

(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1月13日(13.01.2022)



(10) 国際公開番号

WO 2022/009279 A1

(51) 国際特許分類:  
G06F 16/53 (2019.01) G06F 16/532 (2019.01)

(21) 国際出願番号: PCT/JP2020/026450

(22) 国際出願日: 2020年7月6日(06.07.2020)

(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: 日本電気株式会社 (NEC CORPORATION) [JP/JP]; 〒1088001 東京都港区芝五丁目7番1号 Tokyo (JP).

(72) 発明者: 吉田 登 (YOSHIDA Noboru); 〒1088001 東京都港区芝五丁目7番1号 日本電気株式会社内 Tokyo (JP).

(74) 代理人: 速水 進治 (HAYAMI Shinji); 〒1410031 東京都品川区西五反田7丁目9番2号 KDX五反田ビル9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,

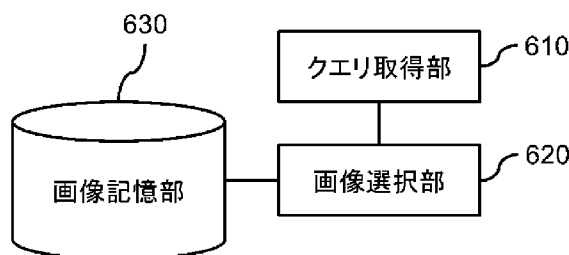
BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, IT, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, 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).

(54) Title: IMAGE SELECTION DEVICE, IMAGE SELECTION METHOD, AND PROGRAM

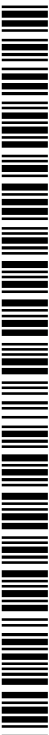
(54) 発明の名称: 画像選択装置、画像選択方法、及びプログラム

105



610 Query acquisition unit  
620 Image selection unit  
630 Image storage unit

(57) Abstract: This image selection device comprises a query acquisition unit (610) and an image selection unit (620). The query acquisition unit (610) acquires query information indicating the posture of a person. The query information includes the relative positions of a plurality of key points that indicate different parts of the body of the person. The image selection unit (620) selects at least one target image from among a plurality of selection target images by using the query information. Specifically, the image selection unit (620) determines the weighting of at least one key point by using a difference between reference posture information indicating the reference relative position of the plurality of key points and the query information. Then, the image selection unit (620) selects a target image by using the relative positions of the plurality of key points of the person included in the selection target images, the query information, and the weighting.



WO 2022/009279 A1

添付公開書類：

- 一 国際調査報告（条約第21条(3)）

---

(57) 要約：画像選択装置は、クエリ取得部（610）及び画像選択部（620）を備えている。クエリ取得部（610）は、人の姿勢を示すクエリ情報を取得する。クエリ情報は、人の身体の互いに異なる部分を示す複数のキーポイントの相対位置を含んでいる。画像選択部（620）は、クエリ情報を用いて、複数の選択対象画像から少なくとも一つの目的画像を選択する。詳細には、画像選択部（620）は、複数のキーポイントの基準相対位置を示す基準姿勢情報と、クエリ情報との差分を用いて、少なくとも一つのキーポイントの重み付けを決定する。そして画像選択部（620）は、選択対象画像に含まれる人の複数のキーポイントの相対位置、クエリ情報、及び重み付けを用いて目的画像を選択する。

## 明 細 書

**発明の名称**：画像選択装置、画像選択方法、及びプログラム

### 技術分野

[0001] 本発明は、画像選択装置、画像選択方法、及びプログラムに関する。

### 背景技術

[0002] 近年、監視システム等において、監視カメラの画像から人物の姿勢や行動等の状態の検出や検索を行う技術が利用されている。関連する技術として、例えば、特許文献1及び2が知られている。特許文献1には、深さ映像に含まれる人物の頭や手足等のキージョイントに基づいて、類似する人物の姿勢を検索する技術が開示されている。特許文献2には、人物の姿勢と関連しないが、画像に付加された傾き等の姿勢情報を利用して類似画像を検索する技術が開示されている。なお、その他に、人物の骨格推定に関連する技術として、非特許文献1が知られている。

[0003] また、特許文献3には、関節により接続された複数の部位を有する物体の姿勢を推定する装置において、姿勢毎に、部位の位置を規定する基準モデルを格納部に格納しておくとともに、姿勢毎に、密集部分の重みが小さくなるように部位に対して重み付けを行うことが記載されている。

### 先行技術文献

#### 特許文献

[0004] 特許文献1：特表2014-522035号公報

特許文献2：特開2006-260405号公報

特許文献3：特開2012-181736号公報

#### 非特許文献

[0005] 非特許文献1：Zhe Cao, Tomas Simon, Shih-En Wei, Yaser Sheikh, "Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields", The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, P. 7291-7299

## 発明の概要

### 発明が解決しようとする課題

[0006] 装置が人の姿勢を示す姿勢情報を用いて画像を選択する場合、所望する姿勢とは異なる姿勢の人を含む画像が検索結果に含まれる可能性がある。本発明の目的は、人の姿勢を示す姿勢情報を用いて画像を選択する場合において、選択精度を上げることにある。

### 課題を解決するための手段

[0007] 本発明によれば、人の姿勢を示すクエリ情報を取得するクエリ取得手段と、

前記クエリ情報を用いて、複数の選択対象画像から少なくとも一つの目的画像を選択する画像選択手段と、

を備え、

前記クエリ情報は、人の身体の互いに異なる部分を示す複数のキーポイントの相対位置を含んでおり、

前記画像選択手段は、

前記複数のキーポイントの基準相対位置を含む基準姿勢情報と、前記クエリ情報との差分を用いて、少なくとも一つの前記キーポイントの重み付けを決定し、

前記選択対象画像に含まれる人の前記複数のキーポイントの相対位置、前記クエリ情報、及び前記重み付けを用いて、前記少なくとも一つの目的画像を選択する、画像選択装置が提供される。

[0008] 本発明によれば、コンピュータが、

人の姿勢を示すクエリ情報を取得するクエリ取得処理と、

前記クエリ情報を用いて、複数の選択対象画像から少なくとも一つの目的画像を選択する画像選択処理と、

を行い

前記クエリ情報は、人の身体の互いに異なる部分を示す複数のキーポイントの相対位置を含んでおり、

前記画像選択処理において、前記コンピュータは、

前記複数のキーポイントの基準相対位置を含む基準姿勢情報と、前記クエリ情報との差分を用いて、少なくとも一つの前記キーポイントの重み付けを決定し、

前記選択対象画像に含まれる人の前記複数のキーポイントの相対位置、前記クエリ情報、及び前記重み付けを用いて、前記少なくとも一つの目的画像を選択する、画像選択方法が提供される。

[0009] 本発明によれば、コンピュータに、

人の姿勢を示すクエリ情報を取得するクエリ取得機能と、

前記クエリ情報を用いて、複数の選択対象画像から少なくとも一つの目的画像を選択する画像選択機能と、  
を持たせ、

前記クエリ情報は、人の身体の互いに異なる部分を示す複数のキーポイントの相対位置を含んでおり、

前記画像選択機能は、

前記複数のキーポイントの基準相対位置を含む基準姿勢情報と、前記クエリ情報との差分を用いて、少なくとも一つの前記キーポイントの重み付けを決定し、

前記選択対象画像に含まれる人の前記複数のキーポイントの相対位置、前記クエリ情報、及び前記重み付けを用いて、前記少なくとも一つの目的画像を選択する、プログラムが提供される。

### 発明の効果

[0010] 本発明によれば、人の姿勢を示す姿勢情報を用いて画像を選択する場合において、選択精度は上がる。

### 図面の簡単な説明

[0011] 上述した目的、およびその他の目的、特徴および利点は、以下に述べる好適な実施の形態、およびそれに付随する以下の図面によってさらに明らかになる。

- [0012] [図1]実施の形態に係る画像処理装置の概要を示す構成図である。
- [図2]実施の形態1に係る画像処理装置の構成を示す構成図である。
- [図3]実施の形態1に係る画像処理方法を示すフローチャートである。
- [図4]実施の形態1に係る分類方法を示すフローチャートである。
- [図5]実施の形態1に係る検索方法を示すフローチャートである。
- [図6]実施の形態1に係る骨格構造の検出例を示す図である。
- [図7]実施の形態1に係る人体モデルを示す図である。
- [図8]実施の形態1に係る骨格構造の検出例を示す図である。
- [図9]実施の形態1に係る骨格構造の検出例を示す図である。
- [図10]実施の形態1に係る骨格構造の検出例を示す図である。
- [図11]実施の形態1に係る分類方法の具体例を示すグラフである。
- [図12]実施の形態1に係る分類結果の表示例を示す図である。
- [図13]実施の形態1に係る検索方法を説明するための図である。
- [図14]実施の形態1に係る検索方法を説明するための図である。
- [図15]実施の形態1に係る検索方法を説明するための図である。
- [図16]実施の形態1に係る検索方法を説明するための図である。
- [図17]実施の形態1に係る検索結果の表示例を示す図である。
- [図18]実施の形態2に係る画像処理装置の構成を示す構成図である。
- [図19]実施の形態2に係る画像処理方法を示すフローチャートである。
- [図20]実施の形態2に係る身長画素数算出方法の具体例1を示すフローチャートである。
- [図21]実施の形態2に係る身長画素数算出方法の具体例2を示すフローチャートである。
- [図22]実施の形態2に係る身長画素数算出方法の具体例2を示すフローチャートである。
- [図23]実施の形態2に係る正規化方法を示すフローチャートである。
- [図24]実施の形態2に係る人体モデルを示す図である。
- [図25]実施の形態2に係る骨格構造の検出例を示す図である。

[図26]実施の形態2に係る骨格構造の検出例を示す図である。

[図27]実施の形態2に係る骨格構造の検出例を示す図である。

[図28]実施の形態2に係る人体モデルを示す図である。

[図29]実施の形態2に係る骨格構造の検出例を示す図である。

[図30]実施の形態2に係る身長画素数算出方法を説明するためのヒストグラムである。

[図31]実施の形態2に係る骨格構造の検出例を示す図である。

[図32]実施の形態2に係る3次元人体モデルを示す図である。

[図33]実施の形態2に係る身長画素数算出方法を説明するための図である。

[図34]実施の形態2に係る身長画素数算出方法を説明するための図である。

[図35]実施の形態2に係る身長画素数算出方法を説明するための図である。

[図36]実施の形態2に係る正規化方法を説明するための図である。

[図37]実施の形態2に係る正規化方法を説明するための図である。

[図38]実施の形態2に係る正規化方法を説明するための図である。

[図39]画像処理装置のハードウェア構成例を示す図である。

[図40]検索方法6に係る検索部の機能構成の一例を示す図である。

[図41] (A) は、基準姿勢情報の一例を示す図であり、(B) は、クエリ情報の一例を示す図である。

[図42]検索方法6において検索部が行う処理の第1例を示すフローチャートである。

[図43]検索方法6において検索部が行う処理の第2例を示すフローチャートである。

[図44]検索方法6において検索部が行う処理の第3例を示すフローチャートである。

### 発明を実施するための形態

[0013] 以下、本発明の実施の形態について、図面を用いて説明する。尚、すべての図面において、同様な構成要素には同様の符号を付し、適宜説明を省略する。

[0014] (実施の形態に至る検討)

近年、ディープラーニング等の機械学習を活用した画像認識技術が様々なシステムに応用されている。例えば、監視カメラの画像により監視を行う監視システムへの適用が進められている。監視システムに機械学習を活用することで、画像から人物の姿勢や行動等の状態をある程度把握することが可能とされつつある。

[0015] しかしながら、このような関連する技術では、必ずしもオンデマンドにユーザが望む人物の状態を把握できない場合がある。例えば、ユーザが検索し把握したい人物の状態を事前に特定できている場合もあれば、未知の状態のように具体的に特定できていない場合もある。そうすると、場合によっては、ユーザが検索したい人物の状態を詳細に指定することができない。また、人物の体の一部が隠れているような場合には検索等を行うことができない。関連する技術では、特定の検索条件のみからしか人物の状態を検索できないため、所望の人物の状態を柔軟に検索や分類することが困難である。

[0016] そこで、発明者らは、オンデマンドに画像からユーザ所望の人物の状態を認識するため、非特許文献1などの骨格推定技術を利用する方法を検討した。非特許文献1に開示されたOpenPose等のように、関連する骨格推定技術では、様々なパターンの正解付けされた画像データを学習することで、人物の骨格を推定する。以下の実施の形態では、このような骨格推定技術を活用することで、人物の状態を柔軟に認識することを可能とする。

[0017] なお、OpenPose等の骨格推定技術により推定される骨格構造は、関節等の特徴的な点である「キーポイント」と、キーポイント間のリンクを示す「ボーン（ボーンリンク）」とから構成される。このため、以下の実施の形態では、骨格構造について「キーポイント」と「ボーン」という用語を用いて説明するが、特に限定されない限り、「キーポイント」は人物の「関節」に対応し、「ボーン」は人物の「骨」に対応している。

[0018] (実施の形態の概要)

図1は、実施の形態に係る画像処理装置10の概要を示している。図1に

示すように、画像処理装置10は、骨格検出部11、特徴量算出部12、及び認識部13を備えている。骨格検出部11は、カメラ等から取得される2次元画像に基づいて、複数の人物の2次元骨格構造を検出する。特徴量算出部12は、骨格検出部11により検出された複数の2次元骨格構造の特徴量を算出する。認識部13は、特徴量算出部12により算出された複数の特徴量の類似度に基づいて、複数の人物の状態の認識処理を行う。認識処理は、人物の状態の分類処理や検索処理（選択処理）等である。このため、画像処理装置10は画像選択装置としても機能する。

[0019] このように、実施の形態では、2次元画像から人物の2次元骨格構造を検出し、この2次元骨格構造から算出される特徴量に基づいて人物の状態の分類や検討等の認識処理を行うことで、所望の人物の状態を柔軟に認識することができる。

[0020] （実施の形態1）以下、図面を参照して実施の形態1について説明する。図2は、本実施の形態に係る画像処理装置100の構成を示している。画像処理装置100は、カメラ200及びデータベース（DB）110とともに画像処理システム1を構成する。画像処理装置100を含む画像処理システム1は、画像から推定される人物の骨格構造に基づき、人物の姿勢や行動等の状態を分類及び検索するシステムである。なお、画像処理装置100も、画像選択装置としても機能する。

[0021] カメラ200は、2次元の画像を生成する監視カメラ等の撮像部である。カメラ200は、所定の箇所に設置されて、設置個所から撮像領域における人物等を撮像する。カメラ200は、撮像した画像（映像）を画像処理装置100へ出力可能に直接接続、もしくはネットワーク等を介して接続されている。なお、カメラ200を画像処理装置100の内部に設けてもよい。

[0022] データベース110は、画像処理装置100の処理に必要な情報（データ）や処理結果等を格納するデータベースである。データベース110は、画像取得部101が取得した画像や、骨格構造検出部102の検出結果、機械学習用のデータ、特徴量算出部103が算出した特徴量、分類部104の分

類結果、検索部105の検索結果等を記憶する。データベース110は、画像処理装置100と必要に応じてデータを入出力可能に直接接続、もしくはネットワーク等を介して接続されている。なお、データベース110をフラッシュメモリなどの不揮発性メモリやハードディスク装置等として、画像処理装置100の内部に設けてもよい。

[0023] 図2に示すように、画像処理装置100は、画像取得部101、骨格構造検出部102、特徴量算出部103、分類部104、検索部105、入力部106、及び表示部107を備えている。なお、各部（ブロック）の構成は一例であり、後述の方法（動作）が可能であれば、その他の各部で構成されてもよい。また、画像処理装置100は、例えば、プログラムを実行するパーソナルコンピュータやサーバ等のコンピュータ装置で実現されるが、1つの装置で実現してもよいし、ネットワーク上の複数の装置で実現してもよい。例えば、入力部106や表示部107等を外部の装置としてもよい。また、分類部104及び検索部105の両方を備えていてもよいし、いずれか一方のみを備えていてもよい。分類部104及び検索部105の両方、もしくは一方は、人物の状態の認識処理を行う認識部である。

[0024] 画像取得部101は、カメラ200が撮像した人物を含む2次元の画像を取得する。画像取得部101は、例えば、所定の監視期間にカメラ200が撮像した、人物を含む画像（複数の画像を含む映像）を取得する。なお、カメラ200からの取得に限らず、予め用意された人物を含む画像をデータベース110等から取得してもよい。

[0025] 骨格構造検出部102は、取得された2次元の画像に基づき、画像内の人物の2次元の骨格構造を検出する。骨格構造検出部102は、取得された画像の中で認識される全ての人物について、骨格構造を検出する。骨格構造検出部102は、機械学習を用いた骨格推定技術を用いて、認識される人物の関節等の特徴に基づき人物の骨格構造を検出する。骨格構造検出部102は、例えば、非特許文献1のOpenPose等の骨格推定技術を用いる。

[0026] 特徴量算出部103は、検出された2次元の骨格構造の特徴量を算出し、

算出した特徴量を、処理対象となった画像に紐づけてデータベース110に格納する。骨格構造の特徴量は、人物の骨格の特徴を示しており、人物の骨格に基づいて人物の状態を分類や検索するための要素となる。通常、この特徴量は、複数のパラメータ（例えば後述する分類要素）を含んでいる。そして特徴量は、骨格構造の全体の特徴量でもよいし、骨格構造の一部の特徴量でもよく、骨格構造の各部のように複数の特徴量を含んでもよい。特徴量の算出方法は、機械学習や正規化等の任意の方法でよく、正規化として最小値や最大値を求めてもよい。一例として、特徴量は、骨格構造を機械学習することで得られた特徴量や、骨格構造の頭部から足部までの画像上の大きさ等である。骨格構造の大きさは、画像上の骨格構造を含む骨格領域の上下方向の高さや面積等である。上下方向（高さ方向または縦方向）は、画像における上下の方向（Y軸方向）であり、例えば、地面（基準面）に対し垂直な方向である。また、左右方向（横方向）は、画像における左右の方向（X軸方向）であり、例えば、地面に対し平行な方向である。

[0027] なお、ユーザが望む分類や検索を行うためには、分類や検索処理に対しロバスト性を有する特徴量を用いることが好ましい。例えば、ユーザが、人物の向きや体型に依存しない分類や検索を望む場合、人物の向きや体型にロバストな特徴量を使用してもよい。同じ姿勢で様々な方向に向いている人物の骨格や同じ姿勢で様々な体型の人物の骨格を学習することや、骨格の上下方向のみの特徴を抽出することで、人物の向きや体型に依存しない特徴量を得ることができる。

[0028] 分類部104は、データベース110に格納された複数の骨格構造を、骨格構造の特徴量の類似度に基づいて分類する（クラスタリングする）。分類部104は、人物の状態の認識処理として、骨格構造の特徴量に基づいて複数の人物の状態を分類しているとも言える。類似度は、骨格構造の特徴量間の距離である。分類部104は、骨格構造の全体の特徴量の類似度により分類してもよいし、骨格構造の一部の特徴量の類似度により分類してもよく、骨格構造の第1の部分（例えば両手）及び第2の部分（例えば両足）の特徴

量の類似度により分類してもよい。なお、各画像における人物の骨格構造の特微量に基づいて人物の姿勢を分類してもよいし、時系列に連続する複数の画像における人物の骨格構造の特微量の変化に基づいて人物の行動を分類してもよい。すなわち、分類部104は、骨格構造の特微量に基づいて人物の姿勢や行動を含む人物の状態を分類できる。例えば、分類部104は、所定の監視期間に撮像された複数の画像における複数の骨格構造を分類対象とする。分類部104は、分類対象の特微量間の類似度を求め、類似度の高い骨格構造が同じクラス（似た姿勢のグループ）となるように分類する。なお、検索と同様に、分類条件をユーザが指定できるようにしてもよい。分類部104は、骨格構造の分類結果をデータベース110に格納するとともに、表示部107に表示する。

[0029] 検索部105は、データベース110に格納された複数の骨格構造の中から、検索クエリ（クエリ状態）の特微量と類似度の高い骨格構造を検索する。検索部105は、人物の状態の認識処理として、骨格構造の特微量に基づいて複数の人物の状態の中から、検索条件（クエリ状態）に該当する人物の状態を検索しているとも言える。分類と同様に、類似度は、骨格構造の特微量間の距離である。検索部105は、骨格構造の全体の特微量の類似度により検索してもよいし、骨格構造の一部の特微量の類似度により検索してもよく、骨格構造の第1の部分（例えば両手）及び第2の部分（例えば両足）の特微量の類似度により検索してもよい。なお、各画像における人物の骨格構造の特微量に基づいて人物の姿勢を検索してもよいし、時系列に連続する複数の画像における人物の骨格構造の特微量の変化に基づいて人物の行動を検索してもよい。すなわち、検索部105は、骨格構造の特微量に基づいて人物の姿勢や行動を含む人物の状態を検索できる。例えば、検索部105は、分類対象と同様に、所定の監視期間に撮像された複数の画像における複数の骨格構造の特微量を検索対象とする。また、分類部104が表示した分類結果の中からユーザが指定した骨格構造（姿勢）を検索クエリ（検索キー）とする。なお、分類結果に限らず、分類されていない複数の骨格構造の中から

検索クエリを選択してもよいし、検索クエリとなる骨格構造をユーザが入力してもよい。検索部105は、検索対象の特徴量の中から、検索クエリの骨格構造の特徴量と類似度の高い特徴量を検索する。検索部105は、特徴量の検索結果をデータベース110に格納するとともに、表示部107に表示する。

[0030] 入力部106は、画像処理装置100を操作するユーザから入力された情報を取得する入力インタフェースである。例えば、ユーザは、監視カメラの画像から不審な状態の人物を監視する監視者である。入力部106は、例えば、GUI (Graphical User Interface) であり、キーボードやマウス、タッチパネル等の入力装置から、ユーザの操作に応じた情報が入力される。例えば、入力部106は、分類部104により分類された骨格構造(姿勢)の中から、指定された人物の骨格構造を検索クエリとして受け付ける。

[0031] 表示部107は、画像処理装置100の動作(処理)の結果等を表示する表示部であり、例えば、液晶ディスプレイや有機EL (Electro Luminescence) ディスプレイ等のディスプレイ装置である。表示部107は、分類部104の分類結果や検索部105の検索結果を類似度等に応じてGUIに表示する。

[0032] 図39は、画像処理装置100のハードウェア構成例を示す図である。画像処理装置100は、バス1010、プロセッサ1020、メモリ1030、ストレージデバイス1040、入出力インタフェース1050、及びネットワークインタフェース1060を有する。

[0033] バス1010は、プロセッサ1020、メモリ1030、ストレージデバイス1040、入出力インタフェース1050、及びネットワークインタフェース1060が、相互にデータを送受信するためのデータ伝送路である。ただし、プロセッサ1020などを互いに接続する方法は、バス接続に限定されない。

[0034] プロセッサ1020は、CPU (Central Processing Unit) やGPU (Graphics Processing Unit) などで実現されるプロセッサである。

- [0035] メモリ1030は、RAM (Random Access Memory) などで実現される主記憶装置である。
- [0036] ストレージデバイス1040は、HDD (Hard Disk Drive)、SSD (Solid State Drive)、メモリカード、又はROM (Read Only Memory) などで実現される補助記憶装置である。ストレージデバイス1040は画像処理装置100の各機能（例えば画像取得部101、骨格構造検出部102、特徴量算出部103、分類部104、検索部105、及び入力部106）を実現するプログラムモジュールを記憶している。プロセッサ1020がこれら各プログラムモジュールをメモリ1030上に読み込んで実行することで、そのプログラムモジュールに対応する各機能が実現される。また、ストレージデバイス1040はデータベース110としても機能することもある。
- [0037] 入出力インタフェース1050は、画像処理装置100と各種入出力機器とを接続するためのインタフェースである。データベース110が画像処理装置100の外部に位置する場合、画像処理装置100は、入出力インタフェース1050を介してデータベース110と接続してもよい。
- [0038] ネットワークインタフェース1060は、画像処理装置100をネットワークに接続するためのインタフェースである。このネットワークは、例えばLAN (Local Area Network) やWAN (Wide Area Network) である。ネットワークインタフェース1060がネットワークに接続する方法は、無線接続であってもよいし、有線接続であってもよい。画像処理装置100は、ネットワークインタフェース1060を介してカメラ200と通信してもよい。データベース110が画像処理装置100の外部に位置する場合、画像処理装置100は、ネットワークインタフェース1060を介してデータベース110と接続してもよい。
- [0039] 図3～図5は、本実施の形態に係る画像処理装置100の動作を示している。図3は、画像処理装置100における画像取得から検索処理までの流れを示し、図4は、図3の分類処理 (S104) の流れを示し、図5は、図3の検索処理 (S105) の流れを示している。

[0040] 図3に示すように、画像処理装置100は、カメラ200から画像を取得する(S101)。画像取得部101は、骨格構造から分類や検索を行うために人物を撮像した画像を取得し、取得した画像をデータベース110に格納する。画像取得部101は、例えば、所定の監視期間に撮像された複数の画像を取得し、複数の画像に含まれる全ての人物について以降の処理を行う。

[0041] 続いて、画像処理装置100は、取得した人物の画像に基づいて人物の骨格構造を検出する(S102)。図6は、骨格構造の検出例を示している。図6に示すように、監視カメラ等から取得した画像には複数の人物が含まれており、画像に含まれる各人物について骨格構造を検出する。

[0042] 図7は、このとき検出する人体モデル300の骨格構造を示しており、図8～図10は、骨格構造の検出例を示している。骨格構造検出部102は、OpenPose等の骨格推定技術を用いて、2次元の画像から図7のような人体モデル(2次元骨格モデル)300の骨格構造を検出する。人体モデル300は、人物の関節等のキーポイントと、各キーポイントを結ぶボーンから構成された2次元モデルである。

[0043] 骨格構造検出部102は、例えば、画像の中からキーポイントとなり得る特徴点を抽出し、キーポイントの画像を機械学習した情報を参照して、人物の各キーポイントを検出する。図7の例では、人物のキーポイントとして、頭A1、首A2、右肩A31、左肩A32、右肘A41、左肘A42、右手A51、左手A52、右腰A61、左腰A62、右膝A71、左膝A72、右足A81、左足A82を検出する。さらに、これらのキーポイントを連結した人物の骨として、頭A1と首A2を結ぶボーンB1、首A2と右肩A31及び左肩A32をそれぞれ結ぶボーンB21及びボーンB22、右肩A31及び左肩A32と右肘A41及び左肘A42をそれぞれ結ぶボーンB31及びボーンB32、右肘A41及び左肘A42と右手A51及び左手A52をそれぞれ結ぶボーンB41及びボーンB42、首A2と右腰A61及び左腰A62をそれぞれ結ぶボーンB51及びボーンB52、右腰A61及び左

腰A 6 2と右膝A 7 1及び左膝A 7 2をそれぞれ結ぶボーンB 6 1及びボーンB 6 2、右膝A 7 1及び左膝A 7 2と右足A 8 1及び左足A 8 2をそれぞれ結ぶボーンB 7 1及びボーンB 7 2を検出する。骨格構造検出部1 0 2は、検出した人物の骨格構造をデータベース1 1 0に格納する。

[0044] 図8は、直立した状態の人物を検出する例である。図8では、直立した人物が正面から撮像されており、正面から見たボーンB 1、ボーンB 5 1及びボーンB 5 2、ボーンB 6 1及びボーンB 6 2、ボーンB 7 1及びボーンB 7 2がそれぞれ重ならず検出され、右足のボーンB 6 1及びボーンB 7 1は左足のボーンB 6 2及びボーンB 7 2よりも多少折れ曲がっている。

[0045] 図9は、しゃがみ込んでいる状態の人物を検出する例である。図9では、しゃがみ込んでいる人物が右側から撮像されており、右側から見たボーンB 1、ボーンB 5 1及びボーンB 5 2、ボーンB 6 1及びボーンB 6 2、ボーンB 7 1及びボーンB 7 2がそれぞれ検出され、右足のボーンB 6 1及びボーンB 7 1と左足のボーンB 6 2及びボーンB 7 2は大きく折れ曲がり、かつ、重なっている。

[0046] 図10は、寝込んでいる状態の人物を検出する例である。図10では、寝込んでいる人物が左斜め前から撮像されており、左斜め前から見たボーンB 1、ボーンB 5 1及びボーンB 5 2、ボーンB 6 1及びボーンB 6 2、ボーンB 7 1及びボーンB 7 2がそれぞれ検出され、右足のボーンB 6 1及びボーンB 7 1と左足のボーンB 6 2及びボーンB 7 2は折れ曲がり、かつ、重なっている。

[0047] 続いて、図3に示すように、画像処理装置1 0 0は、検出された骨格構造の特徴量を算出する(S 1 0 3)。例えば、骨格領域の高さや面積を特徴量とする場合、特徴量算出部1 0 3は、骨格構造を含む領域を抽出し、その領域の高さ(画素数)や面積(画素面積)を求める。骨格領域の高さや面積は、抽出される骨格領域の端部の座標や端部のキーポイントの座標から求められる。特徴量算出部1 0 3は、求めた骨格構造の特徴量をデータベース1 1 0に格納する。なお、この骨格構造の特徴量は、上記したキーポイント及び

ボーンと共に、人物の姿勢を示す姿勢情報としても用いられる。

[0048] 図8の例では、直立した人物の骨格構造から全てのボーンを含む骨格領域を抽出する。この場合、骨格領域の上端は頭部のキーポイントA1、骨格領域の下端は左足のキーポイントA82、骨格領域の左端は右肘のキーポイントA41、骨格領域の右端は左手のキーポイントA52となる。このため、キーポイントA1とキーポイントA82のY座標の差分から骨格領域の高さを求める。また、キーポイントA41とキーポイントA52のX座標の差分から骨格領域の幅を求め、骨格領域の高さと幅から面積を求める。

[0049] 図9の例では、しゃがみ込んだ人物の骨格構造から全てのボーンを含む骨格領域を抽出する。この場合、骨格領域の上端は頭部のキーポイントA1、骨格領域の下端は右足のキーポイントA81、骨格領域の左端は右腰のキーポイントA61、骨格領域の右端は右手のキーポイントA51となる。このため、キーポイントA1とキーポイントA81のY座標の差分から骨格領域の高さを求める。また、キーポイントA61とキーポイントA51のX座標の差分から骨格領域の幅を求め、骨格領域の高さと幅から面積を求める。

[0050] 図10の例では、画像の左右方向に寝込んだ人物の骨格構造から全てのボーンを含む骨格領域を抽出する。この場合、骨格領域の上端は左肩のキーポイントA32、骨格領域の下端は左手のキーポイントA52、骨格領域の左端は右手のキーポイントA51、骨格領域の右端は左足のキーポイントA82となる。このため、キーポイントA32とキーポイントA52のY座標の差分から骨格領域の高さを求める。また、キーポイントA51とキーポイントA82のX座標の差分から骨格領域の幅を求め、骨格領域の高さと幅から面積を求める。

[0051] 続いて、図3に示すように、画像処理装置100は、分類処理を行う(S104)。分類処理では、図4に示すように、分類部104は、算出された骨格構造の特徴量の類似度を算出し(S111)、算出された特徴量に基づいて骨格構造を分類する(S112)。分類部104は、分類対象であるデータベース110に格納されている全ての骨格構造間の特徴量の類似度を求

め、最も類似度が高い骨格構造（姿勢）を同じクラスタに分類する（クラスタリングする）。さらに、分類したクラスタ間の類似度を求めて分類し、所定の数のクラスタとなるまで分類を繰り返す。図11は、骨格構造の特徴量の分類結果のイメージを示している。図11は、2次元の分類要素によるクラスタ分析のイメージであり、2つ分類要素は、例えば、骨格領域の高さと骨格領域の面積等である。図11では、分類の結果、複数の骨格構造の特徴量が3つのクラスタC1～C3に分類されている。クラスタC1～C3は、例えば、立っている姿勢、座っている姿勢、寝ている姿勢のように各姿勢に対応し、似ている姿勢ごとに骨格構造（人物）が分類される。

[0052] 本実施の形態では、人物の骨格構造の特徴量に基づいて分類することにより、多様な分類方法を用いることができる。なお、分類方法は、予め設定されていてもよいし、ユーザが任意に設定できるようにしてもよい。また、後述する検索方法と同じ方法により分類を行ってもよい。つまり、検索条件と同様の分類条件により分類してもよい。例えば、分類部104は、次の分類方法により分類を行う。いずれかの分類方法を用いてもよいし、任意に選択された分類方法を組み合わせてもよい。

[0053] （分類方法1）複数の階層による分類全身の骨格構造による分類や、上半身や下半身の骨格構造による分類、腕や脚の骨格構造による分類等を階層的に組み合わせて分類する。すなわち、骨格構造の第1の部分や第2の部分の特徴量に基づいて分類し、さらに、第1の部分や第2の部分の特徴量に重みづけを行って分類してもよい。

[0054] （分類方法2）時系列に沿った複数枚の画像による分類時系列に連続する複数の画像における骨格構造の特徴量に基づいて分類する。例えば、時系列方向に特徴量を積み重ねて、累積値に基づいて分類してもよい。さらに、連続する複数の画像における骨格構造の特徴量の変化（変化量）に基づいて分類してもよい。

[0055] （分類方法3）骨格構造の左右を無視した分類人物の右側と左側が反対の骨格構造を同じ骨格構造として分類する。

[0056] さらに、分類部104は、骨格構造の分類結果を表示する(S113)。  
分類部104は、データベース110から必要な骨格構造や人物の画像を取得し、分類結果として似ている姿勢(クラスタ)ごとに骨格構造及び人物を表示部107に表示する。図12は、姿勢を3つに分類した場合の表示例を示している。例えば、図12に示すように、表示ウィンドウW1に、姿勢ごとの姿勢領域WA1~WA3を表示し、姿勢領域WA1~WA3にそれぞれ該当する姿勢の骨格構造及び人物(イメージ)を表示する。姿勢領域WA1は、例えば立っている姿勢の表示領域であり、クラスタC1に分類された、立っている姿勢に似た骨格構造及び人物を表示する。姿勢領域WA2は、例えば座っている姿勢の表示領域であり、クラスタC2に分類された、座っている姿勢に似た骨格構造及び人物を表示する。姿勢領域WA3は、例えば寝ている姿勢の表示領域であり、クラスタC2に分類された、寝ている姿勢に似た骨格構造及び人物を表示する。

[0057] 続いて、図3に示すように、画像処理装置100は、検索処理を行う(S105)。検索処理では、図5に示すように、検索部105は、検索条件の入力を受け付け(S121)、検索条件に基づいて骨格構造を検索する(S122)。検索部105は、入力部106から、ユーザの操作に応じて検索条件である検索クエリの入力を受け付ける。分類結果から検索クエリを入力する場合、例えば、図12の表示例では、ユーザは、表示ウィンドウW1に表示されている姿勢領域WA1~WA3の中から検索したい姿勢の骨格構造を指定(選択)する。そうすると、検索部105は、ユーザにより指定された骨格構造を検索クエリとして、検索対象であるデータベース110に格納されている全ての骨格構造の中から特徴量の類似度が高い骨格構造を検索する。検索部105は、検索クエリの骨格構造の特徴量と検索対象の骨格構造の特徴量との類似度を算出し、算出した類似度が所定の閾値よりも高い骨格構造を抽出する。検索クエリの骨格構造の特徴量は、予め算出された特徴量を使用してもよいし、検索時に求めた特徴量を使用してもよい。なお、検索クエリは、ユーザの操作に応じて骨格構造の各部を動かすことで入力しても

よいし、ユーザがカメラの前で実演した姿勢を検索クエリとしてもよい。

[0058] 本実施の形態では、分類方法と同様に、人物の骨格構造の特徴量に基づいて検索することにより、多様な検索方法を用いることができる。なお、検索方法は、予め設定されていてもよいし、ユーザが任意に設定できるようにしてもよい。例えば、検索部105は、次の検索方法により検索を行う。いずれかの検索方法を用いてもよいし、任意に選択された検索方法を組み合わせてもよい。複数の検索方法（検索条件）を論理式（例えばAND（論理積）、OR（論理和）、NOT（否定））により組み合わせて検索してもよい。例えば、検索条件を「（右手を挙げている姿勢）AND（左足を挙げている姿勢）」として検索してもよい。

[0059] （検索方法1）高さ方向の特徴量のみによる検索人物の高さ方向の特徴量のみを用いて検索することで、人物の横方向の変化の影響を抑えることができ、人物の向きや人物の体型の変化に対しロバスト性が向上する。例えば、図13の骨格構造501～503のように、人物の向きや体型が異なる場合でも、高さ方向の特徴量は大きく変化しない。このため、骨格構造501～503では、検索時（分類時）に同じ姿勢であると判断することができる。

[0060] （検索方法2）部分検索画像において人物の体の一部が隠れている場合、認識可能な部分の情報のみを用いて検索する。例えば、図14の骨格構造511及び512のように、左足が隠れていることにより、左足のキーポイントが検出できない場合でも、検出されている他のキーポイントの特徴量を使用して検索できる。このため、骨格構造511及び512では、検索時（分類時）に同じ姿勢であると判断することができる。つまり、全てのキーポイントではなく、一部のキーポイントの特徴量を用いて、分類や検索を行うことができる。図15の骨格構造521及び522の例では、両足の向きが異なっているものの、上半身のキーポイント（A1、A2、A31、A32、A41、A42、A51、A52）の特徴量を検索クエリとすることで、同じ姿勢であると判断することができる。また、検索したい部分（特徴点）に対して、重みを付けて検索してもよいし、類似度判定の閾値を変化させても

よい。体の一部が隠れている場合、隠れた部分を無視して検索してもよいし、隠れた部分を加味して検索してもよい。隠れた部分も含めて検索することで、同じ部位が隠れているような姿勢を検索することができる。

[0061] (検索方法3) 骨格構造の左右を無視した検索人物の右側と左側が反対の骨格構造を同じ骨格構造として検索する。例えば、図16の骨格構造531及び532のように、右手を挙げている姿勢と、左手を挙げている姿勢を同じ姿勢として検索(分類)できる。図16の例では、骨格構造531と骨格構造532は、右手のキーポイントA51、右肘のキーポイントA41、左手のキーポイントA52、左肘のキーポイントA42の位置が異なるものの、その他のキーポイントの位置は同じである。骨格構造531の右手のキーポイントA51及び右肘のキーポイントA41と骨格構造532の左手のキーポイントA52及び左肘のキーポイントA42のうち、一方の骨格構造のキーポイントを左右反転させると、他方の骨格構造のキーポイントと同じ位置となり、また、骨格構造531の左手のキーポイントA52及び左肘のキーポイントA42と骨格構造532の右手のキーポイントA51及び右肘のキーポイントA41のうち、一方の骨格構造のキーポイントを左右反転させると、他方の骨格構造のキーポイントと同じ位置となるため、同じ姿勢と判断する。

[0062] (検索方法4) 縦方向と横方向の特徴量による検索人物の縦方向(Y軸方向)の特徴量のみで検索を行った後、得られた結果をさらに人物の横方向(X軸方向)の特徴量を用いて検索する。

[0063] (検索方法5) 時系列に沿った複数枚の画像による検索時系列に連続する複数の画像における骨格構造の特徴量に基づいて検索する。例えば、時系列方向に特徴量を積み重ねて、累積値に基づいて検索してもよい。さらに、連続する複数の画像における骨格構造の特徴量の変化(変化量)に基づいて検索してもよい。

[0064] さらに、検索部105は、骨格構造の検索結果を表示する(S123)。検索部105は、データベース110から必要な骨格構造や人物の画像を取

得し、検索結果として得られた骨格構造及び人物を表示部107に表示する。例えば、検索クエリ（検索条件）が複数指定されている場合、検索クエリごとに検索結果を表示する。図17は、3つの検索クエリ（姿勢）により検索した場合の表示例を示している。例えば、図17に示すように、表示ウィンドウW2において、左端部に指定された検索クエリQ10、Q20、Q30の骨格構造及び人物を表示し、検索クエリQ10、Q20、Q30の右側に各検索クエリの検索結果Q11、Q21、Q31の骨格構造及び人物を並べて表示する。

[0065] 検索結果を検索クエリの隣から並べて表示する順番は、該当する骨格構造が見つかった順でもよいし、類似度が高い順でもよい。部分検索の部分（特徴点）に重みを付けて検索した場合に、重み付けて計算した類似度順に表示してもよい。ユーザが選択した部分（特徴点）のみから計算した類似度順に表示してもよい。また、検索結果の画像（フレーム）を中心に、時系列の前後の画像（フレーム）を一定時間分切り出して表示してもよい。

[0066] （検索方法6）本検索方法において、検索部105は、検索クエリ（以下、クエリ情報とも記載）として、上記した骨格構造を用いる。この骨格構造は、人の姿勢を示している。そして検索部105は、基準となる姿勢と、クエリ情報との差分を用いて、少なくとも一つのキーポイントの重み付けを決定する。そして検索部105は、この重み付けを用いて、クエリ情報が示す姿勢に類似している人を含む画像（以下、目的画像と記載）を選択する。

[0067] 図40は、本検索方法に係る検索部105の機能構成の一例を示す図である。本図において、検索部105は、クエリ取得部610及び画像選択部620を有している。

[0068] クエリ取得部610は、クエリ情報を取得する。クエリ情報すなわち骨格構造は、複数のキーポイントそれぞれの相対位置を示す情報を含んでいる。上記したように、複数のキーポイントは、いずれも人の身体の互いに異なる部分、例えば関節を示している。クエリ取得部610は、このクエリ情報を、クエリとして入力された画像を処理することにより生成してもよい。また

クエリ取得部610は、骨格情報そのものをクエリ情報として取得してもよい。

[0069] 画像選択部620は、複数の選択対象画像から少なくとも一つの目的画像を選択する。詳細には、画像選択部620は、目的画像を選択する際に、基準となる姿勢を示す基準姿勢情報を用いる。基準姿勢情報は、複数のキーポイントの相対位置の基準的な位置すなわち基準相対位置（標準的な相対位置と表現することもできる）を含んでいる。画像選択部620は、この基準姿勢情報と、クエリ情報との差分を用いて、少なくとも一つのキーポイントの重み付けを決定する。そして画像選択部620は、複数の選択対象画像それぞれに含まれる人の複数のキーポイントの相対位置、クエリ情報、及び重み付けを用いて、少なくとも一つの目的画像を選択する。一例として画像選択部620は、クエリ情報に対する選択対象画像の類似スコア（例えば上記した類似度）を、複数のキーポイントの相対位置を用いて算出する。この際、画像選択部620は、複数のキーポイントそれぞれの類似スコアを算出し、これら複数の類似スコアを用いた演算処理、例えば和の算出を行うが、この演算において、上記した重み付けを用いた処理、例えば加重平均を行う。

[0070] 画像選択部620は、複数のキーポイントのすべてについて重み付けを決定するのが好ましい。この処理において、一部のキーポイントの重み付けの値（例えば係数）は標準値（デフォルト値）から変更されない場合もある。また、一部のキーポイントの重み付けは0になることもある。

[0071] ここで画像選択部620は、基準姿勢情報における当該キーポイントの基準相対位置と、クエリ情報における当該キーポイントの相対位置と、の差が大きくなると、当該キーポイントの重み付けを大きくする。このようにする理由は、上記した差が大きいキーポイントは、クエリ情報が示す姿勢を特徴づけていると推定されるためである。

[0072] ここで、重み付けを決定する際に用いられる各キーポイントの相対位置は、例えば上記したボーンリンク、すなわち人体の構造上隣に位置するキーポイントを基準とした位置で示されてもよい。またこの相対位置は、基準とな

るキーポイント（以下、基準キーポイントと記載）を少なくとも一つ設定し、この基準キーポイントを基準とした位置として示されてもよい。後者の場合、基準キーポイントは、例えば首、右肩、及び左肩の少なくとも一つである。そしてキーポイントの相対位置は、この基準キーポイントを原点とした当該キーポイントの座標で示されてもよいし、基準キーポイントから当該キーポイントに至るまでのボーンリンクで示されてもよい。

[0073] 本図に示す例において、画像選択部620が画像を選択する際の母集団となる複数の画像（以下、選択対象画像と記載）は、画像記憶部630に記憶されている。画像記憶部630に記憶されている選択対象画像は、繰り返し更新される。この更新は、選択対象画像の追加、及び選択対象画像の削除の双方が含まれるが、一般的に、画像記憶部630が記憶している選択対象画像の数は、時間が経過するにつれて増えていく。また、本図に示す例において、画像記憶部630は検索部105すなわち画像処理装置10の一部となっている。ただし画像記憶部630は、画像処理装置10の外部に位置していてもよい。なお、画像記憶部630は、上記したデータベース110の一部であってもよいし、データベース110とは別に設けられていてもよい。

[0074] ここで、基準姿勢情報について説明する。上記したように、基準姿勢情報はキーポイントの重み付けを決定する際に用いられる。基準姿勢情報は、予め定められていてもよいし、画像処理装置10のユーザからの入力に従って画像選択部620が取得してもよいし、画像選択部620が生成してもよい。

[0075] 基準姿勢情報が予め定められている場合、基準姿勢情報が示す姿勢は、例えば直立である。この場合、基準姿勢情報は、人が直立し、かつ両手を体に沿って下げたときの複数のキーポイントの相対位置を示している。基準姿勢情報は、例えば画像記憶部630に記憶されている。ただし、基準姿勢情報は画像記憶部630とは異なる場所に記憶されていてもよい。

[0076] また、画像選択部620がユーザからの入力に従って基準姿勢情報を取得する場合、ユーザから入力される情報は、基準姿勢情報そのものであっても

よいし、予め記憶されている複数の姿勢情報から基準姿勢情報として用いる情報を選択することを示してもよい。後者の例において、複数の姿勢情報のそれぞれは、互いに異なる姿勢に対応しており、かつ当該姿勢における複数のキーポイントの相対位置を含んでいる。なお、ここで用いられる複数の姿勢情報は、画像記憶部630に記憶されていてもよいし、画像記憶部630とは異なる場所に記憶されていてもよい。

[0077] また、画像選択部620は、基準姿勢情報を生成する場合、例えば画像記憶部630が記憶している複数の選択対象画像を処理することにより基準姿勢情報を生成してもよい。ここで行われる処理は、例えば、少なくとも2つの選択対象画像それぞれの複数のキーポイントの相対位置を統計処理することである。ここで行われる統計処理は例えば平均であるが、これに限定されない。なお、画像選択部620は、基準姿勢情報を生成する際に、画像記憶部630が記憶しているすべての選択対象画像を用いてもよい。

[0078] 図41(A)は、基準姿勢情報の一例を示す図であり、図41(B)は、クエリ情報の一例を示す図である。図41(A)に示す例において、基本姿勢情報が示す姿勢は直立である。一方、図41(B)に示すクエリ情報は、左手を上げている。この場合、クエリ情報と基準姿勢情報との差は、左手のキーポイントやボーンリンクが最も大きい。このため、図41(B)に示すクエリ情報を用いて目的画像を選択する場合、画像選択部620は、左手のキーポイントやボーンリンクの重み付けの値を、他のキーポイントやボーンリンクと比べて大きくする。

[0079] なお、クエリ情報が一部のキーポイントを含んでいないことがある。このような例としては、例えばクエリとして用いられる画像において人体の一部が欠損している場合である。画像選択部620は、このようなクエリ情報を用いる場合、欠損しているキーポイントの重み付けの値を0にするのが好ましい。

[0080] 図42は、本検索方法において検索部105が行う処理の第1例を示すフローチャートである。本図に示す例において、画像選択部620は、予め設

定された基準姿勢情報を用いる。

- [0081] まずクエリ取得部610は、クエリ情報を取得する（ステップS300）。すると画像選択部620は、基準姿勢情報を例えば画像記憶部630から読み出し、この基準姿勢情報とクエリ情報との差を用いて、複数のキーポイントそれぞれの重み付けを決定する（ステップS310）。
- [0082] そして画像選択部620は、画像記憶部630が記憶している複数の選択対象画像から、クエリ情報に類似する画像を選択する（ステップS320）。この際、上記したように、画像選択部620は、複数の選択対象画像毎に、クエリ情報に対する類似スコアを算出するが、この類似スコアを算出する際に、ステップS302で算出した重み付けを用いる。そして画像選択部620は類似スコアが基準値以上になった画像を、目的画像として選択する。ただし画像選択部620は、類似スコアが大きい順に選択対象画像を並べ、上位から数えて基準数までの選択対象画像を目的画像として選択してもよい。
- [0083] そして画像選択部620は、選択した目的画像を、例えば表示部107に表示させるために出力する（ステップS330）。
- [0084] 図43は、本検索方法において検索部105が行う処理の第2例を示すフローチャートである。本図に示す例は、画像選択部620が基準姿勢情報を生成する点を除いて、図42に示した例と同様である。
- [0085] 詳細には、クエリ取得部610がクエリ情報を取得する（ステップS300）と、画像選択部620は、画像記憶部630が記憶している選択対象画像を用いて、基準姿勢情報を生成する（ステップS302）。この生成処理の具体例は、上記した通りである。それ以降の処理（ステップS310～S330）は、図42を用いて説明した通りである。
- [0086] なお、上記したように、画像記憶部630が記憶している選択対象画像の数は時間が経過するにつれて増える場合が多い。このため、基準姿勢画像を生成するときに画像記憶部630が記憶しているすべての選択対象画像が使用され、かつ目的画像を選択するたびに基準姿勢情報が生成されるようにす

ると、基準姿勢情報を生成するとき用いられる選択対象画像の数はそのタイミングにおける最大数となる。従って、基準姿勢情報の標準度合いは高くなる。言い換えると基準姿勢情報が示す姿勢の特徴は少なくなる。

[0087] また、ステップS302に示した処理は、ステップS300の前に行われてもよい。

[0088] 図44は、本検索方法において検索部105が行う処理の第3例を示すフローチャートである。本図に示す例は、画像選択部620がユーザの入力に従って基準姿勢情報を取得する点を除いて、図42に示した例と同様である。

[0089] 詳細には、クエリ取得部610がクエリ情報を取得する（ステップS300）と、画像処理装置10のユーザは、画像処理装置10に、基準姿勢情報、又はあらかじめ記憶されている複数の姿勢情報のうち基準姿勢情報となる情報を特定する情報を、入力する。すると画像選択部620は、この入力に従って基準姿勢情報を取得する（ステップS304）。それ以降の処理（ステップS310～S330）は、図42を用いて説明した通りである。

[0090] なお、ステップS304に示した処理は、ステップS300の前に行われてもよい。

[0091] 以上のように、本実施の形態では、2次元画像から人物の骨格構造を検出し、検出した骨格構造の特徴量に基づいて分類や検索を行うことを可能とした。これにより、類似度が高い似た姿勢ごとに分類することができ、また、検索クエリ（検索キー）と類似度が高い似た姿勢を検索することができる。画像から似ている姿勢を分類し表示することで、ユーザが姿勢等を指定することなく、画像中の人物の姿勢を把握することができる。分類結果の中からユーザが検索クエリの姿勢を指定できるため、予めユーザが検索したい姿勢を詳細に把握していない場合でも、所望の姿勢を検索することができる。例えば、人物の骨格構造の全体や一部等を条件として分類や検索を行うことができるため、柔軟な分類や検索が可能となる。

[0092] また、検索方法6によれば、クエリ情報を用いて画像を選択する際に、ク

エリ情報と基準姿勢情報との差を用いて、キーポイントの重み付けを決定する。このようにすると、クエリ情報に含まれる複数のキーポイントのうち、クエリ情報が示す姿勢を特徴づけるキーポイントの重みを大きくすることができる。従って、人の姿勢を示す姿勢情報を用いて画像を選択する場合において、選択精度は上がる。また、画像処理装置10のユーザの負荷も小さい。

[0093] (実施の形態2) 以下、図面を参照して実施の形態2について説明する。本実施の形態では、実施の形態1における特徴量算出の具体例について説明する。本実施の形態では、人物の身長を用いて正規化することで特徴量を求める。その他については、実施の形態1と同様である。

[0094] 図18は、本実施の形態に係る画像処理装置100の構成を示している。図18に示すように、画像処理装置100は、実施の形態1の構成に加えて、さらに身長算出部108を備える。なお、特徴量算出部103と身長算出部108を一つの処理部としてもよい。

[0095] 身長算出部(身長推定部)108は、骨格構造検出部102により検出された2次元の骨格構造に基づき、2次元の画像内の人物の直立時の高さ(身長画素数という)を算出(推定)する。身長画素数は、2次元の画像における人物の身長(2次元画像空間上の人物の全身の長さ)であるとも言える。身長算出部108は、検出された骨格構造の各ボーンの長さ(2次元画像空間上の長さ)から身長画素数(ピクセル数)を求める。

[0096] 以下の例では、身長画素数を求める方法として具体例1~3を用いる。なお、具体例1~3のいずれかの方法を用いてもよいし、任意に選択される複数の方法を組み合わせて用いてもよい。具体例1では、骨格構造の各ボーンのうち、頭部から足部までのボーンの長さを合計することで、身長画素数を求める。骨格構造検出部102(骨格推定技術)が頭頂と足元を出力しない場合は、必要に応じて定数を乗じて補正することもできる。具体例2では、各ボーンの長さ(2次元画像空間上の身長)との関係を示す人体モデルを用いて、身長画素数を算出する。具体例3では、3次元人体モデ

ルを2次元骨格構造にフィッティング（あてはめる）することで、身長画素数を算出する。

[0097] 本実施の形態の特徴量算出部103は、算出された人物の身長画素数に基づいて、人物の骨格構造（骨格情報）を正規化する正規化部である。特徴量算出部103は、正規化した骨格構造の特徴量（正規化値）をデータベース110に格納する。特徴量算出部103は、骨格構造に含まれる各キーポイント（特徴点）の画像上での高さを、身長画素数で正規化する。本実施の形態では、例えば、高さ方向は、画像の2次元座標（X-Y座標）空間における上下の方向（Y軸方向）である。この場合、キーポイントの高さは、キーポイントのY座標の値（画素数）から求めることができる。あるいは、高さ方向は、実世界の3次元座標空間における地面（基準面）に対し垂直な鉛直軸の方向を、2次元座標空間に投影した鉛直投影軸の方向（鉛直投影方向）でもよい。この場合、キーポイントの高さは、実世界における地面に対し垂直な軸を、カメラパラメータに基づいて2次元座標空間に投影した鉛直投影軸を求め、この鉛直投影軸に沿った値（画素数）から求めることができる。なお、カメラパラメータは、画像の撮像パラメータであり、例えば、カメラパラメータは、カメラ200の姿勢、位置、撮像角度、焦点距離等である。カメラ200により、予め長さや位置が分かっている物体を撮像し、その画像からカメラパラメータを求めることができる。撮像された画像の両端ではひずみが発生し、実世界の鉛直方向と画像の上下方向が合わない場合がある。これに対し、画像を撮影したカメラのパラメータを使用することで、実世界の鉛直方向が画像中でどの程度傾いているのかが分かる。このため、カメラパラメータに基づいて画像中に投影した鉛直投影軸に沿ったキーポイントの値を身長で正規化することで、実世界と画像のずれを考慮してキーポイントを特徴量化することができる。なお、左右方向（横方向）は、画像の2次元座標（X-Y座標）空間における左右の方向（X軸方向）であり、または、実世界の3次元座標空間における地面に対し平行な方向を、2次元座標空間に投影した方向である。

- [0098] 図19～図23は、本実施の形態に係る画像処理装置100の動作を示している。図19は、画像処理装置100における画像取得から検索処理までの流れを示し、図20～図22は、図19の身長画素数算出処理(S201)の具体例1～3の流れを示し、図23は、図19の正規化処理(S202)の流れを示している。
- [0099] 図19に示すように、本実施の形態では、実施の形態1における特徴量算出処理(S103)として、身長画素数算出処理(S201)及び正規化処理(S202)を行う。その他については実施の形態1と同様である。
- [0100] 画像処理装置100は、画像取得(S101)及び骨格構造検出(S102)に続いて、検出された骨格構造に基づいて身長画素数算出処理を行う(S201)。この例では、図24に示すように、画像における直立時の人物の骨格構造の高さを身長画素数(h)とし、画像の人物の状態における骨格構造の各キーポイントの高さをキーポイント高さ( $y_i$ )とする。以下、身長画素数算出処理の具体例1～3について説明する。
- [0101] <具体例1>具体例1では、頭部から足部までのボーンの長さをを用いて身長画素数を求める。具体例1では、図20に示すように、身長算出部108は、各ボーンの長さを取得し(S211)、取得した各ボーンの長さを合計する(S212)。
- [0102] 身長算出部108は、人物の頭部から足部の2次元の画像上のボーンの長さを取得し、身長画素数を求める。すなわち、骨格構造を検出した画像から、図24のボーンのうち、ボーンB1(長さL1)、ボーンB51(長さL21)、ボーンB61(長さL31)及びボーンB71(長さL41)、もしくは、ボーンB1(長さL1)、ボーンB52(長さL22)、ボーンB62(長さL32)及びボーンB72(長さL42)の各長さ(画素数)を取得する。各ボーンの長さは、2次元の画像における各キーポイントの座標から求めることができる。これらを合計した、 $L1+L21+L31+L41$ 、もしくは、 $L1+L22+L32+L42$ に補正定数を乗じた値を身長画素数(h)として算出する。両方の値を算出できる場合、例えば、長い方

の値を身長画素数とする。すなわち、各ボーンは正面から撮像された場合が画像中での長さが最も長くなり、カメラに対して奥行き方向に傾くと短く表示される。従って、長いボーンの方が正面から撮像されている可能性が高く、真実の値に近いと考えられる。このため、長い方の値を選択することが好ましい。

[0103] 図25の例では、ボーンB1、ボーンB51及びボーンB52、ボーンB61及びボーンB62、ボーンB71及びボーンB72がそれぞれ重ならずに出検されている。これらのボーンの合計である、 $L1 + L21 + L31 + L41$ 、及び、 $L1 + L22 + L32 + L42$ を求め、例えば、検出されたボーンの長さが長い左足側の $L1 + L22 + L32 + L42$ に補正定数を乗じた値を身長画素数とする。

[0104] 図26の例では、ボーンB1、ボーンB51及びボーンB52、ボーンB61及びボーンB62、ボーンB71及びボーンB72がそれぞれ検出され、右足のボーンB61及びボーンB71と左足のボーンB62及びボーンB72が重なっている。これらのボーンの合計である、 $L1 + L21 + L31 + L41$ 、及び、 $L1 + L22 + L32 + L42$ を求め、例えば、検出されたボーンの長さが長い右足側の $L1 + L21 + L31 + L41$ に補正定数を乗じた値を身長画素数とする。

[0105] 図27の例では、ボーンB1、ボーンB51及びボーンB52、ボーンB61及びボーンB62、ボーンB71及びボーンB72がそれぞれ検出され、右足のボーンB61及びボーンB71と左足のボーンB62及びボーンB72が重なっている。これらのボーンの合計である、 $L1 + L21 + L31 + L41$ 、及び、 $L1 + L22 + L32 + L42$ を求め、例えば、検出されたボーンの長さが長い左足側の $L1 + L22 + L32 + L42$ に補正定数を乗じた値を身長画素数とする。

[0106] 具体例1では、頭から足までのボーンの長さを合計することで身長を求めることができるため、簡易な方法で身長画素数を求めることができる。また、機械学習を用いた骨格推定技術により、少なくとも頭から足までの骨格を

検出できればよいため、しゃがみ込んでいる状態など、必ずしも人物の全体が画像に写っていない場合でも精度よく身長画素数を推定することができる。

[0107] <具体例2>具体例2では、2次元骨格構造に含まれる骨の長さとの関係を示す2次元骨格モデルを用いて身長画素数を求める。

[0108] 図28は、具体例2で用いる、2次元画像空間上の各ボーンの長さとの関係を示す人体モデル（2次元骨格モデル）301である。図28に示すように、平均的な人物の各ボーンの長さとの関係（全身の長さに対する各ボーンの長さの割合）を、人体モデル301の各ボーンに対応付ける。例えば、頭のボーンB1の長さは全身の長さ×0.2（20%）であり、右手のボーンB41の長さは全身の長さ×0.15（15%）であり、右足のボーンB71の長さは全身の長さ×0.25（25%）である。このような人体モデル301の情報をデータベース110に記憶しておくことで、各ボーンの長さから平均的な全身の長さを求めることができる。平均的な人物の人体モデルの他に、年代、性別、国籍等の人物の属性ごとに人体モデルを用意してもよい。これにより、人物の属性に応じて適切に全身の長さ（身長）を求めることができる。

[0109] 具体例2では、図21に示すように、身長算出部108は、各ボーンの長さを取得する（S221）。身長算出部108は、検出された骨格構造において、全てのボーンの長さ（2次元画像空間上の長さ）を取得する。図29は、しゃがみ込んでいる状態の人物を右斜め後ろから撮像し、骨格構造を検出した例である。この例では、人物の顔や左側面が写っていないことから、頭のボーンと左腕及び左手のボーンが検出できていない。このため、検出されているボーンB21、B22、B31、B41、B51、B52、B61、B62、B71、B72の各長さを取得する。

[0110] 続いて、身長算出部108は、図21に示すように、人体モデルに基づき、各ボーンの長さから身長画素数を算出する（S222）。身長算出部10

8は、図28のような、各ボーンと全身の長さとの関係を示す人体モデル301を参照し、各ボーンの長さから身長画素数を求める。例えば、右手のボーンB41の長さが全身の長さ $\times 0.15$ であるため、ボーンB41の長さ $\div 0.15$ によりボーンB41に基づいた身長画素数を求める。また、右足のボーンB71の長さが全身の長さ $\times 0.25$ であるため、ボーンB71の長さ $\div 0.25$ によりボーンB71に基づいた身長画素数を求める。

[0111] このとき参照する人体モデルは、例えば、平均的な人物の人体モデルであるが、年代、性別、国籍等の人物の属性に応じて人体モデルを選択してもよい。例えば、撮像した画像に人物の顔が写っている場合、顔に基づいて人物の属性を識別し、識別した属性に対応する人体モデルを参照する。属性ごとの顔を機械学習した情報を参照し、画像の顔の特徴から人物の属性を認識することができる。また、画像から人物の属性が識別できない場合に、平均的な人物の人体モデルを用いてもよい。

[0112] また、ボーンの長さから算出した身長画素数をカメラパラメータにより補正してもよい。例えばカメラを高い位置において、人物を見下ろすように撮影した場合、二次元骨格構造において肩幅のボーン等の横の長さはカメラの俯角の影響を受けないが、首-腰のボーン等の縦の長さは、カメラの俯角が大きくなる程小さくなる。そうすると、肩幅のボーン等の横の長さから算出した身長画素数が実際より大きくなる傾向がある。そこで、カメラパラメータを活用すると、人物がどの程度の角度でカメラに見下ろされているかがわかるため、この俯角の情報を使って正面から撮影したような二次元骨格構造に補正することができる。これによって、より正確に身長画素数を算出できる。

[0113] 続いて、身長算出部108は、図21に示すように、身長画素数の最適値を算出する(S223)。身長算出部108は、ボーンごとに求めた身長画素数から身長画素数の最適値を算出する。例えば、図30に示すような、ボーンごとに求めた身長画素数のヒストグラムを生成し、その中で大きい身長画素数を選択する。つまり、複数のボーンに基づいて求められた複数の身長

画素数の中で他よりも長い身長画素数を選択する。例えば、上位30%を有効な値とし、図30ではボーンB71、B61、B51による身長画素数を選択する。選択した身長画素数の平均を最適値として求めてもよいし、最も大きい身長画素数を最適値としてもよい。2次元画像のボーンの長さから身長を求めるため、ボーンを正面からできていない場合、すなわち、ボーンがカメラから見て奥行き方向に傾いて撮像された場合、ボーンの長さが正面から撮像した場合よりも短くなる。そうすると、身長画素数が大きい値は、身長画素数が小さい値よりも、正面から撮像された可能性が高く、より尤もらしい値となることから、より大きい値を最適値とする。

[0114] 具体例2では、2次元画像空間上のボーンと全身の長さとの関係を示す人体モデルを用いて、検出した骨格構造のボーンに基づき身長画素数を求めるため、頭から足までの全ての骨格が得られない場合でも、一部のボーンから身長画素数を求めることができる。特に、複数のボーンから求められた値のうち、より大きい値を採用することで、精度よく身長画素数を推定することができる。

[0115] <具体例3>具体例3では、2次元骨格構造を3次元人体モデル（3次元骨格モデル）にフィッティングさせて、フィッティングした3次元人体モデルの身長画素数を用いて全身の骨格ベクトルを求める。

[0116] 具体例3では、図22に示すように、身長算出部108は、まず、カメラ200の撮像した画像に基づき、カメラパラメータを算出する（S231）。身長算出部108は、カメラ200が撮像した複数の画像の中から、予め長さが分かっている物体を抽出し、抽出した物体の大きさ（画素数）からカメラパラメータを求める。なお、カメラパラメータを予め求めておき、求めておいたカメラパラメータを必要に応じて取得してもよい。

[0117] 続いて、身長算出部108は、3次元人体モデルの配置及び高さを調整する（S232）。身長算出部108は、検出された2次元骨格構造に対し、身長画素数算出用の3次元人体モデルを用意し、カメラパラメータに基づいて、同じ2次元画像内に配置する。具体的には、カメラパラメータと、2次

元骨格構造から、「実世界におけるカメラと人物の相対的な位置関係」を特定する。例えば、仮にカメラの位置を座標(0, 0, 0)としたときに、人物が立っている(または座っている)位置の座標(x, y, z)を特定する。そして、特定した人物と同じ位置(x, y, z)に3次元人体モデルを配置して撮像した場合の画像を想定することで、2次元骨格構造と3次元人体モデルを重ね合わせる。

[0118] 図31は、しゃがみ込んでいる人物を左斜め前から撮像し、2次元骨格構造401を検出した例である。2次元骨格構造401は、2次元の座標情報を有する。なお、全てのボーンを検出していることが好ましいが、一部のボーンが検出されていなくてもよい。この2次元骨格構造401に対し、図32のような、3次元人体モデル402を用意する。3次元人体モデル(3次元骨格モデル)402は、3次元の座標情報を有し、2次元骨格構造401と同じ形状の骨格のモデルである。そして、図33のように、検出した2次元骨格構造401に対し、用意した3次元人体モデル402を配置し重ね合わせる。また、重ね合わせるとともに、3次元人体モデル402の高さを2次元骨格構造401に合うように調整する。

[0119] なお、このとき用意する3次元人体モデル402は、図33のように、2次元骨格構造401の姿勢に近い状態のモデルでもよいし、直立した状態のモデルでもよい。例えば、機械学習を用いて2次元画像から3次元空間の姿勢を推定する技術を用いて、推定した姿勢の3次元人体モデル402を生成してもよい。2次元画像の関節と3次元空間の関節の情報を学習することで、2次元画像から3次元の姿勢を推定することができる。

[0120] 続いて、身長算出部108は、図22に示すように、3次元人体モデルを2次元骨格構造にフィッティングする(S233)。身長算出部108は、図34のように、3次元人体モデル402を2次元骨格構造401に重ね合わせた状態で、3次元人体モデル402と2次元骨格構造401の姿勢が一致するように、3次元人体モデル402を変形させる。すなわち、3次元人体モデル402の身長、体の向き、関節の角度を調整し、2次元骨格構造4

01との差異がなくなるように最適化する。例えば、3次元人体モデル402の関節を、人の可動範囲で回転させていき、また、3次元人体モデル402の全体を回転させたり、全体のサイズを調整する。なお、3次元人体モデルと2次元骨格構造のフィッティング（あてはめ）は、2次元空間（2次元座標）上で行う。すなわち、2次元空間に3次元人体モデルを写像し、変形させた3次元人体モデルが2次元空間（画像）でどのように変化するかを考慮して、3次元人体モデルを2次元骨格構造に最適化する。

[0121] 続いて、身長算出部108は、図22に示すように、フィッティングさせた3次元人体モデルの身長画素数を算出する（S234）。身長算出部108は、図35のように、3次元人体モデル402と2次元骨格構造401の差異がなくなり、姿勢が一致すると、その状態の3次元人体モデル402の身長画素数を求める。最適化された3次元人体モデル402を直立させた状態として、カメラパラメータに基づき、2次元空間上の全身の長さを求める。例えば、3次元人体モデル402を直立させた場合の頭から足までのボーンの長さ（画素数）により身長画素数を算出する。具体例1と同様に、3次元人体モデル402の頭部から足部までのボーンの長さを合計してもよい。

[0122] 具体例3では、カメラパラメータに基づいて3次元人体モデルを2次元骨格構造にフィッティングさせて、その3次元人体モデルに基づいて身長画素数を求めることで、全てのボーンが正面に写っていない場合、すなわち、全てのボーンが斜めに映っているため誤差が大きい場合でも、精度よく身長画素数を推定することができる。

[0123] <正規化処理>図19に示すように、画像処理装置100は、身長画素数算出処理に続いて、正規化処理（S202）を行う。正規化処理では、図23に示すように、特徴量算出部103は、キーポイント高さを算出する（S241）。特徴量算出部103は、検出された骨格構造に含まれる全てのキーポイントのキーポイント高さ（画素数）を算出する。キーポイント高さは、骨格構造の最下端（例えばいずれかの足のキーポイント）からそのキーポイントまでの高さ方向の長さ（画素数）である。ここでは、一例として、キ

ーポイント高さを、画像におけるキーポイントのY座標から求める。なお、上記のように、キーポイント高さは、カメラパラメータに基づいた鉛直投影軸に沿った方向の長さから求めてもよい。例えば、図24の例で、首のキーポイントA2の高さ( $y_i$ )は、キーポイントA2のY座標から右足のキーポイントA81または左足のキーポイントA82のY座標を引いた値である。

[0124] 続いて、特徴量算出部103は、正規化のための基準点を特定する(S242)。基準点は、キーポイントの相対的な高さを表すための基準となる点である。基準点は、予め設定されていてもよいし、ユーザが選択できるようにしてもよい。基準点は、骨格構造の中心もしくは中心よりも高い(画像の上下方向における上である)ことが好ましく、例えば、首のキーポイントの座標を基準点とする。なお、首に限らず頭やその他のキーポイントの座標を基準点としてもよい。キーポイントに限らず、任意の座標(例えば骨格構造の中心座標等)を基準点としてもよい。

[0125] 続いて、特徴量算出部103は、キーポイント高さ( $y_i$ )を身長画素数で正規化する(S243)。特徴量算出部103は、各キーポイントのキーポイント高さ、基準点、身長画素数を用いて、各キーポイントを正規化する。具体的には、特徴量算出部103は、基準点に対するキーポイントの相対的な高さを身長画素数により正規化する。ここでは、高さ方向のみに着目する例として、Y座標のみを抽出し、また、基準点を首のキーポイントとして正規化を行う。具体的には、基準点(首のキーポイント)のY座標を( $y_c$ )として、次の式(1)を用いて、特徴量(正規化値)を求める。なお、カメラパラメータに基づいた鉛直投影軸を用いる場合は、( $y_i$ )及び( $y_c$ )を鉛直投影軸に沿った方向の値に変換する。

[数1]

$$f_i = (y_i - y_c) / h \quad \dots (1)$$

[0126] 例えば、キーポイントが18個の場合、各キーポイントの18点の座標(

$(x_0, y_0)$ 、 $(x_1, y_1)$ 、 $\dots$ 、 $(x_{17}, y_{17})$  を、上記式 (1) を用いて、次のように 18 次元の特徴量に変換する。

[数2]

$$\begin{aligned} f_0 &= (y_0 - y_c) / h \\ f_1 &= (y_1 - y_c) / h \\ &\vdots \\ f_{17} &= (y_{17} - y_c) / h \end{aligned} \quad \dots (2)$$

[0127] 図36は、特徴量算出部103が求めた各キーポイントの特徴量の例を示している。この例では、首のキーポイントA2を基準点とするため、キーポイントA2の特徴量は0.0となり、首と同じ高さの右肩のキーポイントA31及び左肩のキーポイントA32の特徴量も0.0である。首よりも高い頭の特徴量は-0.2である。首よりも低い右手のキーポイントA51及び左手のキーポイントA52の特徴量は0.4であり、右足のキーポイントA81及び左足のキーポイントA82の特徴量は0.9である。この状態から人物が左手を挙げると、図37のように左手が基準点よりも高くなるため、左手のキーポイントA52の特徴量は-0.4となる。一方で、Y軸の座標のみを用いて正規化を行っているため、図38のように、図36に比べて骨格構造の幅が変わっても特徴量は変わらない。すなわち、本実施の形態の特徴量（正規化値）は、骨格構造（キーポイント）の高さ方向（Y方向）の特徴を示しており、骨格構造の横方向（X方向）の変化に影響を受けない。

[0128] 以上のように、本実施の形態では、2次元画像から人物の骨格構造を検出し、検出した骨格構造から求めた身長画素数（2次元画像空間上の直立時の高さ）を用いて、骨格構造の各キーポイントを正規化する。この正規化された特徴量を用いることで、分類や検索等を行った場合のロバスト性を向上することができる。すなわち、本実施の形態の特徴量は、上記のように人物の横方向の変化に影響を受けないため、人物の向きや人物の体型の変化に対しロバスト性が高い。

[0129] さらに、本実施の形態では、OpenPose等の骨格推定技術を用いて人物の骨格構造を検出することで実現できるため、人物の姿勢等を学習する学習データを用意する必要がない。また、骨格構造のキーポイントを正規化し、データベースに格納しておくことで、人物の姿勢等の分類や検索が可能となるため、未知な姿勢に対しても分類や検索を行うことができる。また、骨格構造のキーポイントを正規化することで、明確でわかりやすい特徴量を得ることができるため、機械学習のようにブラックボックス型のアルゴリズムと異なり、処理結果に対するユーザの納得性が高い。

[0130] 以上、図面を参照して本発明の実施形態について述べたが、これらは本発明の例示であり、上記以外の様々な構成を採用することもできる。

[0131] また、上述の説明で用いた複数のフローチャートでは、複数の工程（処理）が順番に記載されているが、各実施形態で実行される工程の実行順序は、その記載の順番に制限されない。各実施形態では、図示される工程の順番を内容的に支障のない範囲で変更することができる。また、上述の各実施形態は、内容が相反しない範囲で組み合わせることができる。

[0132] 上記の実施形態の一部または全部は、以下の付記のようにも記載されうるが、以下に限られない。

1. 人の姿勢を示すクエリ情報を取得するクエリ取得手段と、

前記クエリ情報を用いて、複数の選択対象画像から少なくとも一つの目的画像を選択する画像選択手段と、

を備え、

前記クエリ情報は、人の身体の互いに異なる部分を示す複数のキーポイントの相対位置を含んでおり、

前記画像選択手段は、

前記複数のキーポイントの基準相対位置を含む基準姿勢情報と、前記クエリ情報との差分を用いて、少なくとも一つの前記キーポイントの重み付けを決定し、

前記選択対象画像に含まれる人の前記複数のキーポイントの相対位置、

前記クエリ情報、及び前記重み付けを用いて、前記少なくとも一つの目的画像を選択する、画像選択装置。

2. 上記1に記載の画像選択装置において、

前記画像選択手段は、前記複数のキーポイントのそれぞれにおいて、前記基準姿勢情報における当該キーポイントの前記基準相対位置と、前記クエリ情報における当該キーポイントの相対位置と、の差が大きくなると、当該キーポイントの前記重み付けを大きくする画像選択装置。

3. 上記1又は2に記載の画像選択装置において、

前記基準姿勢情報は、人が直立し、かつ両手を体に沿って下げたときの前記複数のキーポイントの相対位置を示す画像選択装置。

4. 上記1又は2に記載の画像選択装置において、

前記画像選択手段は、複数の選択対象画像を処理することにより、前記基準姿勢情報を生成する画像選択装置。

5. 上記4に記載の画像選択装置において、

前記画像選択手段は、少なくとも2つの前記選択対象画像それぞれの前記複数のキーポイントの相対位置を統計処理することにより、前記基準相対位置を決定する画像選択装置。

6. 上記1又は2に記載の画像選択装置において、

前記画像選択手段は、ユーザからの入力に従って前記基準姿勢情報を取得する、画像選択装置。

7. コンピュータが、

人の姿勢を示すクエリ情報を取得するクエリ取得処理と、

前記クエリ情報を用いて、複数の選択対象画像から少なくとも一つの目的画像を選択する画像選択処理と、

を行い

前記クエリ情報は、人の身体の互いに異なる部分を示す複数のキーポイントの相対位置を含んでおり、

前記画像選択処理において、前記コンピュータは、

前記複数のキーポイントの基準相対位置を含む基準姿勢情報と、前記クエリ情報との差分を用いて、少なくとも一つの前記キーポイントの重み付けを決定し、

前記選択対象画像に含まれる人の前記複数のキーポイントの相対位置、前記クエリ情報、及び前記重み付けを用いて、前記少なくとも一つの目的画像を選択する、画像選択方法。

8. 上記7に記載の画像選択方法において、

前記画像選択処理において、前記コンピュータは、前記複数のキーポイントのそれぞれにおいて、前記基準姿勢情報における当該キーポイントの前記基準相対位置と、前記クエリ情報における当該キーポイントの相対位置と、の差が大きくなると、当該キーポイントの前記重み付けを大きくする画像選択方法。

9. 上記7又は8に記載の画像選択方法において、

前記基準姿勢情報は、人が直立し、かつ両手を体に沿って下げたときの前記複数のキーポイントの相対位置を示す画像選択方法。

10. 上記7又は8に記載の画像選択方法において、

前記画像選択処理において、前記コンピュータは、複数の選択対象画像を処理することにより、前記基準姿勢情報を生成する画像選択方法。

11. 上記10に記載の画像選択方法において、

前記画像選択処理において、前記コンピュータは、少なくとも2つの前記選択対象画像それぞれの前記複数のキーポイントの相対位置を統計処理することにより、前記基準相対位置を決定する画像選択方法。

12. 上記7又は8に記載の画像選択方法において、

前記画像選択処理において、前記コンピュータは、ユーザからの入力に従って前記基準姿勢情報を取得する、画像選択方法。

13. コンピュータに、

人の姿勢を示すクエリ情報を取得するクエリ取得機能と、

前記クエリ情報を用いて、複数の選択対象画像から少なくとも一つの目

的画像を選択する画像選択機能と、  
を持たせ、

前記クエリ情報は、人の身体の互いに異なる部分を示す複数のキーポイントの相対位置を含んでおり、

前記画像選択機能は、

前記複数のキーポイントの基準相対位置を含む基準姿勢情報と、前記クエリ情報との差分を用いて、少なくとも一つの前記キーポイントの重み付けを決定し、

前記選択対象画像に含まれる人の前記複数のキーポイントの相対位置、前記クエリ情報、及び前記重み付けを用いて、前記少なくとも一つの目的画像を選択する、プログラム。

14. 上記13に記載のプログラムにおいて、

前記画像選択機能は、前記複数のキーポイントのそれぞれにおいて、前記基準姿勢情報における当該キーポイントの前記基準相対位置と、前記クエリ情報における当該キーポイントの相対位置と、の差が大きくなると、当該キーポイントの前記重み付けを大きくするプログラム。

15. 上記13又は14に記載のプログラムにおいて、

前記基準姿勢情報は、人が直立し、かつ両手を体に沿って下げたときの前記複数のキーポイントの相対位置を示すプログラム。

16. 上記13又は14に記載のプログラムにおいて、

前記画像選択機能は、複数の選択対象画像を処理することにより、前記基準姿勢情報を生成するプログラム。

17. 上記16に記載のプログラムにおいて、

前記画像選択機能は、少なくとも2つの前記選択対象画像それぞれの前記複数のキーポイントの相対位置を統計処理することにより、前記基準相対位置を決定するプログラム。

18. 上記13又は14に記載のプログラムにおいて、

前記画像選択機能は、ユーザからの入力に従って前記基準姿勢情報を取得

する、プログラム。

## 符号の説明

[0133]	1	画像処理システム
	1 0	画像処理装置（画像選択装置）
	1 1	骨格検出部
	1 2	特徴量算出部
	1 3	認識部
	1 0 0	画像処理装置（画像選択装置）
	1 0 1	画像取得部
	1 0 2	骨格構造検出部
	1 0 3	特徴量算出部
	1 0 4	分類部
	1 0 5	検索部
	1 0 6	入力部
	1 0 7	表示部
	1 0 8	身長算出部
	1 1 0	データベース
	2 0 0	カメラ
	3 0 0、3 0 1	人体モデル
	4 0 1	2次元骨格構造
	4 0 2	3次元人体モデル
	6 1 0	クエリ取得部
	6 2 0	画像選択部
	6 3 0	画像記憶部

## 請求の範囲

- [請求項1] 人の姿勢を示すクエリ情報を取得するクエリ取得手段と、  
前記クエリ情報を用いて、複数の選択対象画像から少なくとも一つの目的画像を選択する画像選択手段と、  
を備え、  
前記クエリ情報は、人の身体の互いに異なる部分を示す複数のキーポイントの相対位置を含んでおり、  
前記画像選択手段は、  
前記複数のキーポイントの基準相対位置を含む基準姿勢情報と、  
前記クエリ情報との差分を用いて、少なくとも一つの前記キーポイントの重み付けを決定し、  
前記選択対象画像に含まれる人の前記複数のキーポイントの相対位置、前記クエリ情報、及び前記重み付けを用いて、前記少なくとも一つの目的画像を選択する、画像選択装置。
- [請求項2] 請求項1に記載の画像選択装置において、  
前記画像選択手段は、前記複数のキーポイントのそれぞれにおいて、前記基準姿勢情報における当該キーポイントの前記基準相対位置と、前記クエリ情報における当該キーポイントの相対位置と、の差が大きくなると、当該キーポイントの前記重み付けを大きくする画像選択装置。
- [請求項3] 請求項1又は2に記載の画像選択装置において、  
前記基準姿勢情報は、人が直立し、かつ両手を体に沿って下げたときの前記複数のキーポイントの相対位置を示す画像選択装置。
- [請求項4] 請求項1又は2に記載の画像選択装置において、  
前記画像選択手段は、複数の選択対象画像を処理することにより、前記基準姿勢情報を生成する画像選択装置。
- [請求項5] 請求項4に記載の画像選択装置において、  
前記画像選択手段は、少なくとも2つの前記選択対象画像それぞれ

の前記複数のキーポイントの相対位置を統計処理することにより、前記基準相対位置を決定する画像選択装置。

[請求項6] 請求項1又は2に記載の画像選択装置において、  
前記画像選択手段は、ユーザからの入力に従って前記基準姿勢情報を取得する、画像選択装置。

[請求項7] コンピュータが、  
人の姿勢を示すクエリ情報を取得するクエリ取得処理と、  
前記クエリ情報を用いて、複数の選択対象画像から少なくとも一つの目的画像を選択する画像選択処理と、  
を行い

前記クエリ情報は、人の身体の互いに異なる部分を示す複数のキーポイントの相対位置を含んでおり、

前記画像選択処理において、前記コンピュータは、

前記複数のキーポイントの基準相対位置を含む基準姿勢情報と、  
前記クエリ情報との差分を用いて、少なくとも一つの前記キーポイントの重み付けを決定し、

前記選択対象画像に含まれる人の前記複数のキーポイントの相対位置、前記クエリ情報、及び前記重み付けを用いて、前記少なくとも一つの目的画像を選択する、画像選択方法。

[請求項8] コンピュータに、  
人の姿勢を示すクエリ情報を取得するクエリ取得機能と、  
前記クエリ情報を用いて、複数の選択対象画像から少なくとも一つの目的画像を選択する画像選択機能と、  
を持たせ、

前記クエリ情報は、人の身体の互いに異なる部分を示す複数のキーポイントの相対位置を含んでおり、

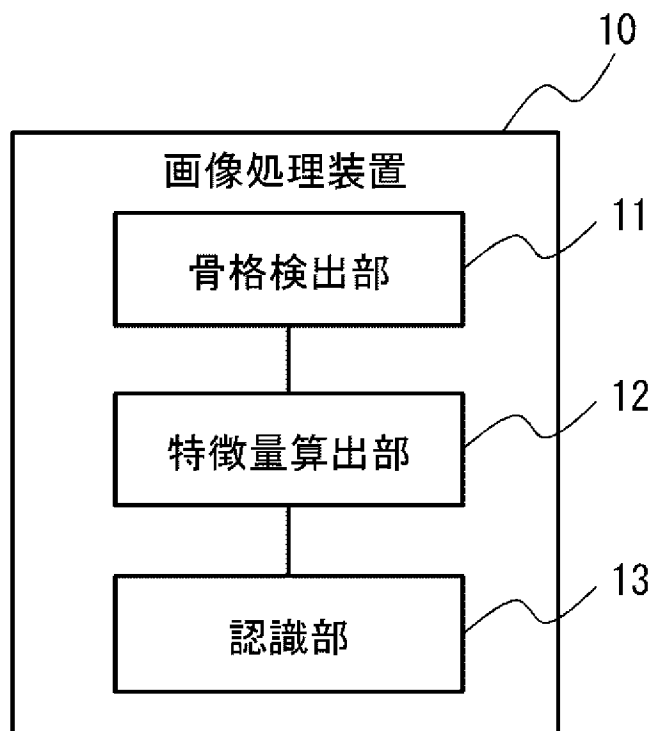
前記画像選択機能は、

前記複数のキーポイントの基準相対位置を含む基準姿勢情報と、

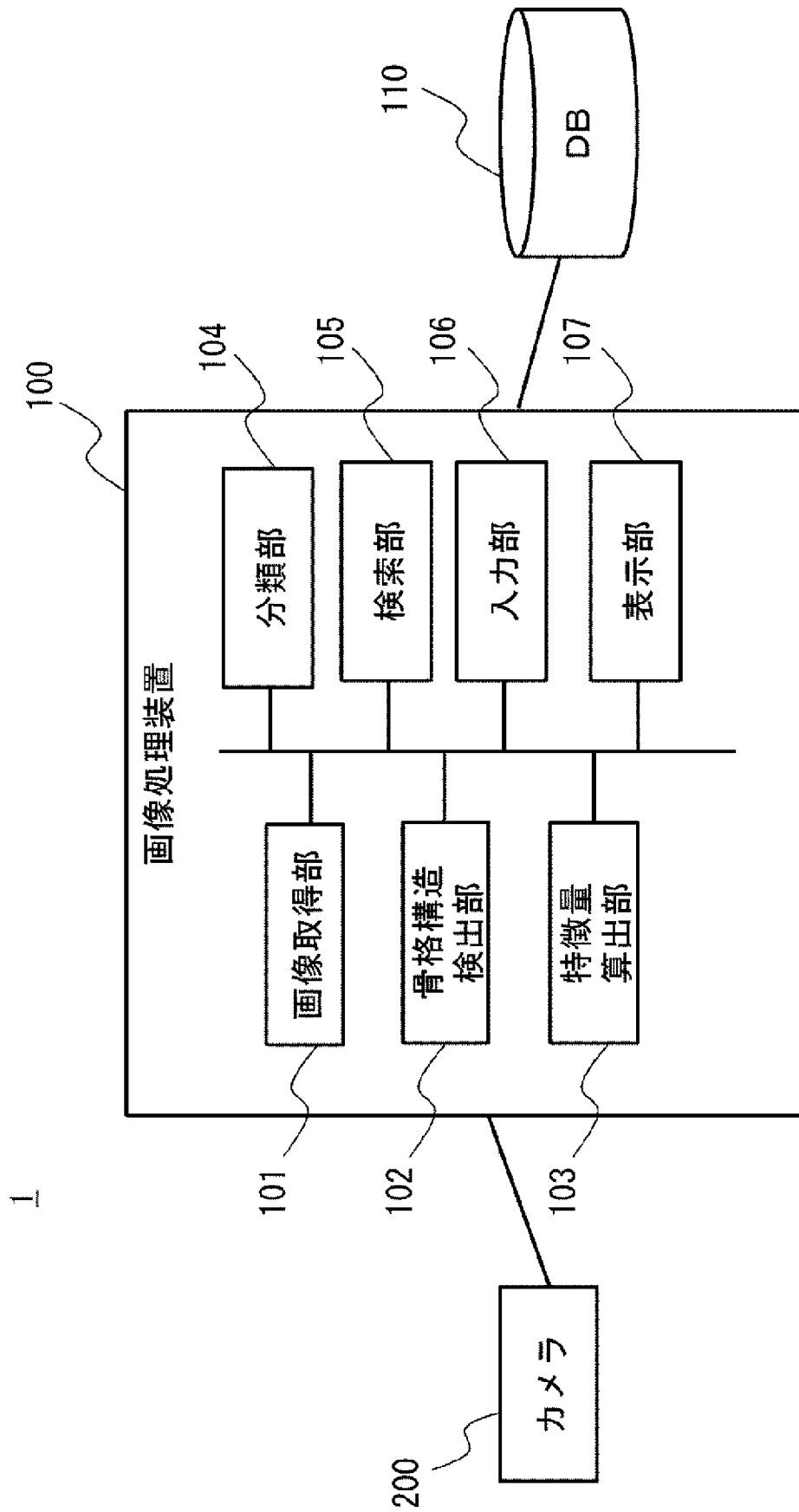
前記クエリ情報との差分を用いて、少なくとも一つの前記キーポイントの重み付けを決定し、

前記選択対象画像に含まれる人の前記複数のキーポイントの相対位置、前記クエリ情報、及び前記重み付けを用いて、前記少なくとも一つの目的画像を選択する、プログラム。

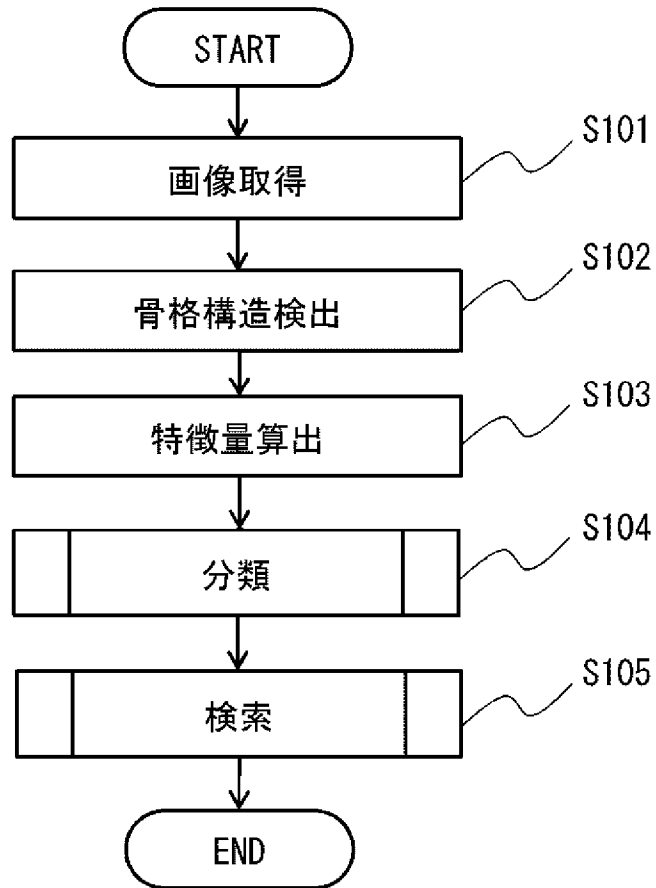
[図1]



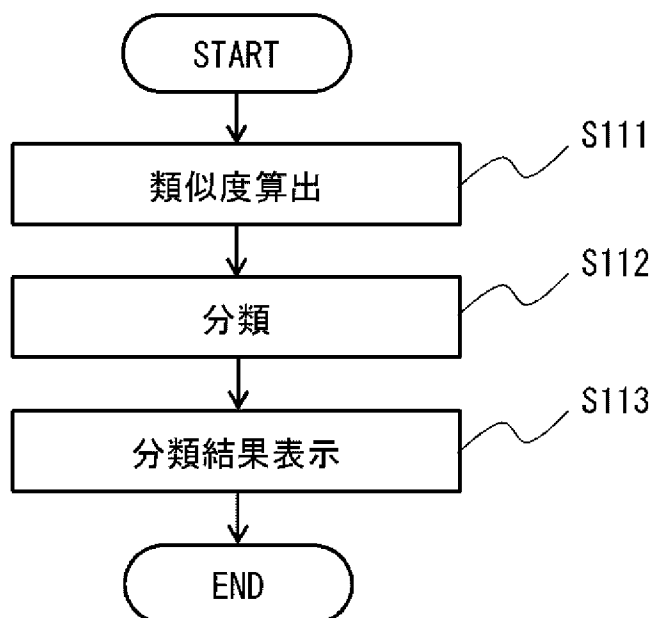
[図2]



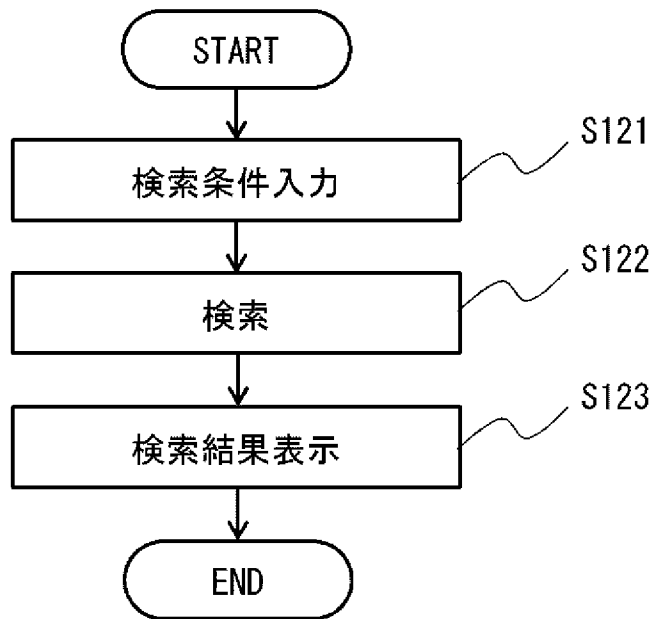
[図3]



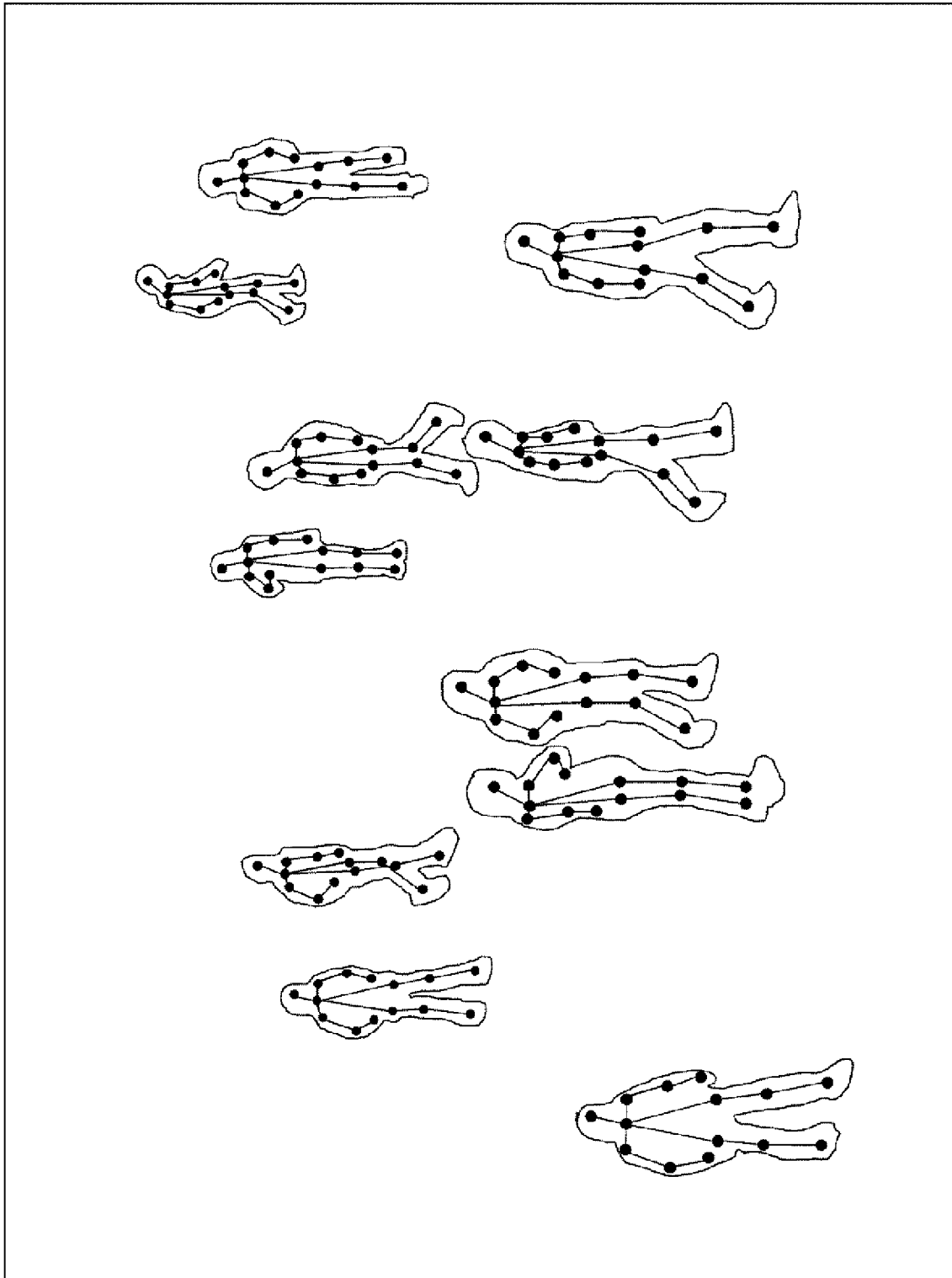
[図4]



[図5]

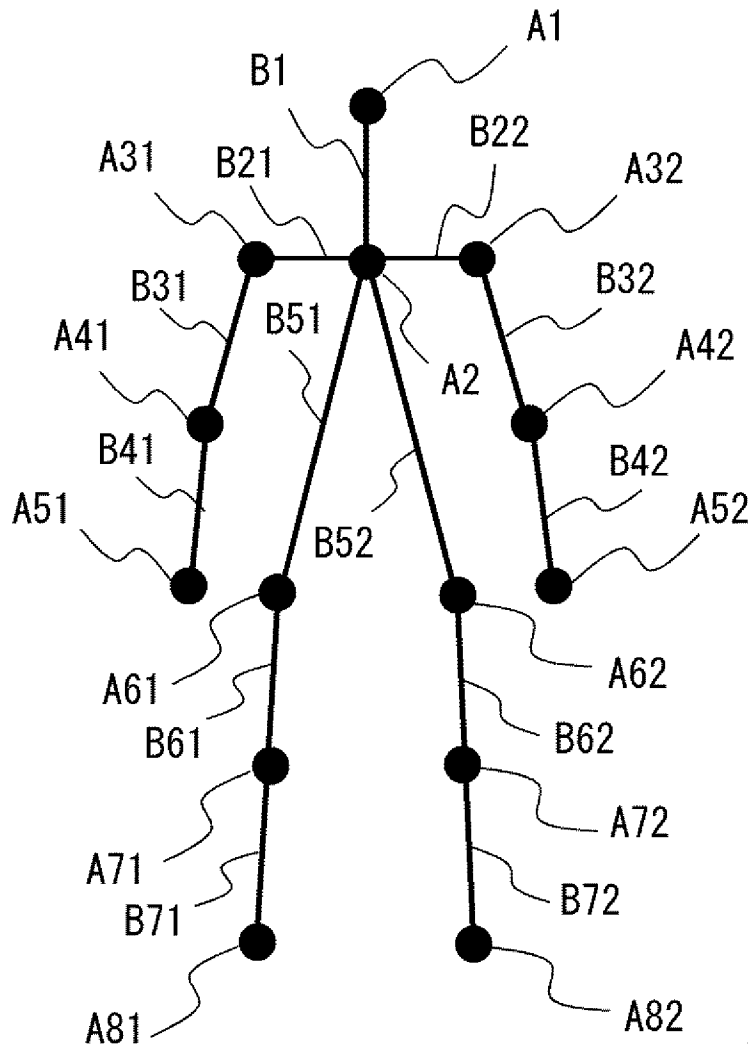


[図6]

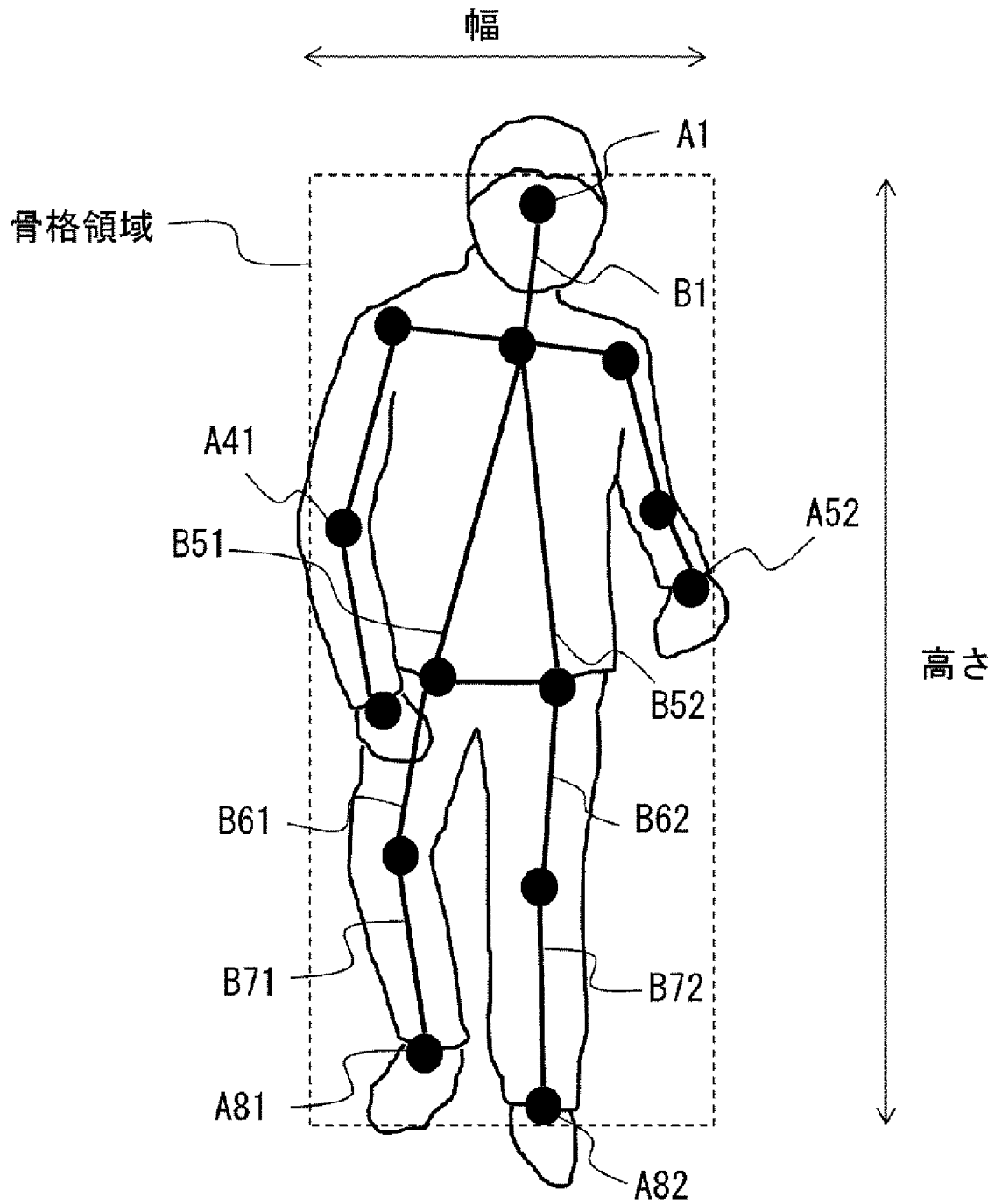


[図7]

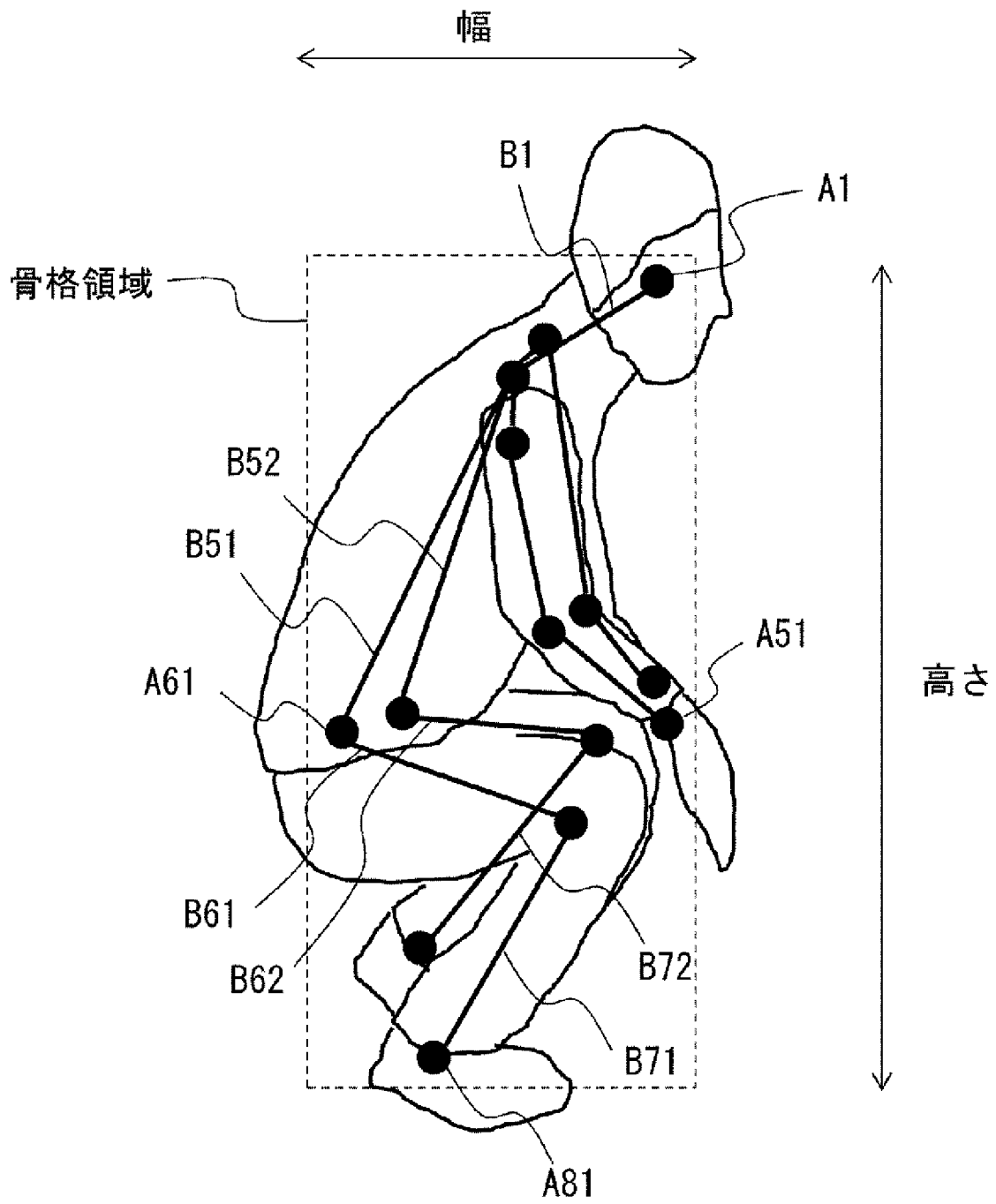
300



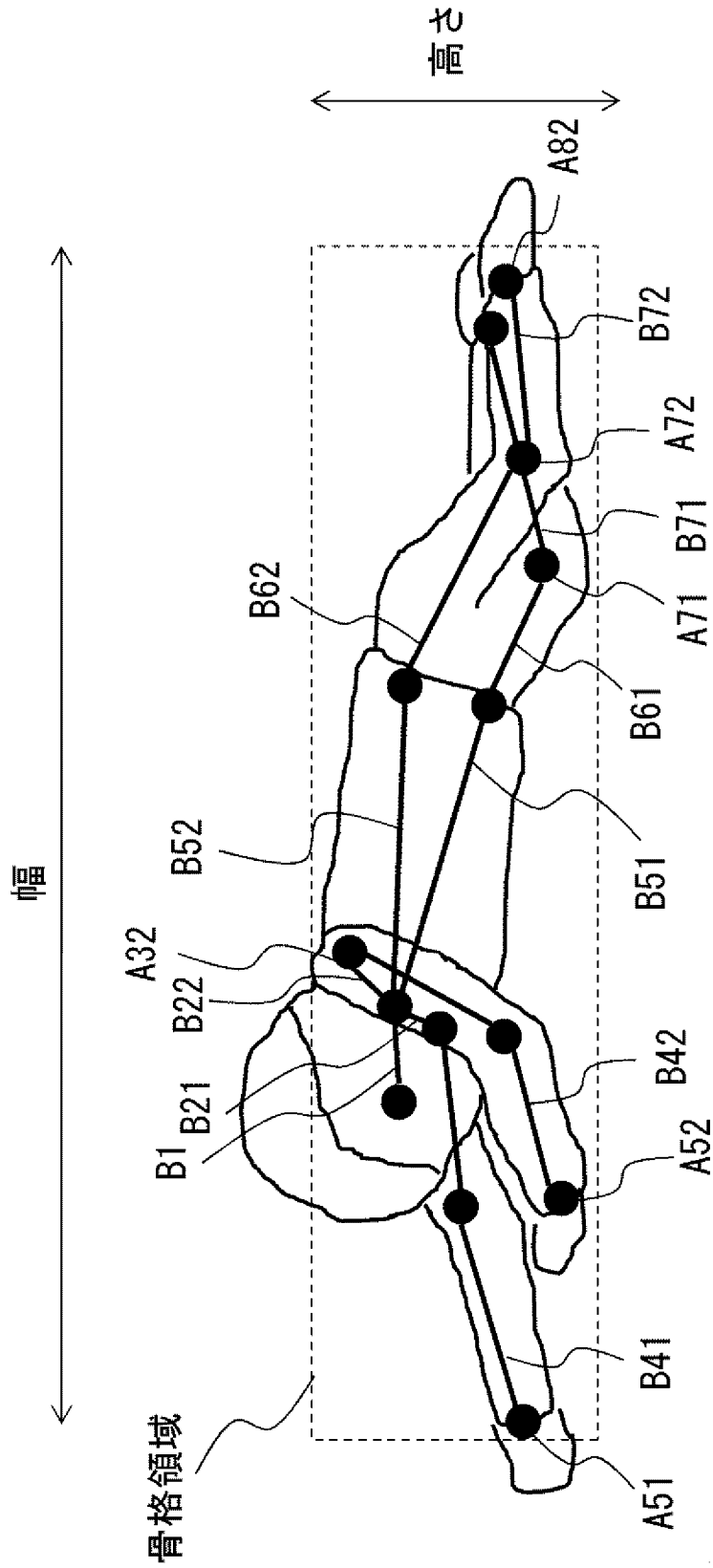
[図8]



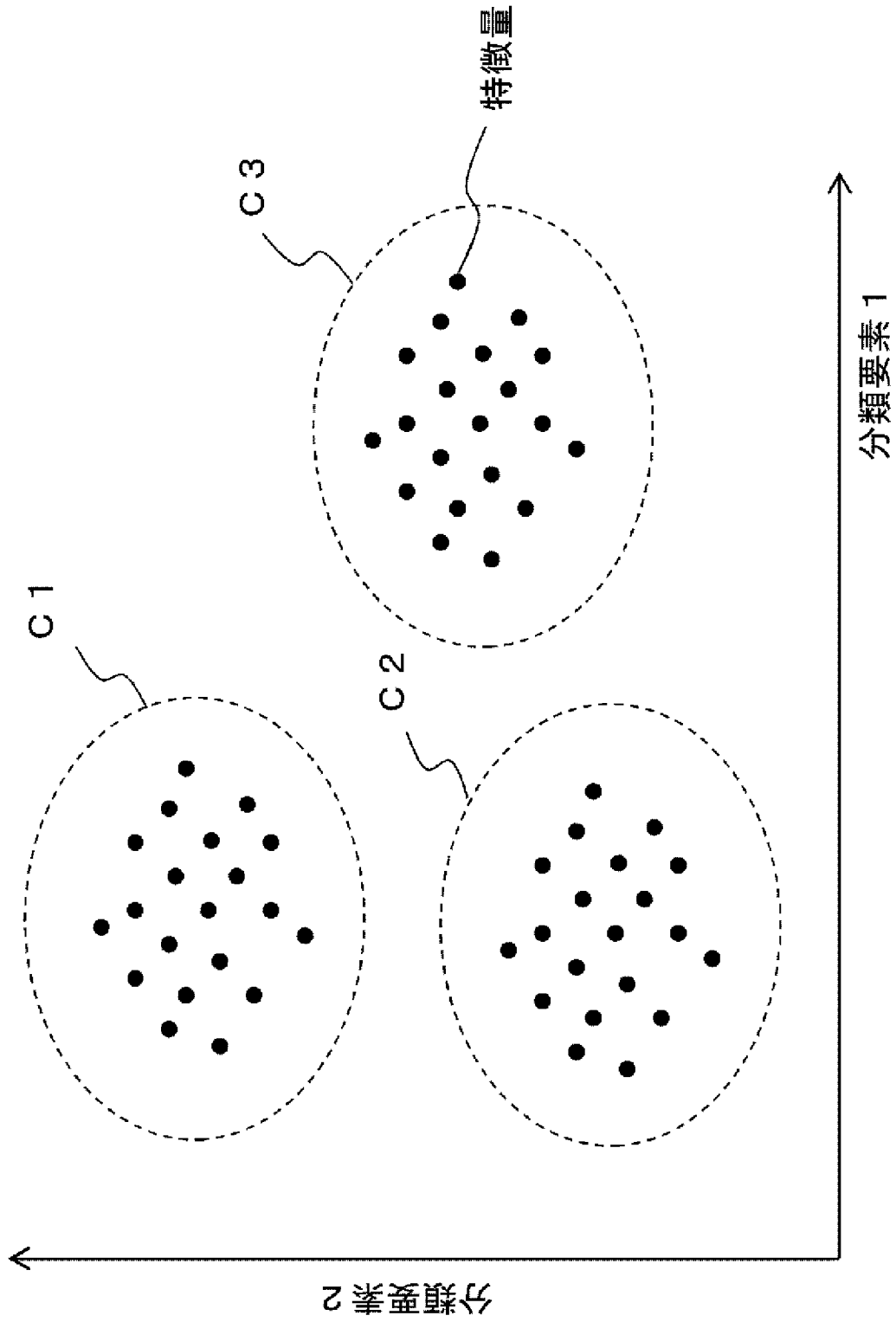
[図9]



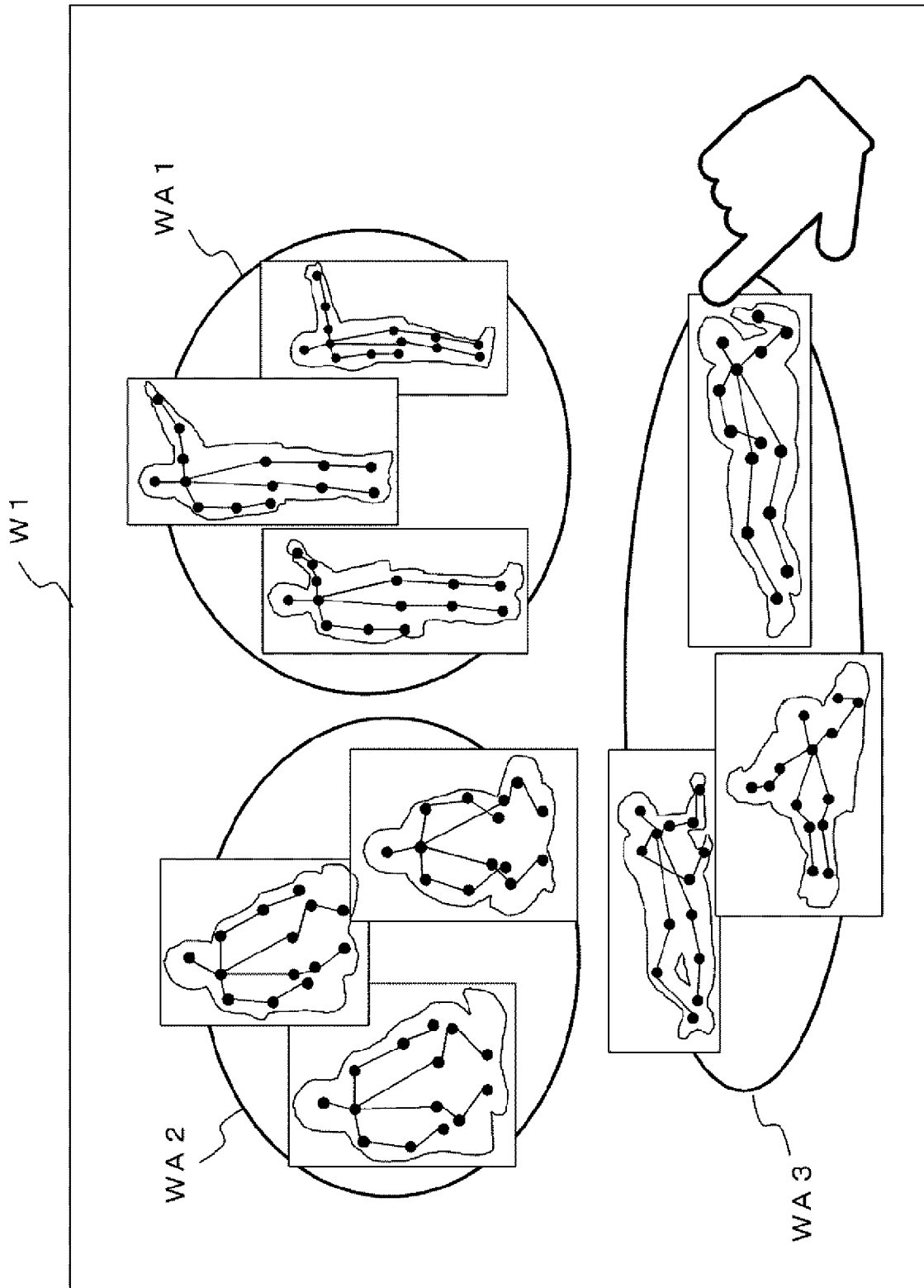
[図10]



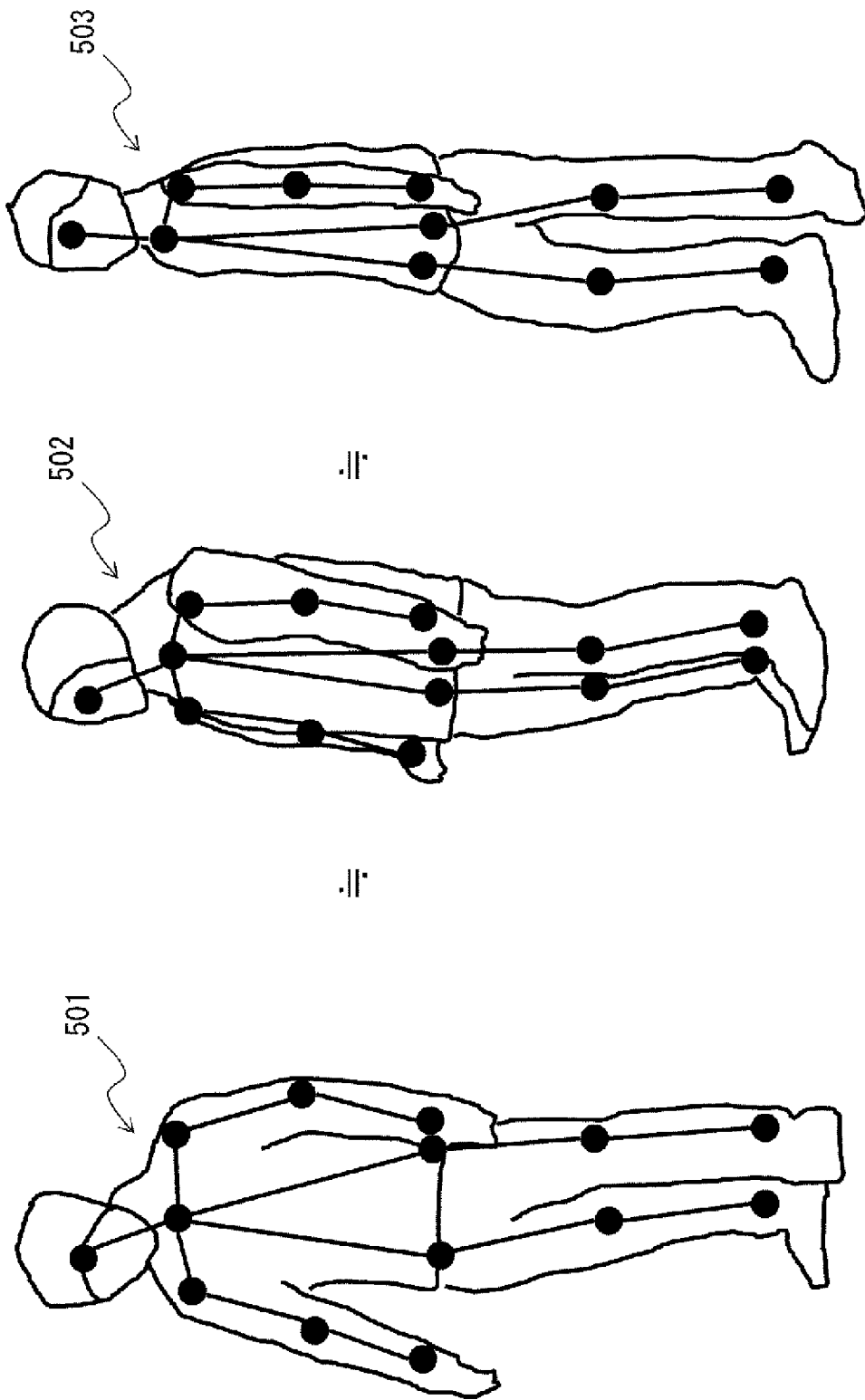
[図11]



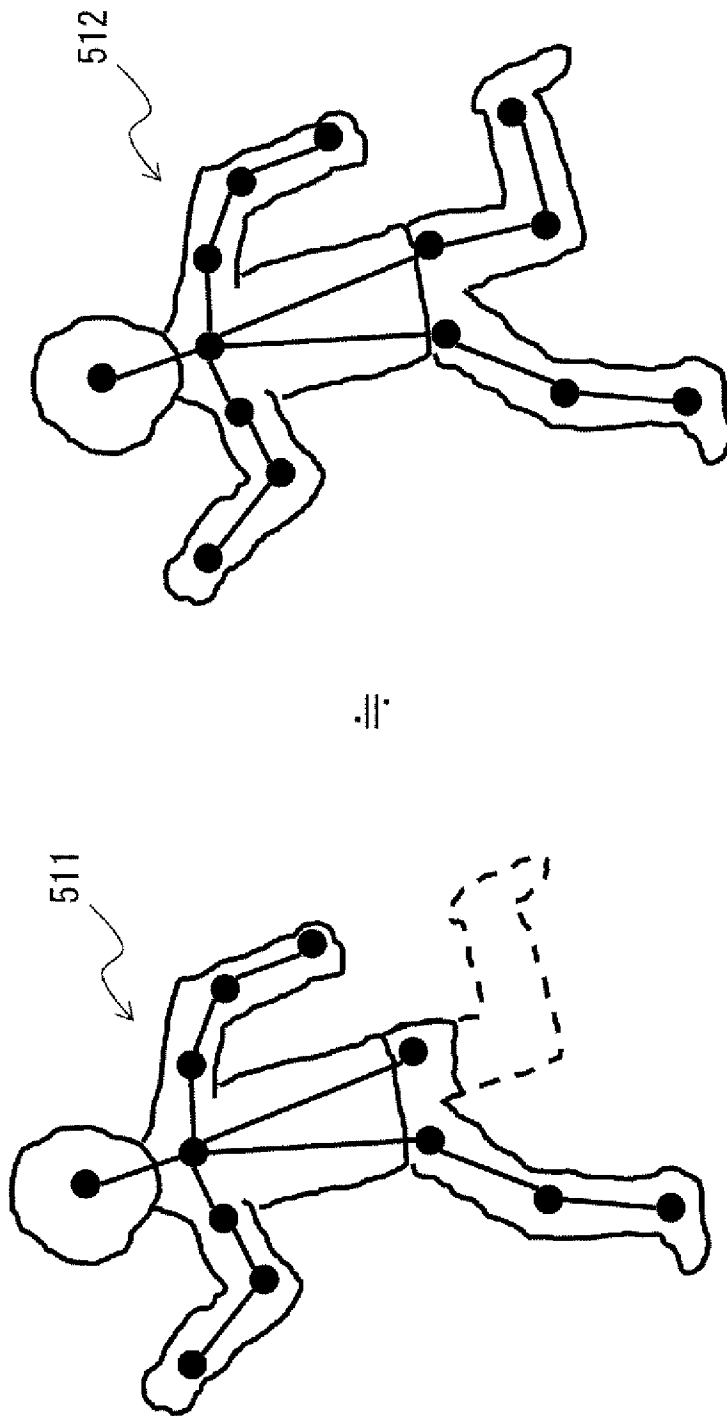
[図12]



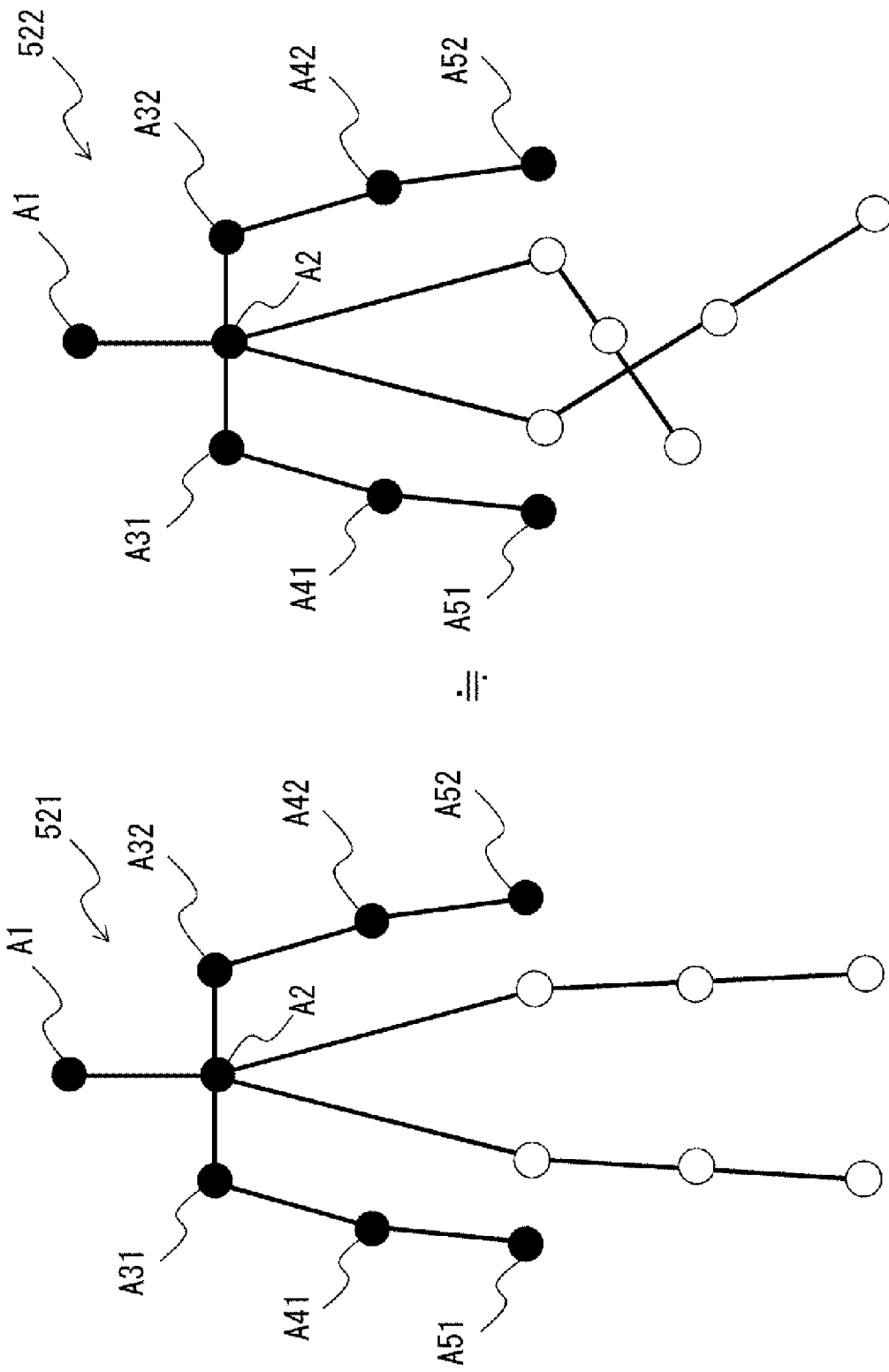
[図13]



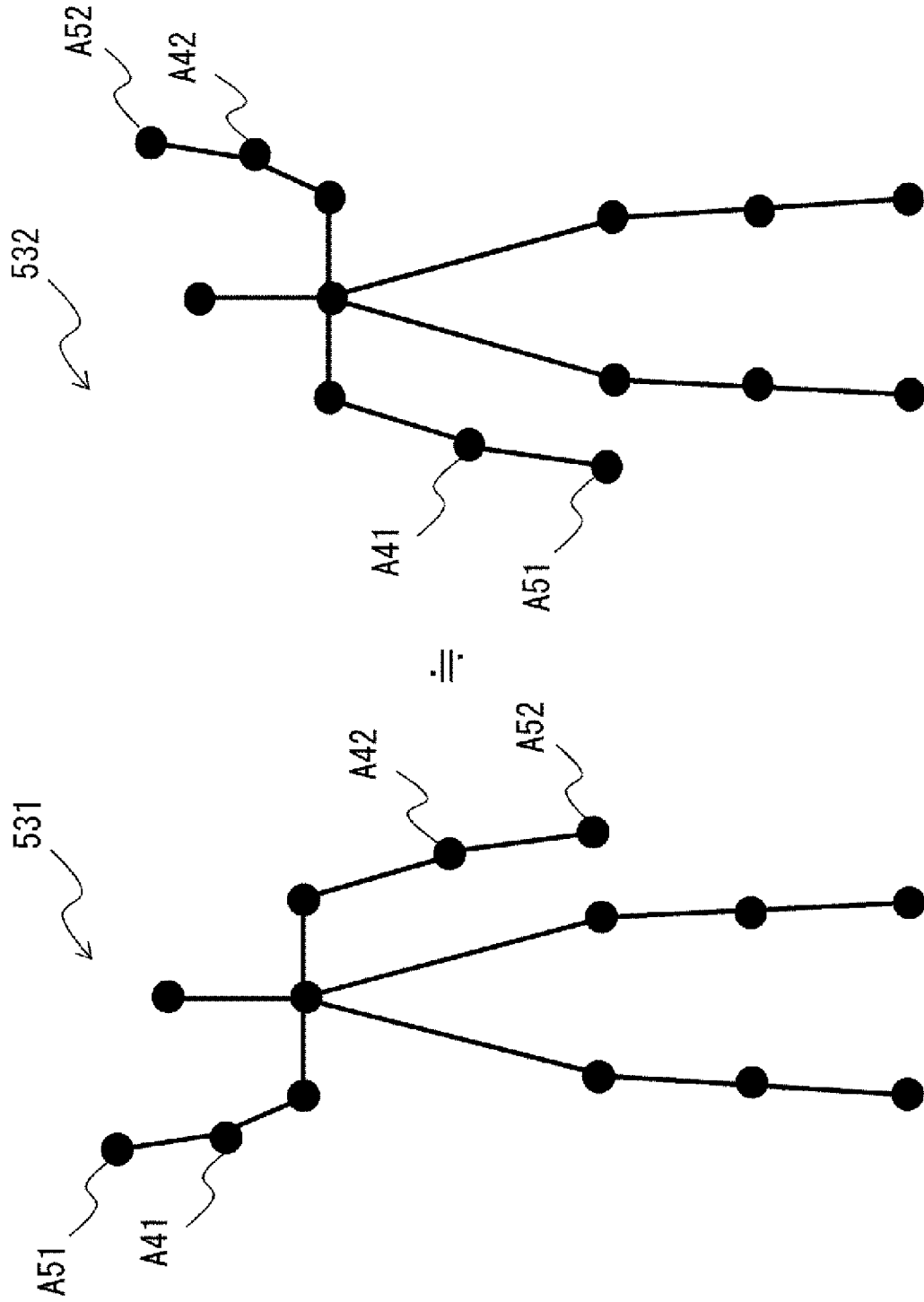
[図14]



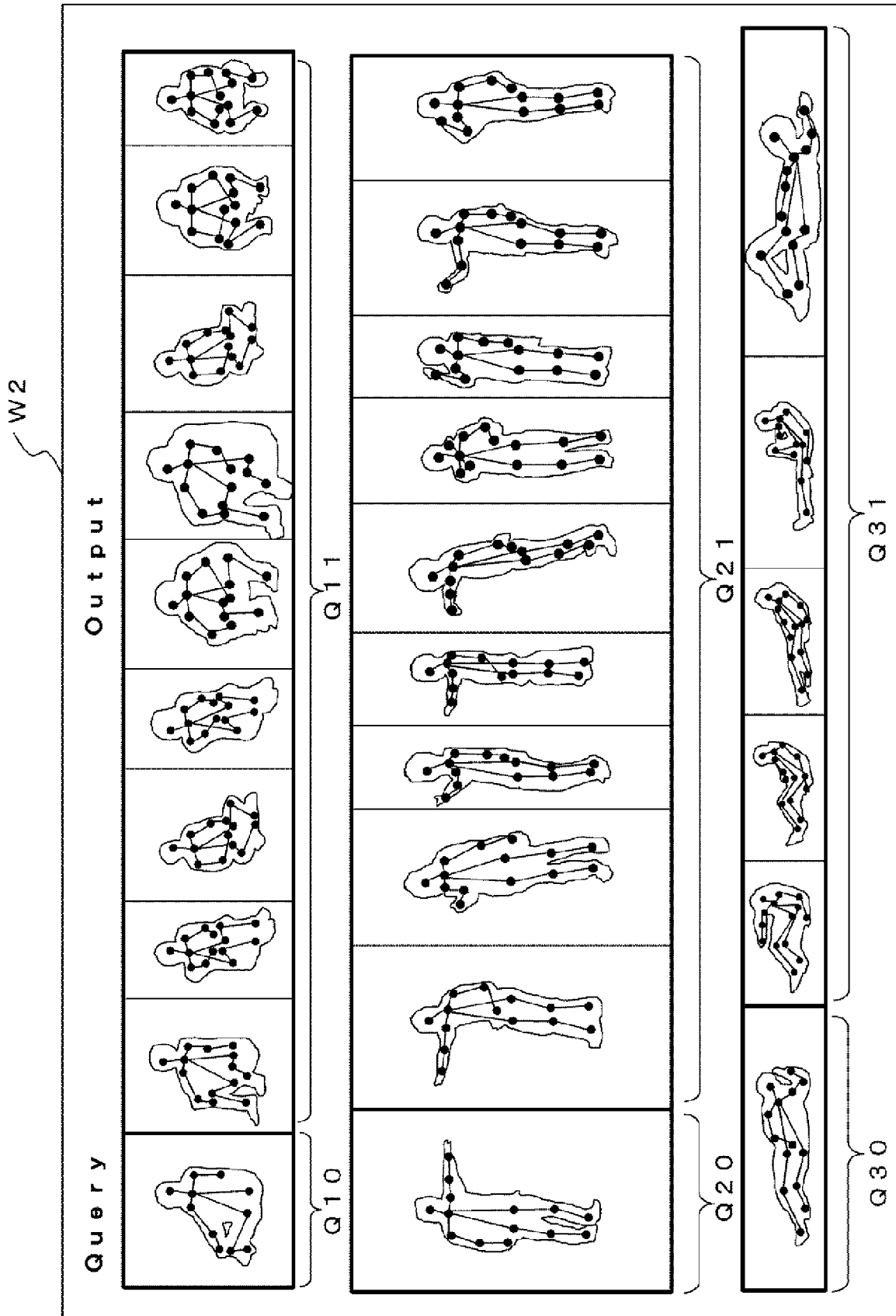
[図15]



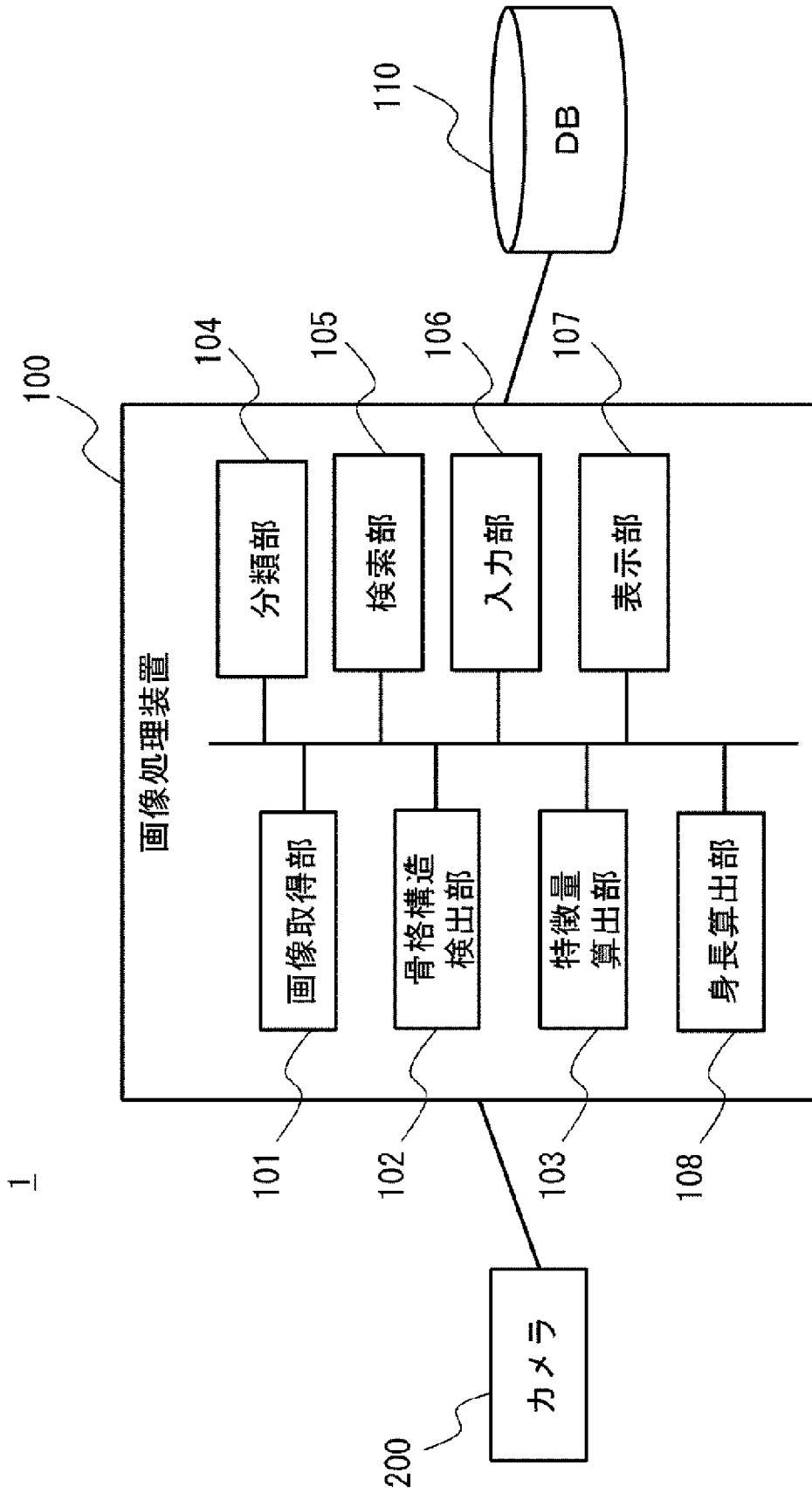
[図16]



[図17]

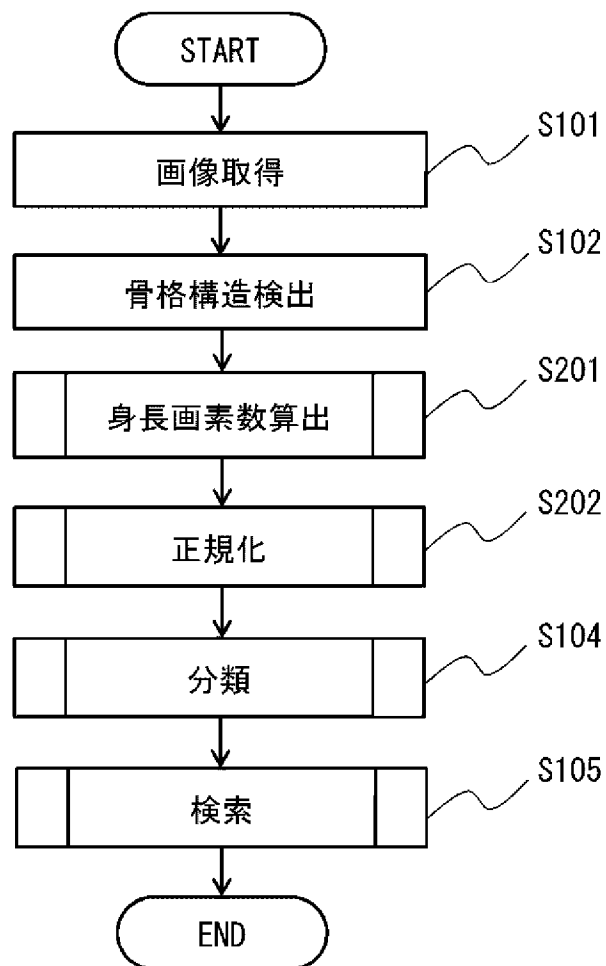


[図18]

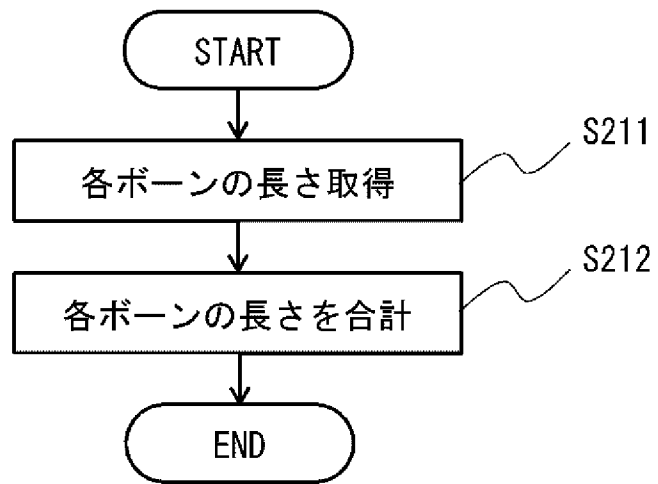


1

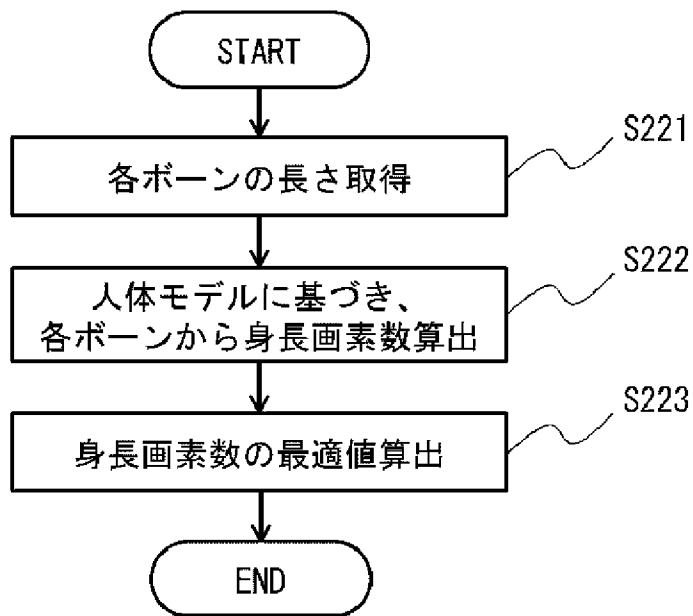
[図19]



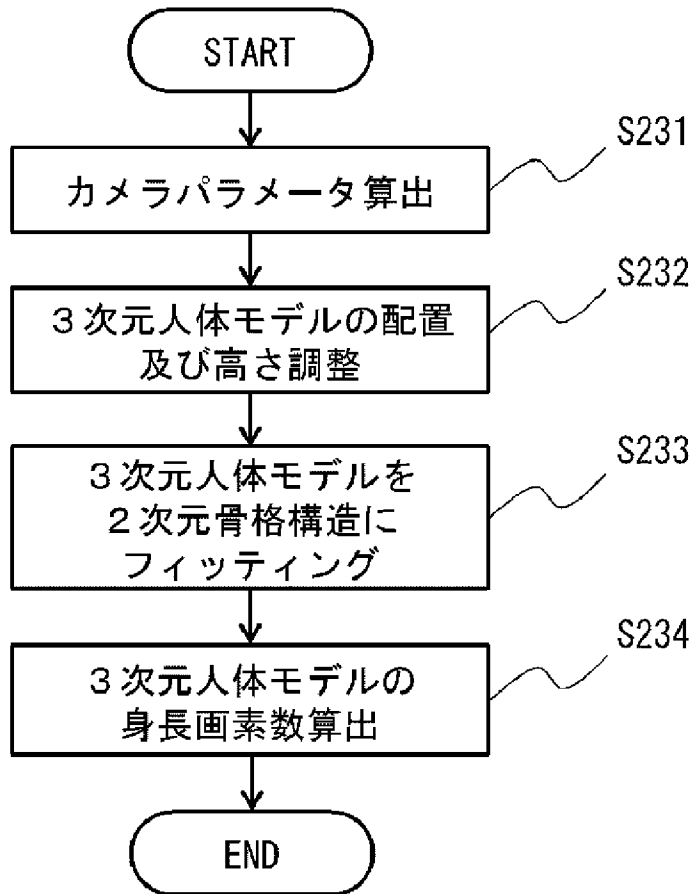
[図20]



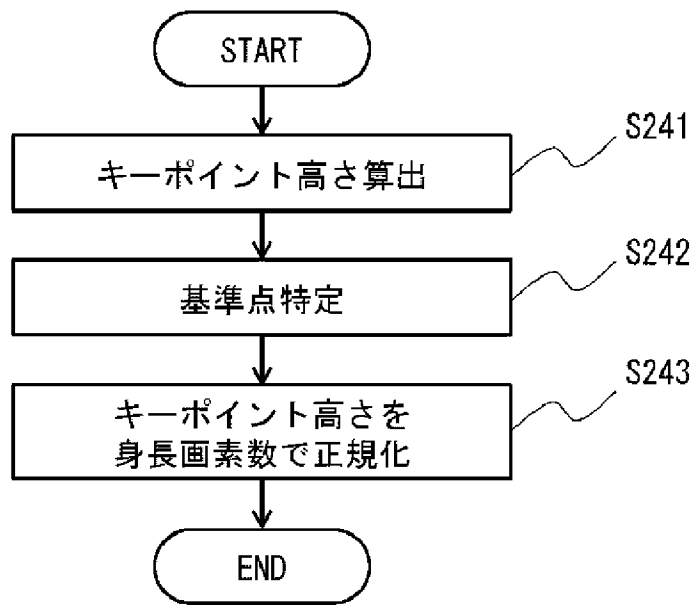
[図21]



[図22]

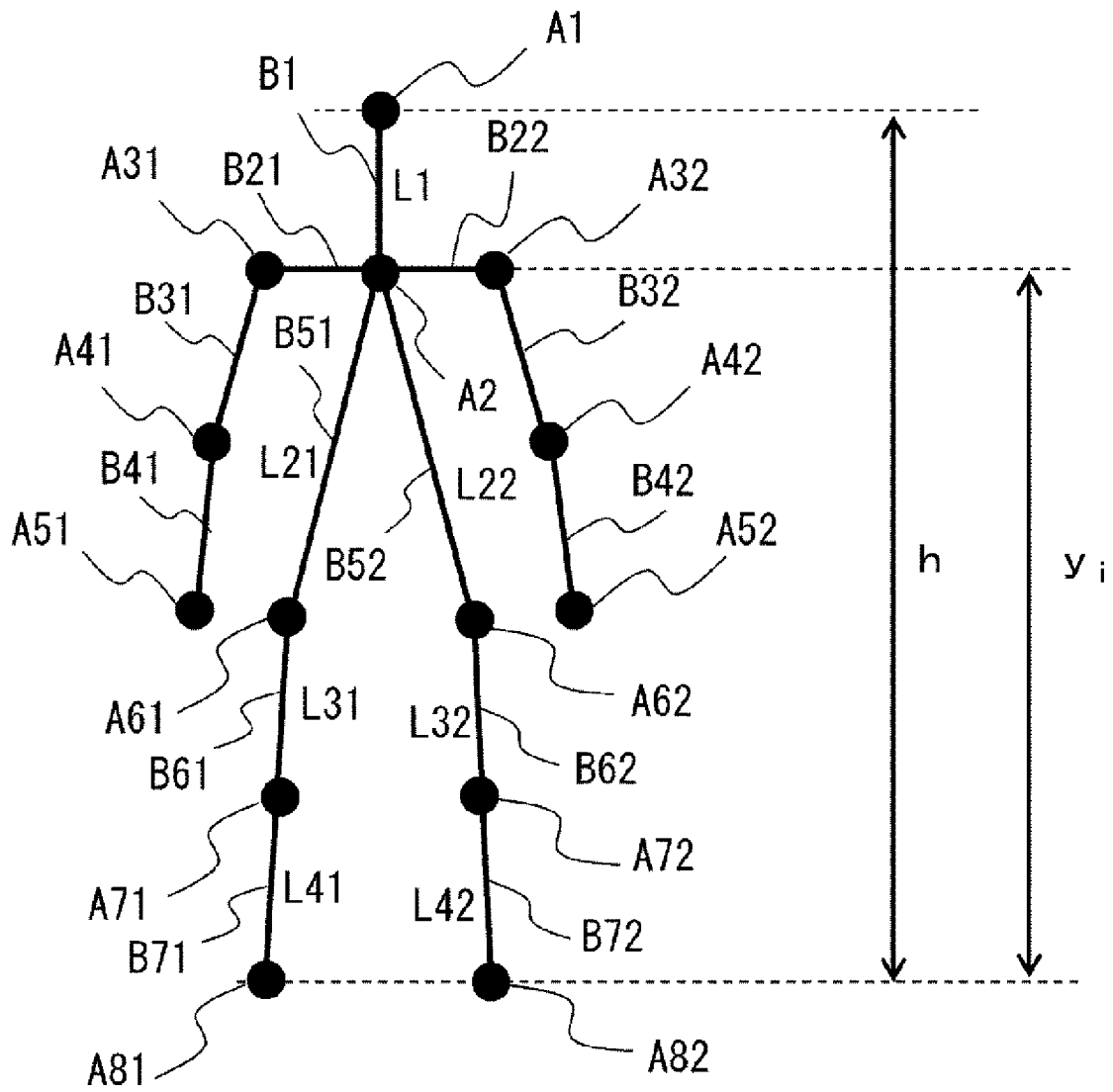


[図23]

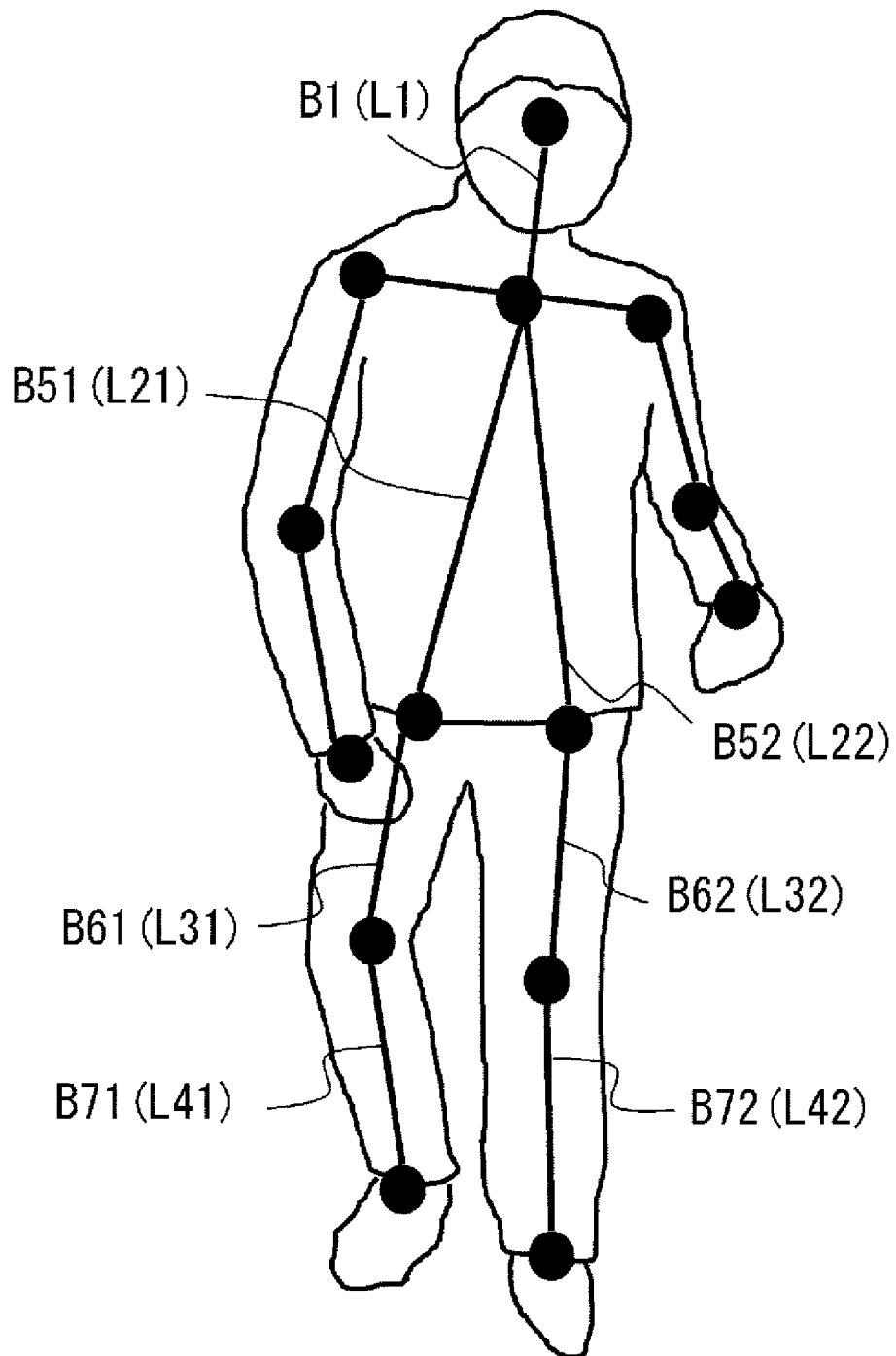


[図24]

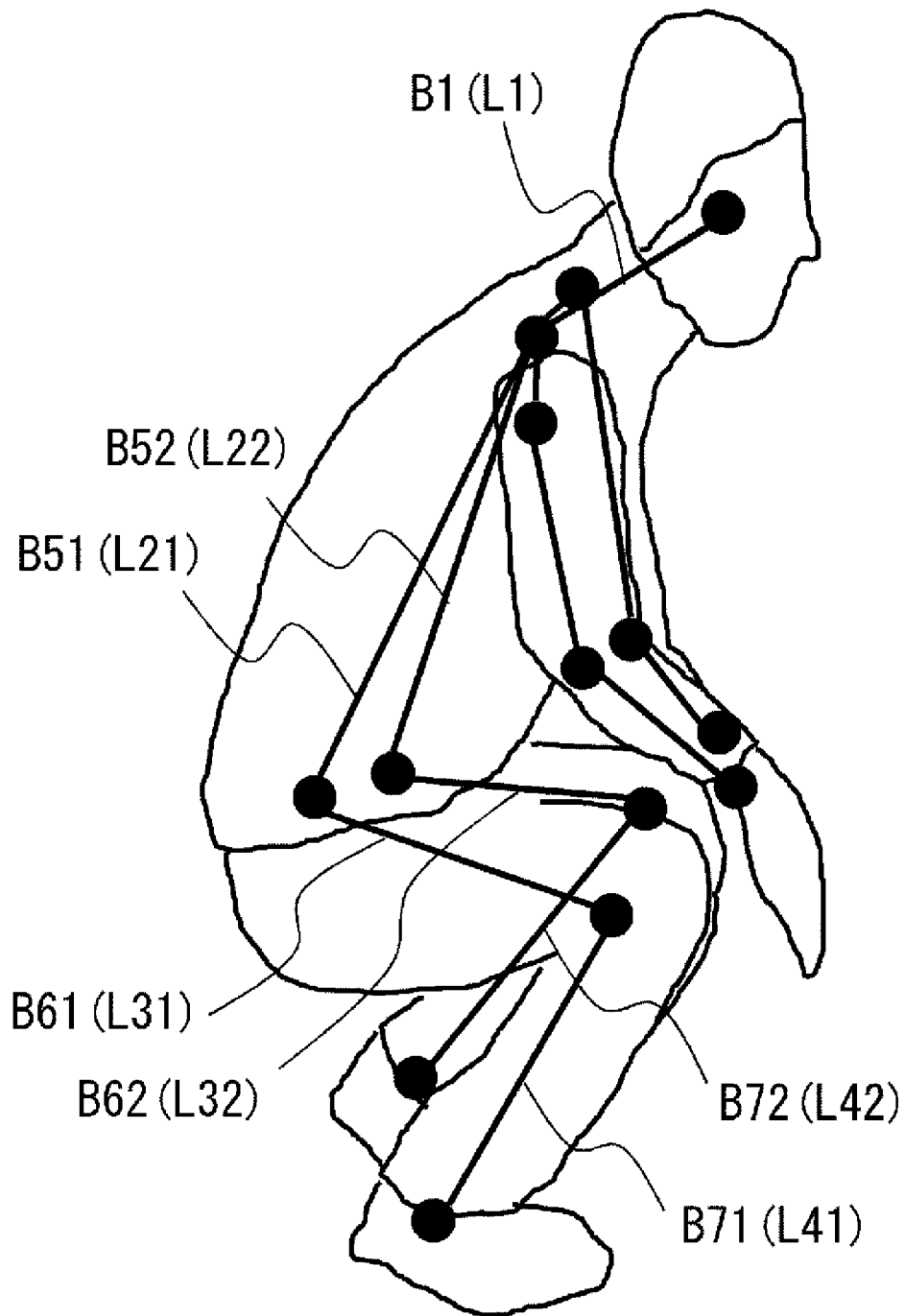
300



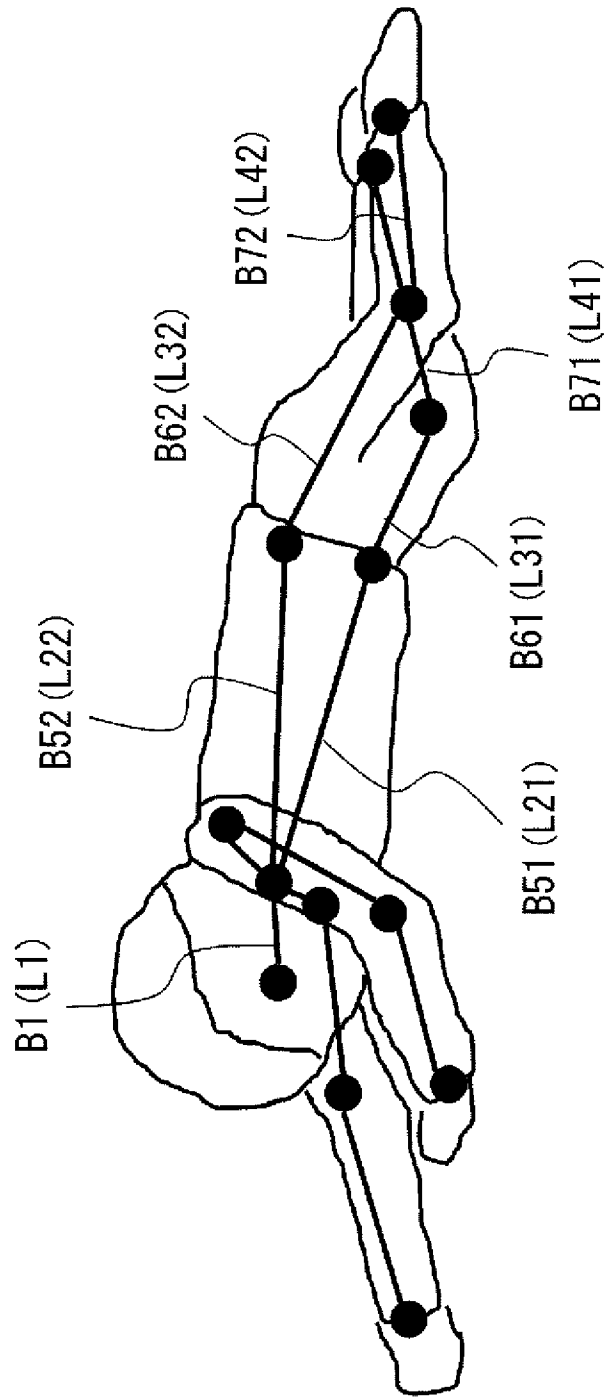
[図25]



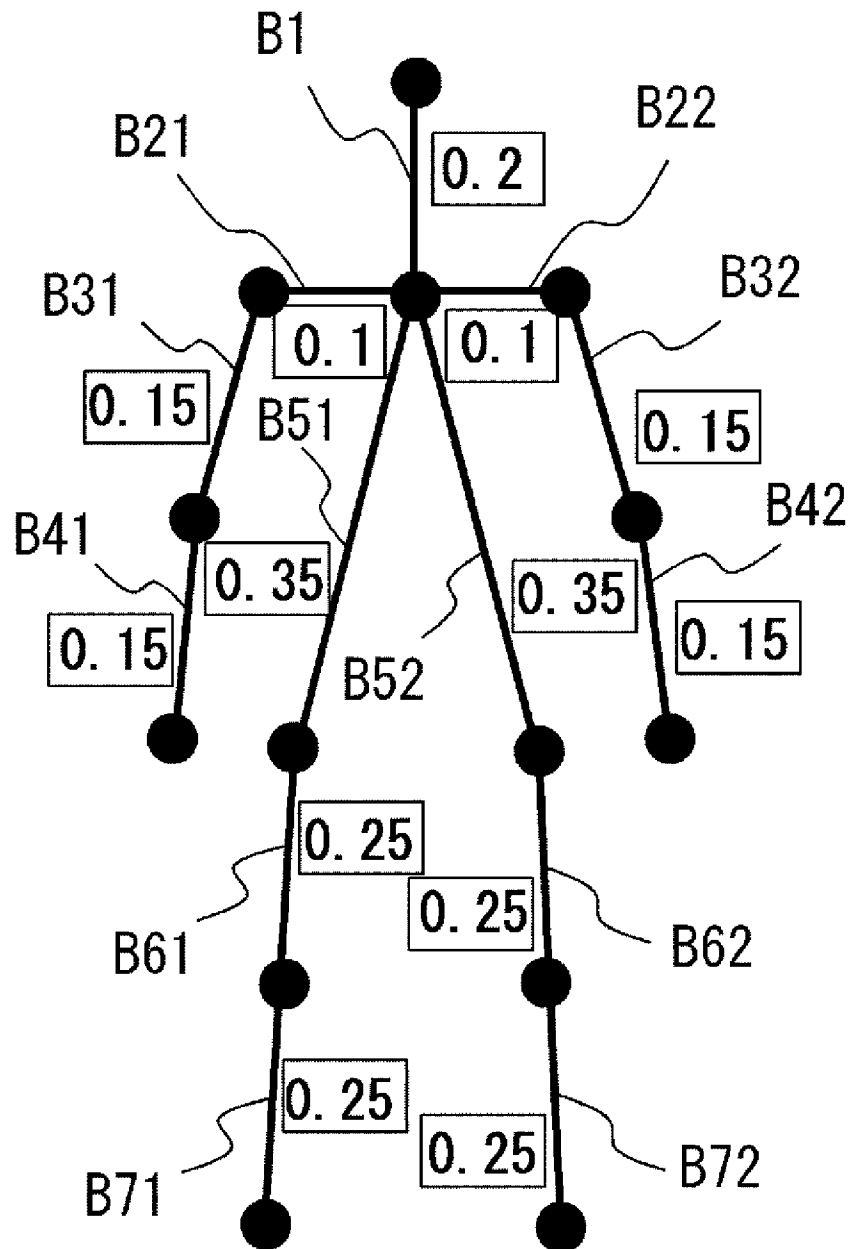
[図26]



[図27]

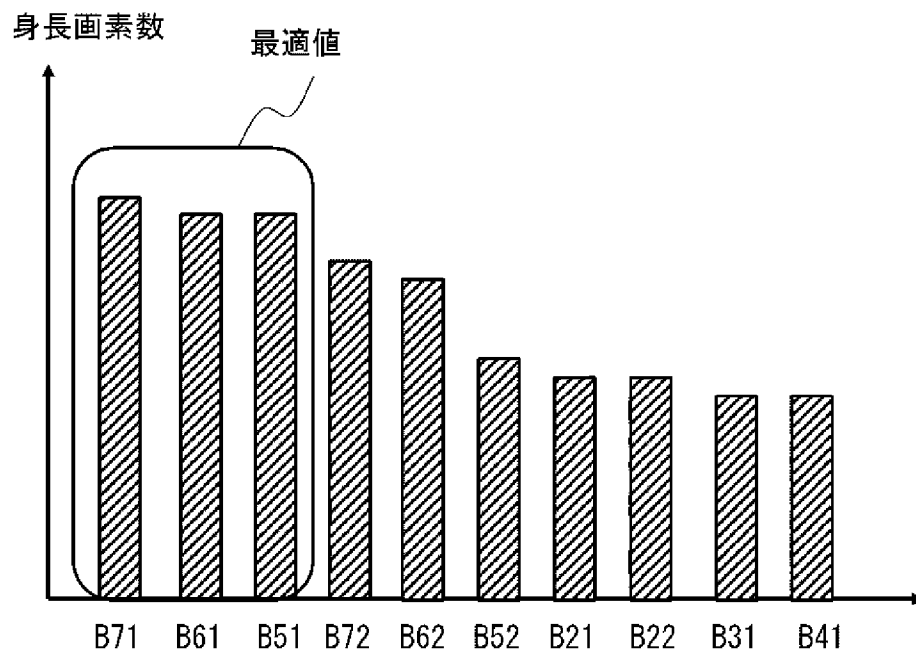


[図28]

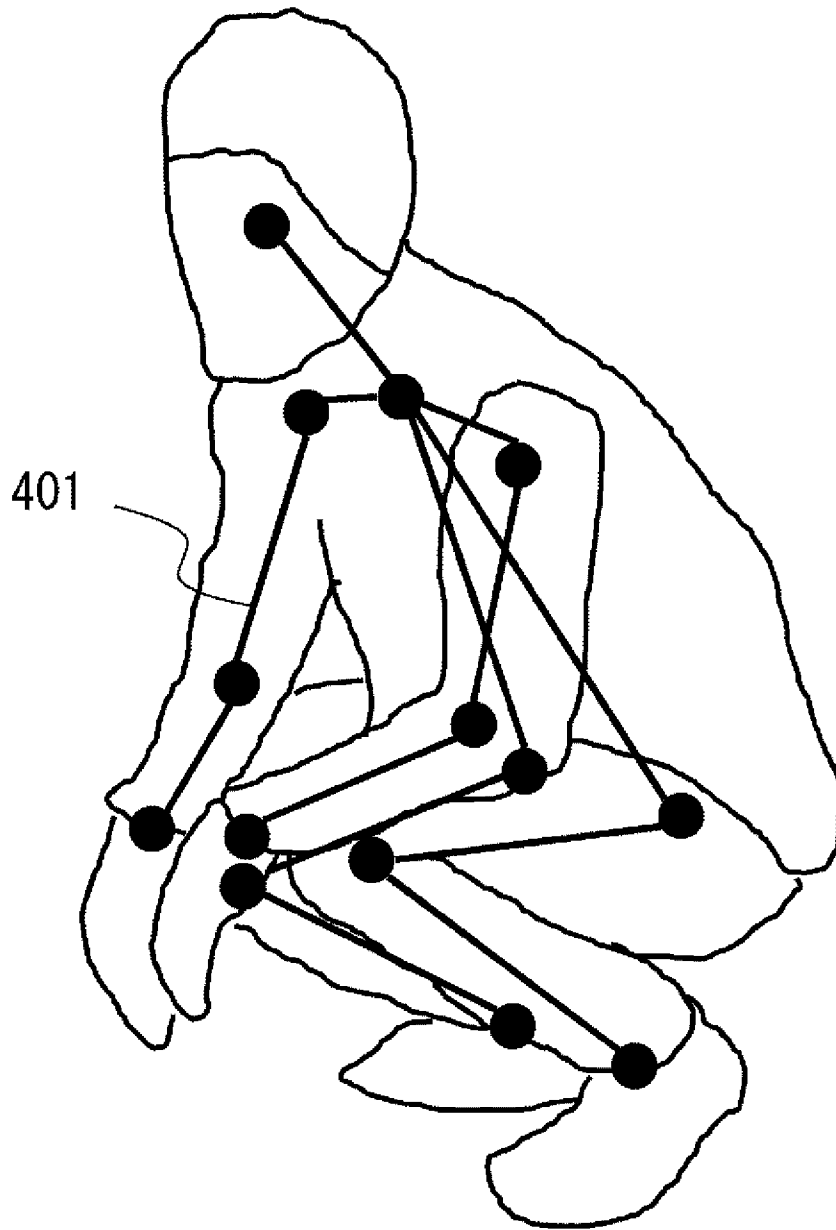
301



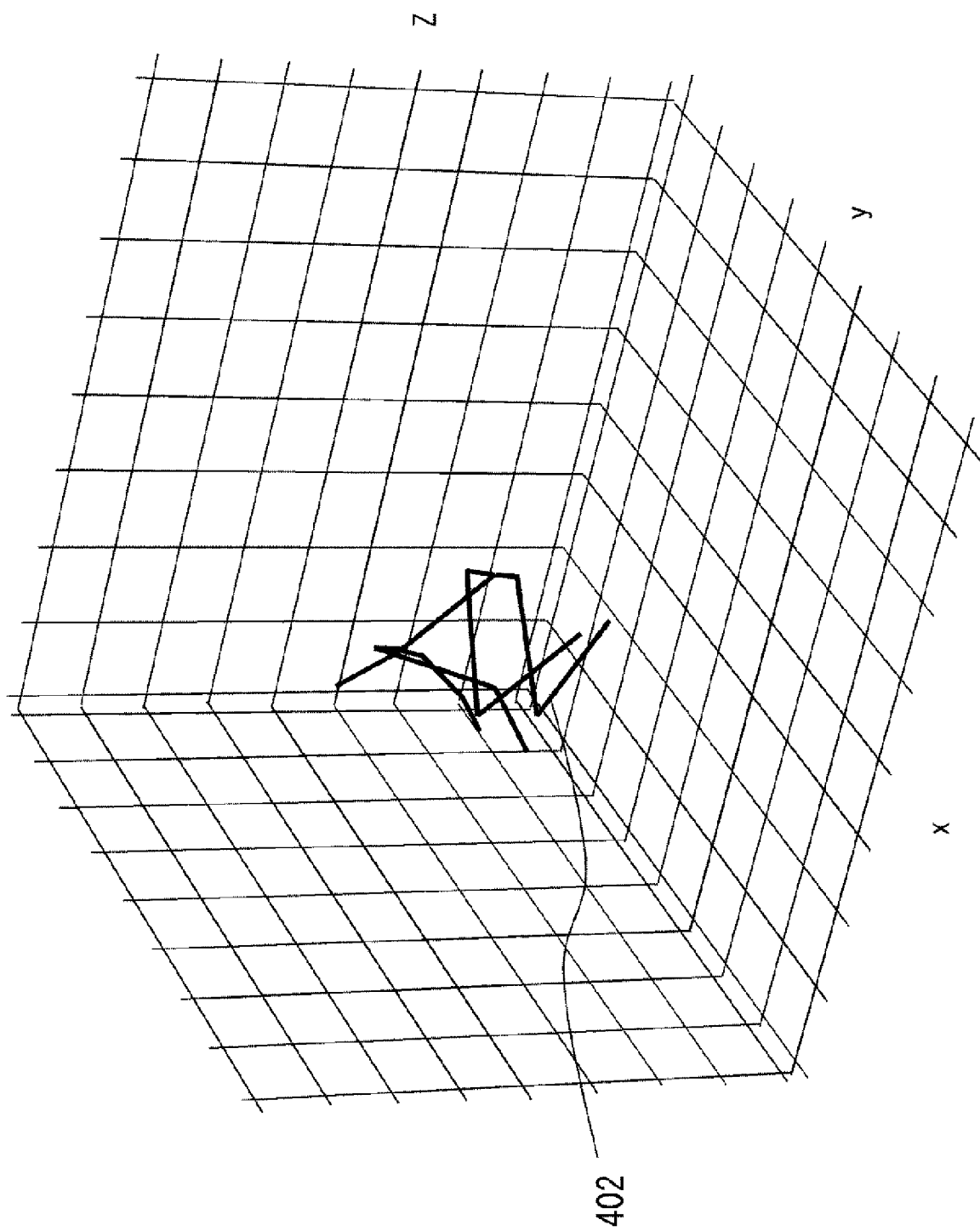
[図30]



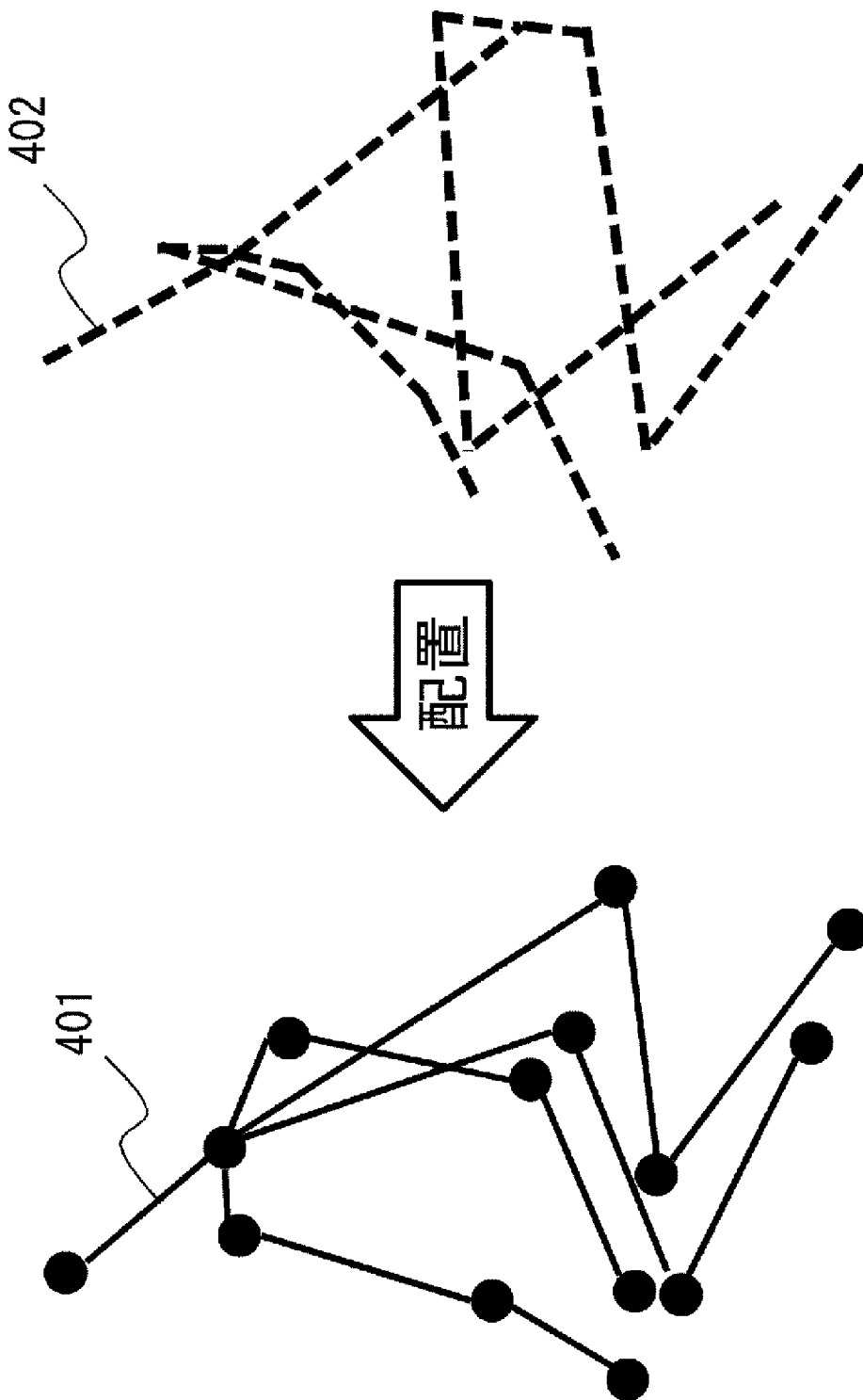
[図31]



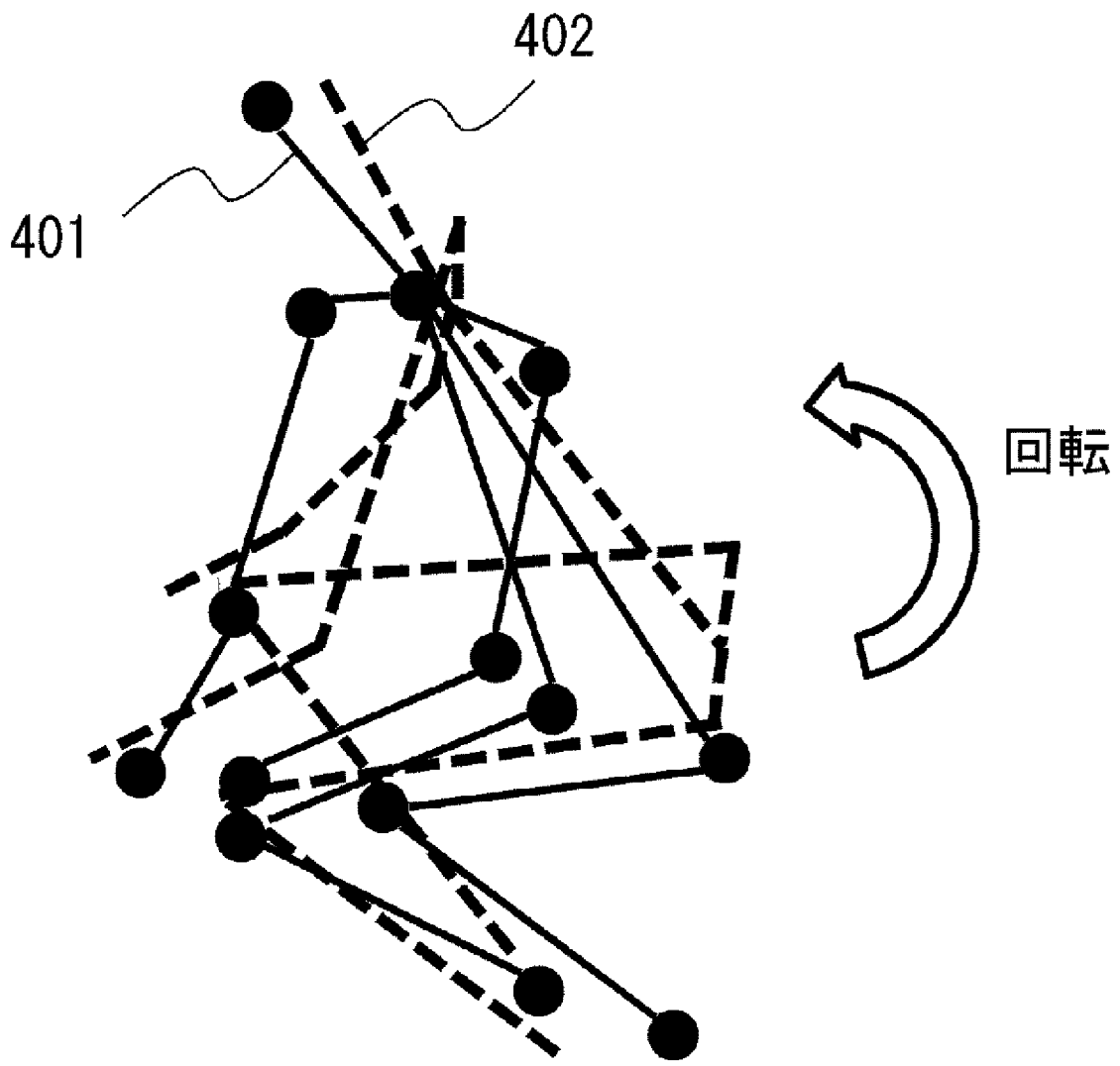
[図32]



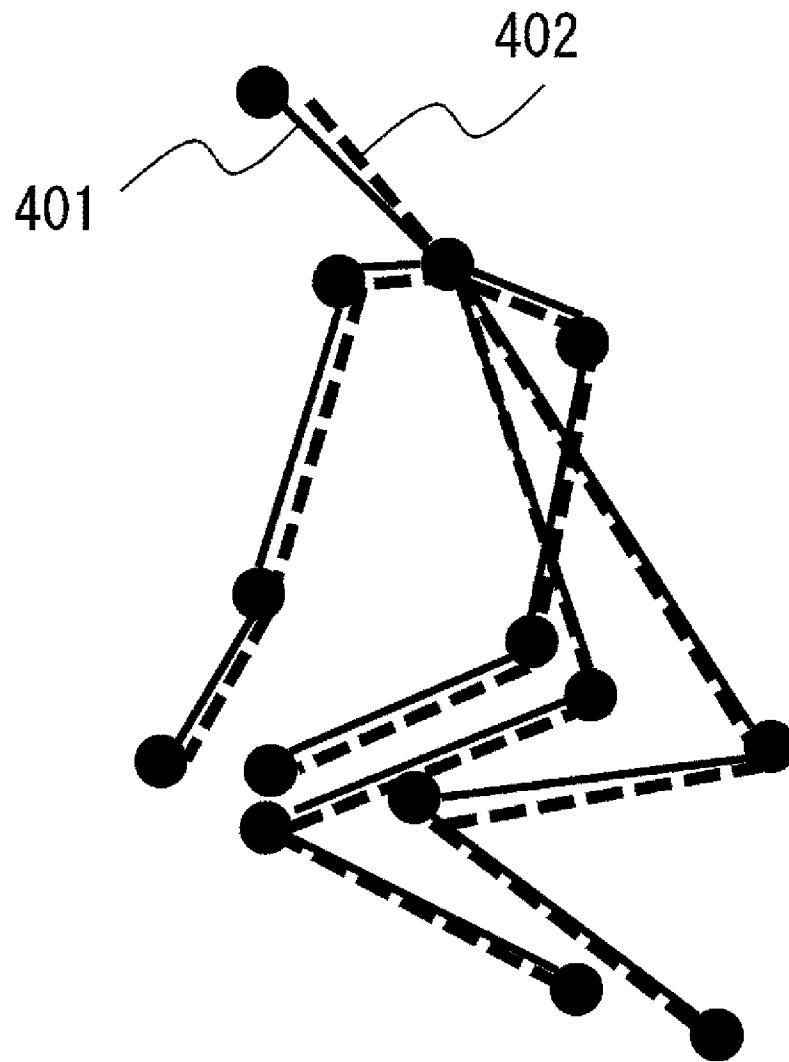
[図33]



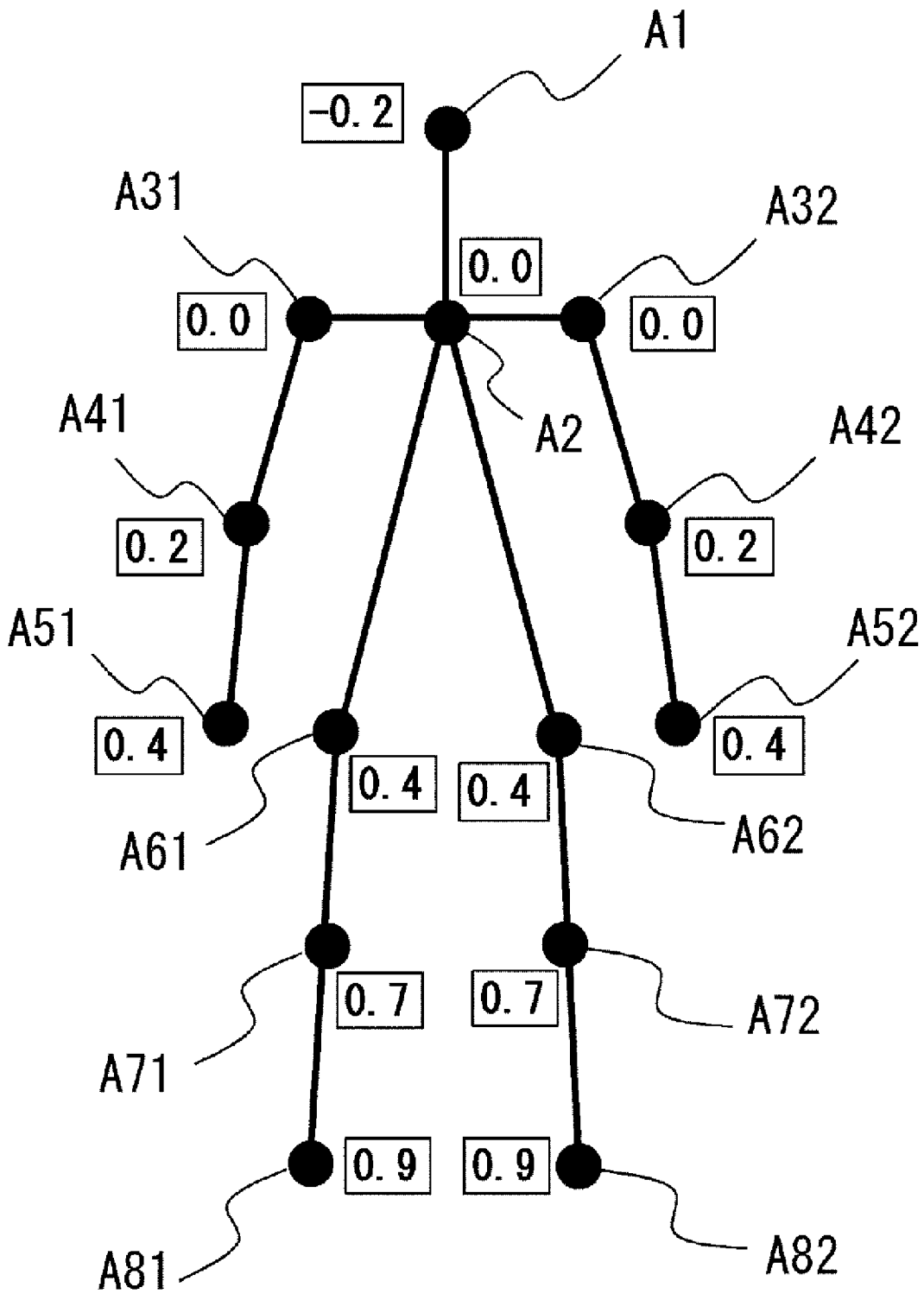
[図34]



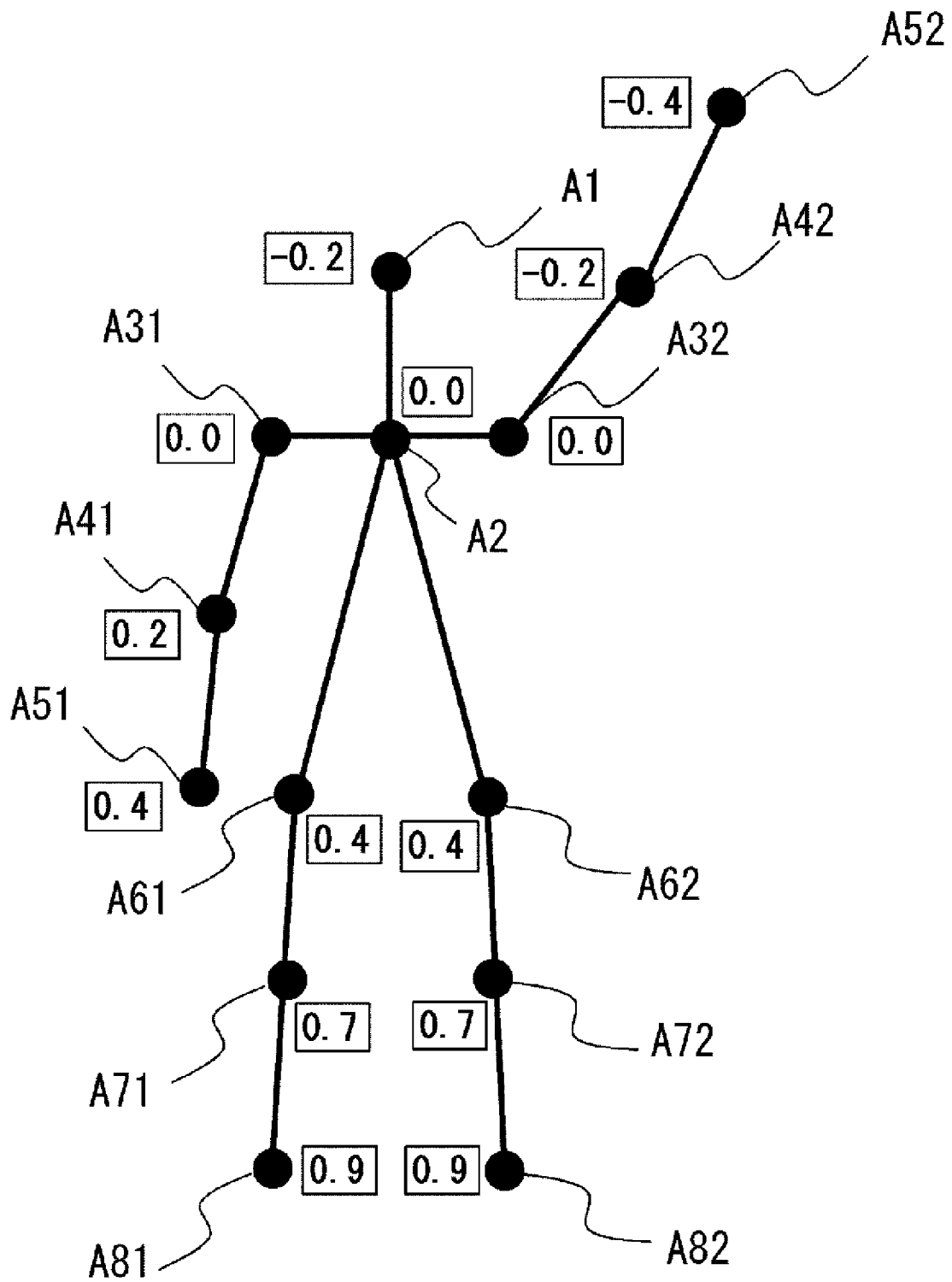
[図35]



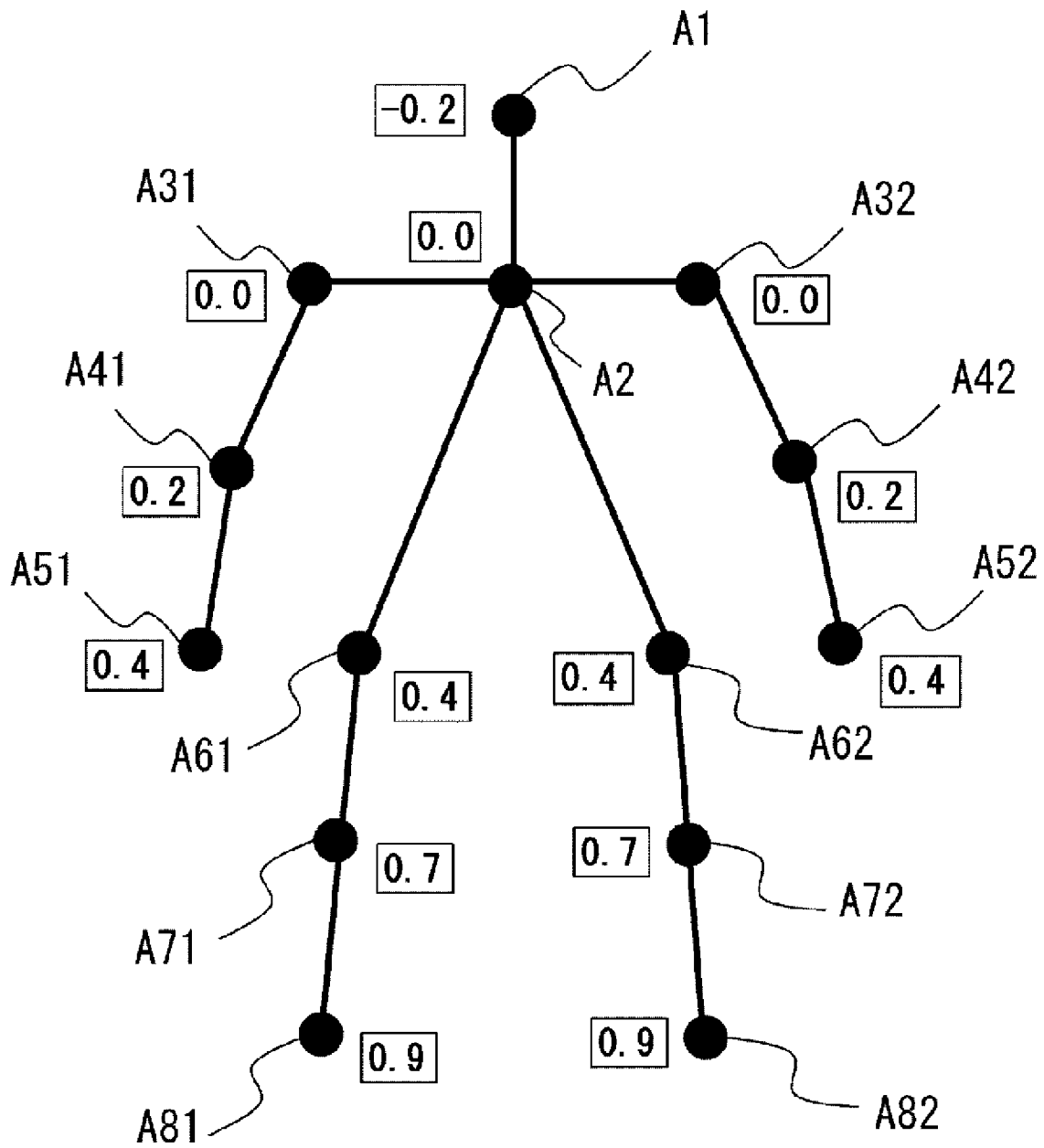
[図36]



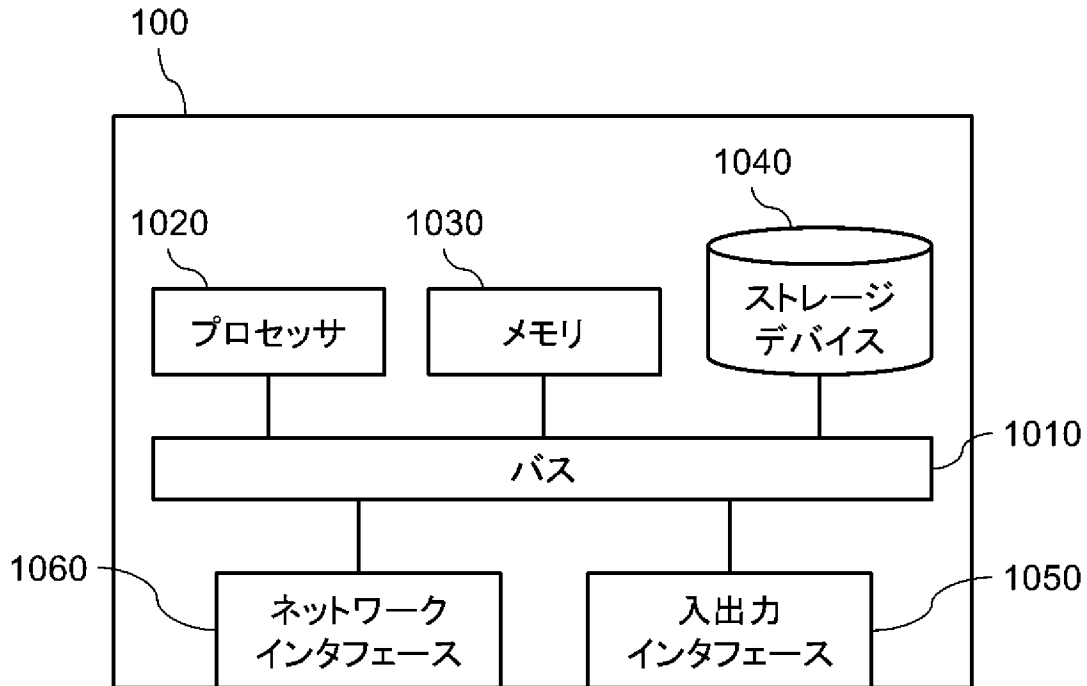
[図37]



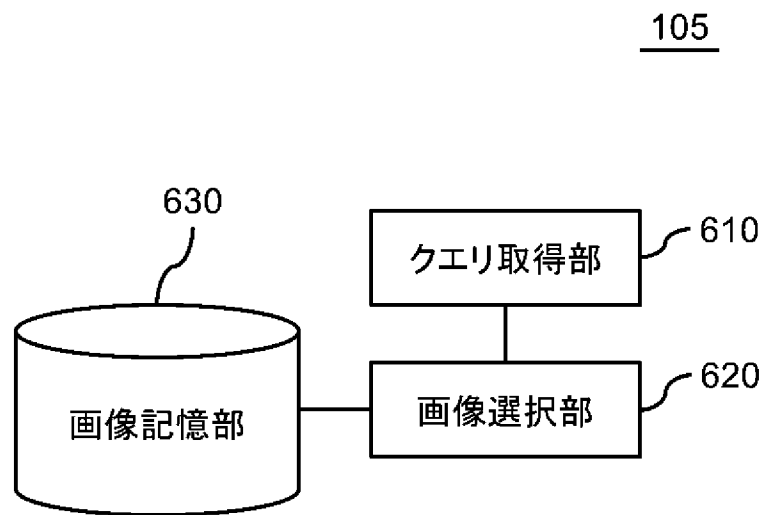
[図38]



[図39]

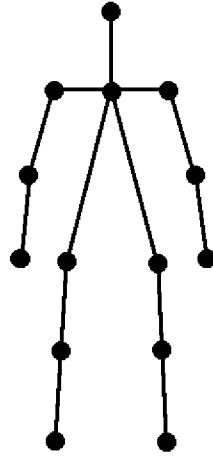


[図40]

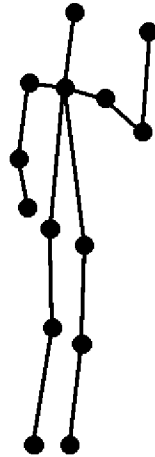


[図41]

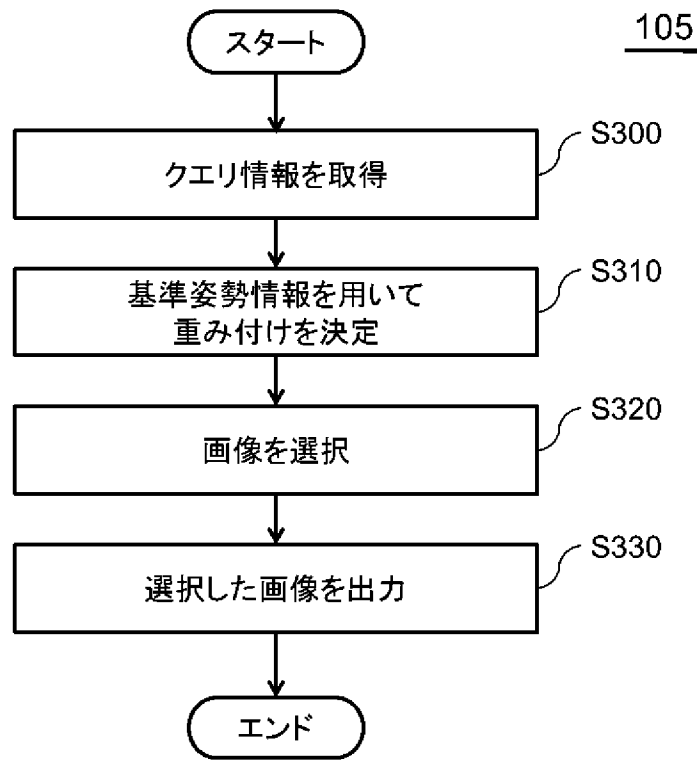
(A) 基準姿勢情報



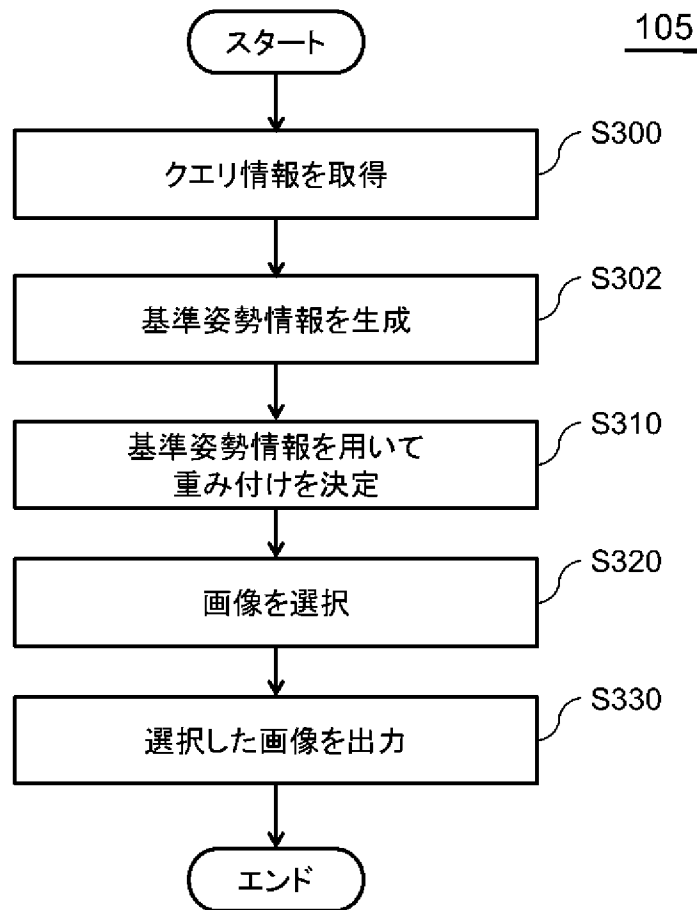
(B) クエリ情報



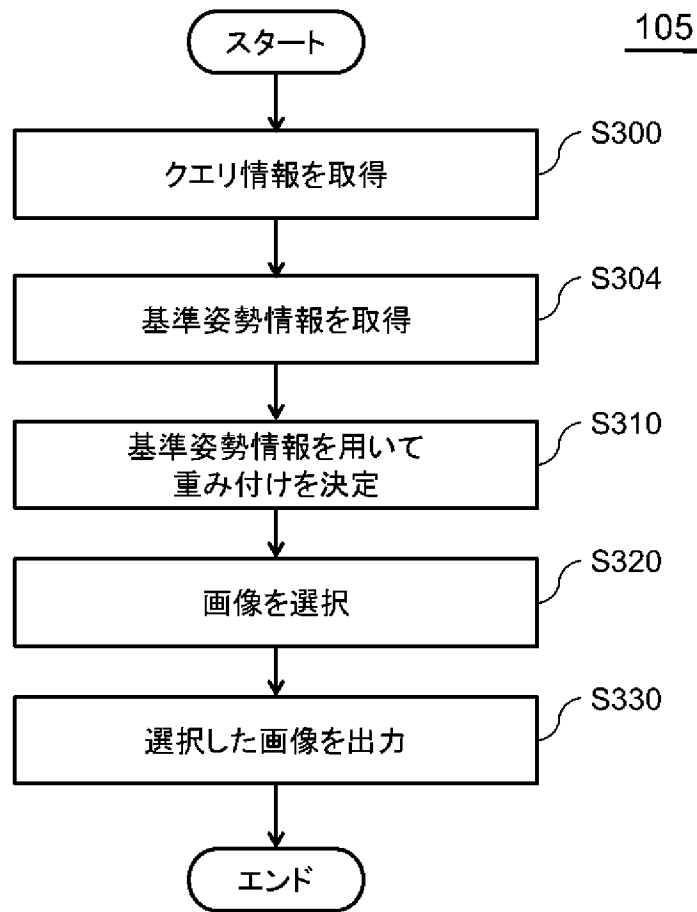
[図42]



[図43]



[図44]



**INTERNATIONAL SEARCH REPORT**

International application No.

PCT/JP2020/026450

<p><b>A. CLASSIFICATION OF SUBJECT MATTER</b>                  Int. Cl. G06F16/53 (2019.01) i, G06F16/532 (2019.01) i                  FI: G06F16/53, G06F16/532</p> <p>According to International Patent Classification (IPC) or to both national classification and IPC</p>								
<p><b>B. FIELDS SEARCHED</b></p> <p>Minimum documentation searched (classification system followed by classification symbols)                  Int. Cl. G06F16/53, G06F16/532</p> <p>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</p> <p>Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)</p>								
<p><b>C. DOCUMENTS CONSIDERED TO BE RELEVANT</b></p> <table border="1" style="width:100%; border-collapse: collapse;"> <thead> <tr> <th style="width:10%;">Category*</th> <th style="width:70%;">Citation of document, with indication, where appropriate, of the relevant passages</th> <th style="width:20%;">Relevant to claim No.</th> </tr> </thead> <tbody> <tr> <td align="center">X</td> <td>JP 2019-91138 A (HITACHI, LTD.) 13 June 2019, paragraphs [0010]-[0031], [0056], [0063]</td> <td align="center">1-8</td> </tr> </tbody> </table>			Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.	X	JP 2019-91138 A (HITACHI, LTD.) 13 June 2019, paragraphs [0010]-[0031], [0056], [0063]	1-8
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.						
X	JP 2019-91138 A (HITACHI, LTD.) 13 June 2019, paragraphs [0010]-[0031], [0056], [0063]	1-8						
<p><input type="checkbox"/> Further documents are listed in the continuation of Box C.      <input checked="" type="checkbox"/> See patent family annex.</p>								
<p>* Special categories of cited documents:</p> <table style="width:100%;"> <tr> <td style="width:50%;"> <p>“A” document defining the general state of the art which is not considered to be of particular relevance</p> <p>“E” earlier application or patent but published on or after the international filing date</p> <p>“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)</p> <p>“O” document referring to an oral disclosure, use, exhibition or other means</p> <p>“P” document published prior to the international filing date but later than the priority date claimed</p> </td> <td style="width:50%;"> <p>“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</p> <p>“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</p> <p>“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</p> <p>“&amp;” document member of the same patent family</p> </td> </tr> </table>			<p>“A” document defining the general state of the art which is not considered to be of particular relevance</p> <p>“E” earlier application or patent but published on or after the international filing date</p> <p>“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)</p> <p>“O” document referring to an oral disclosure, use, exhibition or other means</p> <p>“P” document published prior to the international filing date but later than the priority date claimed</p>	<p>“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</p> <p>“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</p> <p>“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</p> <p>“&amp;” document member of the same patent family</p>				
<p>“A” document defining the general state of the art which is not considered to be of particular relevance</p> <p>“E” earlier application or patent but published on or after the international filing date</p> <p>“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)</p> <p>“O” document referring to an oral disclosure, use, exhibition or other means</p> <p>“P” document published prior to the international filing date but later than the priority date claimed</p>	<p>“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</p> <p>“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</p> <p>“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</p> <p>“&amp;” document member of the same patent family</p>							
<p>Date of the actual completion of the international search 02.09.2020</p>		<p>Date of mailing of the international search report 15.09.2020</p>						
<p>Name and mailing address of the ISA/ Japan Patent Office 3-4-3, Kasumigaseki, Chiyoda-ku, Tokyo 100-8915, Japan</p>		<p>Authorized officer</p> <p>Telephone No.</p>						

**INTERNATIONAL SEARCH REPORT**  
Information on patent family members

International application No.  
PCT/JP2020/026450

Patent Documents referred to in the Report	Publication Date	Patent Family	Publication Date
JP 2019-91138 A	13.06.2019	EP 3483741 A1 paragraphs [0009]- [0030], [0056], [0063]	

A. 発明の属する分野の分類（国際特許分類（IPC）） G06F 16/53(2019.01)i; G06F 16/532(2019.01)i FI: G06F16/53; G06F16/532		
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）） G06F16/53; G06F16/532		
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
X	JP 2019-91138 A（株式会社日立製作所）13.06.2019（2019 - 06 - 13） 段落[0010]-[0031], [0056], [0063]	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 02.09.2020	国際調査報告の発送日 15.09.2020	
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） 三橋 竜太郎 5N 6302 電話番号 03-3581-1101 内線 3586	

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/JP2020/026450

引用文献	公表日	パテントファミリー文献	公表日
JP 2019-91138 A	13.06.2019	EP 3483741 A1 段落[0009]-[0030], [0056], [0063]	