use, exhibition or other means	"&" document member of the same patent family	
"P" document published prior to the international filing date but later than the priority date claimed		
Date of the actual completion of the international search 18 December 2020 (18.12.2020)	Date of mailing of the international search report 28 December 2020 (28.12.2020)	
Name and mailing address of the ISA/ Japan Patent Office 3-4-3, Kasumigaseki, Chiyoda-ku, Tokyo 100-8915, Japan	Authorized officer Telephone No.	

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application no.

PCT/JP2020/044524

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
JP 2019-79487 A	23 May 2019	(Family: none)	
JP 2019-191974 A	31 Oct. 2019	(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 FI: G06T7/00 350C		
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 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 - 2020年 日本国実用新案登録公報 1996 - 2020年 日本国登録実用新案公報 1994 - 2020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 2019-79487 A（日本電信電話株式会社）23.05.2019（2019 - 05 - 23） 段落[0015]-[0059]	1-12
A	JP 2019-191974 A（株式会社 デイ・エヌ・エー）31.10.2019（2019 - 10 - 31） 全文	1-12
A	ISKAKOV, Karim et al., Learnable Triangulation of Human Pose, Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2019, Computer Vision Foundation, 2019, pp.7718-7727 全文	1-12
<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 18.12.2020	国際調査報告の発送日 28.12.2020	
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） 岡本 俊威 5H 9178 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/JP2020/044524

引用文献	公表日	パテントファミリー文献	公表日
JP 2019-79487 A	23.05.2019	(ファミリーなし)	
JP 2019-191974 A	31.10.2019	(ファミリーなし)	