

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7396364号
(P7396364)

(45)発行日 令和5年12月12日(2023.12.12)

(24)登録日 令和5年12月4日(2023.12.4)

(51)国際特許分類 F I
G 0 6 T 7/00 (2017.01) G 0 6 T 7/00 6 6 0 B

請求項の数 22 (全27頁)

(21)出願番号	特願2021-553986(P2021-553986)	(73)特許権者	000004237 日本電気株式会社 東京都港区芝五丁目7番1号
(86)(22)出願日	令和1年10月31日(2019.10.31)	(74)代理人	100103894 弁理士 家入 健
(86)国際出願番号	PCT/JP2019/042805	(72)発明者	吉田 登 東京都港区芝五丁目7番1号 日本電気株式会社内
(87)国際公開番号	WO2021/084687	(72)発明者	西村 祥治 東京都港区芝五丁目7番1号 日本電気株式会社内
(87)国際公開日	令和3年5月6日(2021.5.6)	審査官	片岡 利延
審査請求日	令和4年3月28日(2022.3.28)		

最終頁に続く

(54)【発明の名称】 画像処理装置、画像処理方法及び画像処理プログラム

(57)【特許請求の範囲】

【請求項1】

取得される2次元画像に基づいて人物の2次元骨格構造を検出する骨格検出手段と、
前記検出された2次元骨格構造に基づいて2次元画像空間上の前記人物の直立時の高さを推定する推定手段と、

前記推定された直立時の高さに基づいて前記検出された2次元骨格構造を正規化する正規化手段と、

を備え、

前記正規化手段は、前記2次元骨格構造に含まれる特徴点の高さを前記直立時の高さにより正規化し、

前記正規化手段は、前記2次元骨格構造内の基準点に対する前記特徴点の相対的な高さを前記直立時の高さにより正規化する、

画像処理装置。

【請求項2】

前記特徴点の高さは、前記2次元画像空間を表すX-Y座標におけるY軸方向の高さである、

請求項1に記載の画像処理装置。

【請求項3】

前記特徴点の高さは、前記2次元画像の撮像パラメータに基づいて、実世界の3次元空間における基準面に対する鉛直方向を、前記2次元画像空間に投影した鉛直投影方向の高

さである、

請求項 1 に記載の画像処理装置。

【請求項 4】

前記基準点は、前記 2 次元骨格構造における中心よりも前記 2 次元画像空間で上の点である、

請求項 1 に記載の画像処理装置。

【請求項 5】

前記基準点は、前記 2 次元骨格構造における首部または頭部の特徴点である、

請求項 4 に記載の画像処理装置。

【請求項 6】

前記推定手段は、前記 2 次元骨格構造に含まれる 2 次元画像空間上の骨の長さに基づいて、前記人物の直立時の高さを推定する、

請求項 1 乃至 5 のいずれか一項に記載の画像処理装置。

【請求項 7】

前記推定手段は、前記 2 次元骨格構造に含まれる足部から頭部までの前記骨の長さの合計に基づいて、前記人物の直立時の高さを推定する、

請求項 6 に記載の画像処理装置。

【請求項 8】

前記推定手段は、前記骨の長さと 2 次元画像空間上の前記人物の全身の長さとの関係を示す 2 次元骨格モデルに基づいて、前記人物の直立時の高さを推定する、

請求項 6 に記載の画像処理装置。

【請求項 9】

取得される 2 次元画像に基づいて人物の 2 次元骨格構造を検出する骨格検出手段と、
前記検出された 2 次元骨格構造に基づいて 2 次元画像空間上の前記人物の直立時の高さを推定する推定手段と、

前記推定された直立時の高さに基づいて前記検出された 2 次元骨格構造を正規化する正規化手段と、

を備え、

前記推定手段は、前記 2 次元画像の撮像パラメータに基づいて前記 2 次元骨格構造にフィッティングさせた 3 次元骨格モデルに基づいて、前記人物の直立時の高さを推定する、

画像処理装置。

【請求項 10】

前記骨格検出手段は、前記 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し、前記正規化手段は、前記複数の人物の 2 次元骨格構造を正規化し、

前記正規化された複数の 2 次元骨格構造の正規化値に基づいて、前記複数の人物の状態の認識処理を行う認識手段をさらに備える、

請求項 1 乃至 9 のいずれか一項に記載の画像処理装置。

【請求項 11】

前記認識手段は、前記認識処理として、前記複数の人物の状態を分類する、

請求項 10 に記載の画像処理装置。

【請求項 12】

前記認識手段は、前記 2 次元骨格構造の全体または一部の正規化値に基づいて、前記複数の人物の状態を分類する、

請求項 11 に記載の画像処理装置。

【請求項 13】

前記骨格検出手段は、時系列に連続する複数の 2 次元画像から、前記複数の人物の 2 次元骨格構造を検出し、

前記認識手段は、前記複数の 2 次元画像から検出された 2 次元骨格構造の正規化値に基づいて、前記複数の人物の状態を分類する、

請求項 11 または 12 に記載の画像処理装置。

10

20

30

40

50

【請求項 14】

前記認識手段は、前記複数の2次元画像における前記2次元骨格構造の正規化値の変化に基づいて、前記複数の人物の状態を分類する、

請求項 13 に記載の画像処理装置。

【請求項 15】

前記認識手段は、前記認識処理として、前記複数の人物の状態からクエリ状態を検索する、

請求項 10 に記載の画像処理装置。

【請求項 16】

前記認識手段は、前記2次元骨格構造の全体または一部の正規化値に基づいて、前記クエリ状態を検索する、

請求項 15 に記載の画像処理装置。

【請求項 17】

前記骨格検出手段は、時系列に連続する複数の2次元画像から、前記複数の人物の2次元骨格構造を検出し、

前記認識手段は、前記複数の2次元画像から検出された2次元骨格構造の正規化値に基づいて、前記クエリ状態を検索する、

請求項 15 または 16 に記載の画像処理装置。

【請求項 18】

前記認識手段は、前記複数の2次元画像における前記2次元骨格構造の正規化値の変化に基づいて、前記クエリ状態を検索する、

請求項 17 に記載の画像処理装置。

【請求項 19】

取得される2次元画像に基づいて人物の2次元骨格構造を検出し、

前記検出された2次元骨格構造に基づいて2次元画像空間上の前記人物の直立時の高さを推定し、

前記推定された直立時の高さに基づいて前記検出された2次元骨格構造を正規化し、
前記正規化では、前記2次元骨格構造に含まれる特徴点の高さを前記直立時の高さにより正規化し、

前記正規化では、前記2次元骨格構造内の基準点に対する前記特徴点の相対的な高さを前記直立時の高さにより正規化する、

画像処理方法。

【請求項 20】

取得される2次元画像に基づいて人物の2次元骨格構造を検出し、

前記検出された2次元骨格構造に基づいて2次元画像空間上の前記人物の直立時の高さを推定し、

前記推定された直立時の高さに基づいて前記検出された2次元骨格構造を正規化し、
前記推定では、前記2次元画像の撮像パラメータに基づいて前記2次元骨格構造にフィッティングさせた3次元骨格モデルに基づいて、前記人物の直立時の高さを推定する、
画像処理方法。

【請求項 21】

取得される2次元画像に基づいて人物の2次元骨格構造を検出し、

前記検出された2次元骨格構造に基づいて2次元画像空間上の前記人物の直立時の高さを推定し、

前記推定された直立時の高さに基づいて前記検出された2次元骨格構造を正規化し、
前記正規化では、前記2次元骨格構造に含まれる特徴点の高さを前記直立時の高さにより正規化し、

前記正規化では、前記2次元骨格構造内の基準点に対する前記特徴点の相対的な高さを前記直立時の高さにより正規化する、

処理をコンピュータに実行させるための画像処理プログラム。

10

20

30

40

50

【請求項 2 2】

取得される 2 次元画像に基づいて人物の 2 次元骨格構造を検出し、
前記検出された 2 次元骨格構造に基づいて 2 次元画像空間上の前記人物の直立時の高さを推定し、
前記推定された直立時の高さに基づいて前記検出された 2 次元骨格構造を正規化し、
前記推定では、前記 2 次元画像の撮像パラメータに基づいて前記 2 次元骨格構造にフィッティングさせた 3 次元骨格モデルに基づいて、前記人物の直立時の高さを推定する、
処理をコンピュータに実行させるための画像処理プログラム。

【発明の詳細な説明】

【技術分野】

10

【0001】

本発明は、画像処理装置、画像処理方法及び画像処理プログラムが格納された非一時的なコンピュータ可読媒体に関する。

【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、人物を撮像した画像から人物の姿勢を推定し、推定された姿勢に類似した姿勢を含む画像を検索する技術が開示されている。特許文献 2 には、人物を撮像した画像から人物の状態を検出し、検出された状態に基づいて人物の身長を推定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。

20

【先行技術文献】

【特許文献】

【0003】

【文献】特開 2019 - 091138 号公報

【文献】国際公開第 2019 / 064375 号

【非特許文献】

【0004】

【文献】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

30

【発明の概要】

【発明が解決しようとする課題】

【0005】

上記特許文献 1 のような関連する技術では、類似した姿勢を検索するため、人物の姿勢の特徴を示す特徴量を用いているが、特定の観点からの検索しか考慮されていないため、多様な観点からの検索に対しロバスト性が低い場合がある。このため、関連する技術では、人物の状態の検索や分類等の状態認識処理のロバスト性が低いという問題がある。

【0006】

40

本開示は、このような課題に鑑み、人物の状態認識処理のロバスト性を向上することが可能な画像処理装置、画像処理方法及び画像処理プログラムが格納された非一時的なコンピュータ可読媒体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】

【0007】

本開示に係る画像処理装置は、取得される 2 次元画像に基づいて人物の 2 次元骨格構造を検出する骨格検出手段と、前記検出された 2 次元骨格構造に基づいて 2 次元画像空間上の前記人物の直立時の高さを推定する推定手段と、前記推定された直立時の高さに基づいて前記検出された 2 次元骨格構造を正規化する正規化手段と、を備えるものである。

【0008】

50

本開示に係る画像処理方法は、取得される２次元画像に基づいて人物の２次元骨格構造を検出し、前記検出された２次元骨格構造に基づいて２次元画像空間上の前記人物の直立時の高さを推定し、前記推定された直立時の高さに基づいて前記検出された２次元骨格構造を正規化するものである。

【 0 0 0 9 】

本開示に係る画像処理プログラムが格納された非一時的なコンピュータ可読媒体は、取得される２次元画像に基づいて人物の２次元骨格構造を検出し、前記検出された２次元骨格構造に基づいて２次元画像空間上の前記人物の直立時の高さを推定し、前記推定された直立時の高さに基づいて前記検出された２次元骨格構造を正規化する処理をコンピュータに実行させるための画像処理プログラムが格納された非一時的なコンピュータ可読媒体である。

10

【発明の効果】

【 0 0 1 0 】

本開示によれば、人物の状態認識処理のロバスト性を向上することが可能な画像処理装置、画像処理方法及び画像処理プログラムが格納された非一時的なコンピュータ可読媒体を提供することができる。

【図面の簡単な説明】

【 0 0 1 1 】

【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。

20

【図 3】実施の形態 1 に係る画像処理装置の他の構成を示す構成図である。

【図 4】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 5】実施の形態 1 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 6】実施の形態 1 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 7】実施の形態 1 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 8】実施の形態 1 に係る正規化方法を示すフローチャートである。

【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 に係る人体モデルを示す図である。

【図 11】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 12】実施の形態 1 に係る骨格構造の検出例を示す図である。

30

【図 13】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 14】実施の形態 1 に係る人体モデルを示す図である。

【図 15】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 16】実施の形態 1 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 17】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 18】実施の形態 1 に係る３次元人体モデルを示す図である。

【図 19】実施の形態 1 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 20】実施の形態 1 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 21】実施の形態 1 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 22】実施の形態 1 に係る正規化方法を説明するための図である。

40

【図 23】実施の形態 1 に係る正規化方法を説明するための図である。

【図 24】実施の形態 1 に係る正規化方法を説明するための図である。

【図 25】実施の形態 1 に係る画像処理方法の効果を説明するための図である。

【図 26】実施の形態 1 に係る画像処理方法の効果を説明するための図である。

【図 27】実施の形態 1 に係る画像処理方法の効果を説明するための図である。

【図 28】実施の形態 1 に係る画像処理方法の効果を説明するための図である。

【図 29】実施の形態に係るコンピュータのハードウェアの概要を示す構成図である。

【発明を実施するための形態】

【 0 0 1 2 】

以下、図面を参照して実施の形態について説明する。各図面においては、同一の要素に

50

は同一の符号が付されており、必要に応じて重複説明は省略される。

【0013】

(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。

【0014】

しかしながら、このような関連する技術では、機械学習による事前準備が前提となるため、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。すなわち、関連する技術では、予め人物の状態が定義された画像(座り込んでいる画像や手を挙げている画像等)を大量に学習する必要がある。そうすると、事前に定義することができない人物の状態を監視対象としたい場合には、機械学習では対応することが困難である。

10

【0015】

また、関連する技術では、多様な観点からの検索が考慮されていない。例えば、同じ姿勢であれば、どの向きの姿勢でも同じ姿勢と判定することや、画角の影響をなるべく抑えることをユーザが望む場合がある。ディープラーニングを用いることにより姿勢情報を画角や人物の向きにロバストな特徴量に変換する方法もあるが、この方法では、上記のように大量の学習データが必要であり、未知の(学習データにない)姿勢に弱いという問題がある。また、例えば人物の体の一部が隠れている場合に検索を行うことができず、部分一致等の柔軟な検索にも対応することができない。

20

【0016】

そこで、発明者らは、人物の状態を機械学習するような事前準備を行うことなく、柔軟に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、人物の向きや画角等による影響に対しロバスト性の高い状態認識処理を可能とする。

【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。

30

【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、推定部12、正規化部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人物の2次元骨格構造を検出する。推定部12は、骨格検出部11により検出された2次元骨格構造に基づいて、2次元画像空間上の人物の直立時の高さを推定する。正規化部13は、推定部12により推定された直立時の高さに基づいて、骨格検出部11により検出された2次元骨格構造を正規化する。

40

【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次元骨格構造から推定する人物の直立時の高さに基づいて2次元骨格構造を正規化する。これにより、この正規化された正規化値(特徴量)を用いた検索等の状態認識処理において、人物の向き等に対するロバスト性を向上することができる。

【0020】

(実施の形態1)

50

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を正規化するシステムである。また、正規化された骨格構造の特徴量を用いて、人物の状態の分類や検索等を行うことができる。

【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。

10

【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、正規化部 104 が正規化した特徴量等を記憶する。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。

20

【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、身長算出部 103、正規化部 104 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。なお、必要に応じて、ユーザからの操作を入力する入力部や、処理結果等を表示する表示部を備えていてもよい。

【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。

30

【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 Open Pose 等の骨格推定技術を用いる。

40

【0026】

身長算出部 (身長推定部) 103 は、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 103 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。

【0027】

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

50

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、各ボーンの長さと全身の長さ（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。

【0028】

正規化部 104 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。この例では、正規化部 104 は、骨格構造に含まれる各キーポイント（特徴点）の画像上の高さを、身長画素数で正規化する。正規化部 104 は、正規化され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。例えば、高さ方向（上下方向または縦方向）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。

【0029】

なお、画像処理装置 100 は、図 3 に示すように、さらに、分類部 105 及び検索部 106 のいずれか、または両方を備えていてもよい。分類部 105 及び検索部 106 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。分類部 105 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 105 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。また、検索部 106 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）の特徴量と類似度の高い骨格構造を検索する。検索部 106 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。

【0030】

類似度は、骨格構造の特徴量間の距離である。分類部 105 及び検索部 106 は、骨格構造の全体の特徴量の類似度により分類及び検索してもよいし、骨格構造の一部の特徴量の類似度により分類及び検索してもよい。また、各画像における人物の骨格構造の特徴量に基づいて人物の姿勢を分類及び検索してもよいし、時間的に連続する複数の画像における人物の骨格構造の特徴量の変化に基づいて人物の行動を分類及び検索してもよい。すなわち、分類部 105 及び検索部 106 は、骨格構造の特徴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及び検索できる。

【0031】

図 4 ~ 図 8 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 4 は、画像処理装置 100 における画像取得から正規化情報を出力するまでの流れを示し、図 5

10

20

30

40

50

～図7は、図4の身長画素数算出処理(S103)の具体例1～3の流れを示し、図8は、図4の正規化処理(S104)の流れを示している。

【0032】

図4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、骨格構造から正規化処理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。

【0033】

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9は、骨格構造の検出例を示している。図9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。

10

【0034】

図10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11～図13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10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。この例では、図10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。

20

【0035】

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10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A62と右膝A71及び左膝A72をそれぞれ結ぶボーンB61及びボーンB62、右膝A71及び左膝A72と右足A81及び左足A82をそれぞれ結ぶボーンB71及びボーンB72を検出する。

30

【0036】

図11は、直立した状態の人物を検出する例である。図11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、右足のボーンB61及びボーンB71は左足のボーンB62及びボーンB72よりも多少折れ曲がっている。

40

【0037】

図12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12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は大きく折れ曲がり、かつ、重なっている。

【0038】

図13は、寝込んでいる状態の人物を検出する例である。図13では、寝込んでいる人

50

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。

【 0 0 3 9 】

続いて、画像処理装置 1 0 0 は、図 4 に示すように、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 (S 1 0 3)。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。

【 0 0 4 0 】

< 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 5 に示すように、身長算出部 1 0 3 は、各ボーンの長さを取得し (S 1 1 1)、取得した各ボーンの長さを合計する (S 1 1 2)。

【 0 0 4 1 】

身長算出部 1 0 3 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 1 0 のボーンのうち、ボーン B 1 (長さ L 1)、ボーン B 5 1 (長さ L 2 1)、ボーン B 6 1 (長さ L 3 1) 及びボーン B 7 1 (長さ L 4 1)、もしくは、ボーン B 1 (長さ L 1)、ボーン B 5 2 (長さ L 2 2)、ボーン B 6 2 (長さ L 3 2) 及びボーン B 7 2 (長さ L 4 2) 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、 $L 1 + L 2 1 + L 3 1 + L 4 1$ 、もしくは、 $L 1 + L 2 2 + L 3 2 + L 4 2$ 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。

【 0 0 4 2 】

図 1 1 の例では、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ている。これらのボーンの合計である、 $L 1 + L 2 1 + L 3 1 + L 4 1$ 、及び、 $L 1 + L 2 2 + L 3 2 + L 4 2$ を求め、例えば、検出されたボーンの長さが長い左足側の $L 1 + L 2 2 + L 3 2 + L 4 2$ に補正定数を乗じた値を身長画素数とする。

【 0 0 4 3 】

図 1 2 の例では、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 が重なっている。これらのボーンの合計である、 $L 1 + L 2 1 + L 3 1 + L 4 1$ 、及び、 $L 1 + L 2 2 + L 3 2 + L 4 2$ を求め、例えば、検出されたボーンの長さが長い右足側の $L 1 + L 2 1 + L 3 1 + L 4 1$ に補正定数を乗じた値を身長画素数とする。

【 0 0 4 4 】

図 1 3 の例では、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 が重なっている。これらのボーンの合計である、 $L 1 + L 2 1 + L 3 1 + L 4 1$ 、及び、 $L 1 + L 2 2 + L 3 2 + L 4 2$ を求め、例えば、検出されたボーンの長さが長い左足側の $L 1 + L 2 2 + L 3 2 + L 4 2$ に補正定数を乗じた値を身長画素数とする。

【 0 0 4 5 】

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

10

20

30

40

50

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。

【 0 0 4 6 】

< 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。

【 0 0 4 7 】

図 1 4 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) 3 0 1 である。図 1 4 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 3 0 1 の各ボーンに対応付ける。例えば、頭のボーン B 1 の長さは全身の長さ $\times 0.2$ (2 0 %) であり、右手のボーン B 4 1 の長さは全身の長さ $\times 0.15$ (1 5 %) であり、右足のボーン B 7 1 の長さは全身の長さ $\times 0.25$ (2 5 %) である。このような人体モデル 3 0 1 の情報をデータベース 1 1 0 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。

【 0 0 4 8 】

具体例 2 では、図 6 に示すように、身長算出部 1 0 3 は、各ボーンの長さを取得する (S 1 2 1) 。身長算出部 1 0 3 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 1 5 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 B 2 1、B 2 2、B 3 1、B 4 1、B 5 1、B 5 2、B 6 1、B 6 2、B 7 1、B 7 2 の各長さを取得する。

【 0 0 4 9 】

続いて、身長算出部 1 0 3 は、図 6 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 (S 1 2 2) 。身長算出部 1 0 3 は、図 1 4 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 3 0 1 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 B 4 1 の長さが全身の長さ $\times 0.15$ であるため、ボーン B 4 1 の長さ $\div 0.15$ によりボーン B 4 1 に基づいた身長画素数を求める。また、右足のボーン B 7 1 の長さが全身の長さ $\times 0.25$ であるため、ボーン B 7 1 の長さ $\div 0.25$ によりボーン B 7 1 に基づいた身長画素数を求める。

【 0 0 5 0 】

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。

【 0 0 5 1 】

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、2 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 2 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。

【 0 0 5 2 】

10

20

30

40

50

続いて、身長算出部 103 は、図 6 に示すように、身長画素数の最適値を算出する (S123)。身長算出部 103 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 16 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 16 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。

10

【0053】

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。

【0054】

<具体例 3>

20

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。

【0055】

具体例 3 では、図 7 に示すように、身長算出部 103 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 (S131)。身長算出部 103 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。

【0056】

30

続いて、身長算出部 103 は、3次元人体モデルの配置及び高さを調整する (S132)。身長算出部 103 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。

【0057】

図 17 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 18 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 19 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。

40

【0058】

なお、このとき用意する 3次元人体モデル 402 は、図 19 のように、2次元骨格構造

50

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。

【0059】

続いて、身長算出部103は、図7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(S133)。身長算出部103は、図20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401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。

10

【0060】

続いて、身長算出部103は、図7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(S134)。身長算出部103は、図21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。

20

【0061】

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。

30

【0062】

<正規化処理>

図4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(S104)を行い、正規化処理の結果得られた正規化値をデータベース110に格納する(S105)。正規化処理では、図8に示すように、正規化部104は、キーポイント高さを算出する(S141)。正規化部104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10の例で、首のキーポイントA2の高さ(y_1)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。

40

【0063】

続いて、正規化部104は、正規化のための基準点を特定する(S142)。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。また、キーポイントに限らず、任意の座標(例えば骨格構造の

50

中心座標等)を基準点としてもよい。

【0064】

続いて、正規化部104は、キーポイント高さ(y_i)を身長画素数で正規化する(S143)。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、正規化部104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。

10

【数1】

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

【0065】

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。

20

【数2】

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

30

【0066】

図22は、正規化部104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。首よりも高い頭のキーポイントA1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23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24のように、図22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。

40

50

【 0 0 6 7 】

< 実施の形態 1 の効果 >

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(特徴点)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。

【 0 0 6 8 】

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。例えば、図25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、特徴量は大きく変化しない。このため、骨格構造501~503では、分類や検索時に同じ姿勢であると判断することができる。

10

【 0 0 6 9 】

図26は検索結果の例を示しており、検索クエリQ10、Q20、Q30のそれぞれの検索結果Q11、Q21、Q31を示している。検索クエリQ10のように座っている姿勢を検索すると、検索結果Q11のように様々な方向を向いて座っている姿勢が似ている姿勢として判断される。同様に、検索クエリQ20のように立ち上がって右手を水平に伸ばしている姿勢を検索すると、検索結果Q21のように立ち上がって様々な方向に右手を伸ばしている姿勢が似ている姿勢として判断される。また、検索クエリQ30のように寝込んでいる姿勢を検索すると、検索結果Q31のように様々な方向を向いて寝込んでいる姿勢が似ている姿勢として判断される。

20

【 0 0 7 0 】

また、本実施の形態の特徴量は、各キーポイントを正規化した値であるため、体の一部が隠れている画像に対してロバスト性が高い。例えば、図27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できる。このため、骨格構造511及び512では、分類や検索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28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。

30

【 0 0 7 1 】

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。

【 0 0 7 2 】

なお、上述の実施形態における各構成は、ハードウェア又はソフトウェア、もしくはその両方によって構成され、1つのハードウェア又はソフトウェアから構成してもよいし、複数のハードウェア又はソフトウェアから構成してもよい。画像処理装置10及び100の機能(処理)を、図29に示すような、CPU(Central Processing Unit)等のプロセッサ21及び記憶装置であるメモリ22を有するコンピュータ20により実現してもよい。例えば、メモリ22に実施形態における方法を行うためのプログラム(画像処理プログラム)を格納し、各機能を、メモリ22に格納されたプログラムをプロセッサ21で実行することにより実現してもよい。

40

【 0 0 7 3 】

これらのプログラムは、様々なタイプの非一時的なコンピュータ可読媒体(non-transi

50

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

10

【0074】

また、本開示は上記実施の形態に限られたものではなく、趣旨を逸脱しない範囲で適宜変更することが可能である。例えば、上記では人物の骨格構造を検出したが、骨格構造を有する人物以外の動物 (哺乳類、爬虫類、鳥類、両生類、魚類等) を検出対象としてもよい。

【0075】

以上、実施の形態を参照して本開示を説明したが、本開示は上記実施の形態に限定されるものではない。本開示の構成や詳細には、本開示のスコープ内で当業者が理解し得る様々な変更をすることができる。

20

【0076】

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

(付記 1)

取得される2次元画像に基づいて人物の2次元骨格構造を検出する骨格検出手段と、
前記検出された2次元骨格構造に基づいて2次元画像空間上の前記人物の直立時の高さを推定する推定手段と、
前記推定された直立時の高さに基づいて前記検出された2次元骨格構造を正規化する正規化手段と、
を備える画像処理装置。

30

(付記 2)

前記正規化手段は、前記2次元骨格構造に含まれる特徴点の高さを前記直立時の高さにより正規化する、
付記1に記載の画像処理装置。

(付記 3)

前記特徴点の高さは、前記2次元画像空間を表すX-Y座標におけるY軸方向の高さである、
付記2に記載の画像処理装置。

(付記 4)

前記特徴点の高さは、前記2次元画像の撮像パラメータに基づいて、実世界の3次元空間における基準面に対する鉛直方向を、前記2次元画像空間に投影した鉛直投影方向の高さである、
付記2に記載の画像処理装置。

40

(付記 5)

前記正規化手段は、前記2次元骨格構造内の基準点に対する前記特徴点の相対的な高さを前記直立時の高さにより正規化する、
付記2乃至4のいずれかに記載の画像処理装置。

(付記 6)

前記基準点は、前記2次元骨格構造における中心よりも前記2次元画像空間で上の点で

50

ある、

付記 5 に記載の画像処理装置。

(付記 7)

前記基準点は、前記 2 次元骨格構造における首部または頭部の特徴点である、

付記 6 に記載の画像処理装置。

(付記 8)

前記推定手段は、前記 2 次元骨格構造に含まれる 2 次元画像空間上の骨の長さに基づいて、前記人物の直立時の高さを推定する、

付記 1 乃至 7 のいずれかに記載の画像処理装置。

(付記 9)

前記推定手段は、前記 2 次元骨格構造に含まれる足部から頭部までの前記骨の長さの合計に基づいて、前記人物の直立時の高さを推定する、

付記 8 に記載の画像処理装置。

(付記 10)

前記推定手段は、前記骨の長さと 2 次元画像空間上の前記人物の全身の長さとの関係を示す 2 次元骨格モデルに基づいて、前記人物の直立時の高さを推定する、

付記 8 に記載の画像処理装置。

(付記 11)

前記推定手段は、前記 2 次元画像の撮像パラメータに基づいて前記 2 次元骨格構造にフイッティングさせた 3 次元骨格モデルに基づいて、前記人物の直立時の高さを推定する、

付記 1 乃至 7 のいずれかに記載の画像処理装置。

(付記 12)

前記骨格検出手段は、前記 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し、前記正規化手段は、前記複数の人物の 2 次元骨格構造を正規化し、

前記正規化された複数の 2 次元骨格構造の正規化値に基づいて、前記複数の人物の状態の認識処理を行う認識手段をさらに備える、

付記 1 乃至 11 のいずれかに記載の画像処理装置。

(付記 13)

前記認識手段は、前記認識処理として、前記複数の人物の状態を分類する、

付記 12 に記載の画像処理装置。

(付記 14)

前記認識手段は、前記 2 次元骨格構造の全体または一部の正規化値に基づいて、前記複数の人物の状態を分類する、

付記 13 に記載の画像処理装置。

(付記 15)

前記骨格検出手段は、時系列に連続する複数の 2 次元画像から、前記複数の人物の 2 次元骨格構造を検出し、

前記認識手段は、前記複数の 2 次元画像から検出された 2 次元骨格構造の正規化値に基づいて、前記複数の人物の状態を分類する、

付記 13 または 14 に記載の画像処理装置。

(付記 16)

前記認識手段は、前記複数の 2 次元画像における前記 2 次元骨格構造の正規化値の変化に基づいて、前記複数の人物の状態を分類する、

付記 15 に記載の画像処理装置。

(付記 17)

前記認識手段は、前記認識処理として、前記複数の人物の状態からクエリ状態を検索する、

付記 12 に記載の画像処理装置。

(付記 18)

前記認識手段は、前記 2 次元骨格構造の全体または一部の正規化値に基づいて、前記ク

10

20

30

40

50

エリ状態を検索する、

付記 17 に記載の画像処理装置。

(付記 19)

前記骨格検出手段は、時系列に連続する複数の 2 次元画像から、前記複数の人物の 2 次元骨格構造を検出し、

前記認識手段は、前記複数の 2 次元画像から検出された 2 次元骨格構造の正規化値に基づいて、前記クエリ状態を検索する、

付記 17 または 18 に記載の画像処理装置。

(付記 20)

前記認識手段は、前記複数の 2 次元画像における前記 2 次元骨格構造の正規化値の変化に基づいて、前記クエリ状態を検索する、

付記 19 に記載の画像処理装置。

(付記 21)

取得される 2 次元画像に基づいて人物の 2 次元骨格構造を検出し、

前記検出された 2 次元骨格構造に基づいて 2 次元画像空間上の前記人物の直立時の高さを推定し、

前記推定された直立時の高さに基づいて前記検出された 2 次元骨格構造を正規化する、
画像処理方法。

(付記 22)

前記正規化では、前記 2 次元骨格構造に含まれる特徴点の高さを前記直立時の高さにより正規化する、

付記 21 に記載の画像処理方法。

(付記 23)

取得される 2 次元画像に基づいて人物の 2 次元骨格構造を検出し、

前記検出された 2 次元骨格構造に基づいて 2 次元画像空間上の前記人物の直立時の高さを推定し、

前記推定された直立時の高さに基づいて前記検出された 2 次元骨格構造を正規化する、
処理をコンピュータに実行させるための画像処理プログラム。

(付記 24)

前記正規化では、前記 2 次元骨格構造に含まれる特徴点の高さを前記直立時の高さにより正規化する、

付記 23 に記載の画像処理プログラム。

【符号の説明】

【0077】

1 画像処理システム

10 画像処理装置

11 骨格検出部

12 推定部

13 正規化部

20 コンピュータ

21 プロセッサ

22 メモリ

100 画像処理装置

101 画像取得部

102 骨格構造検出部

103 身長算出部

104 正規化部

105 分類部

106 検索部

110 データベース

10

20

30

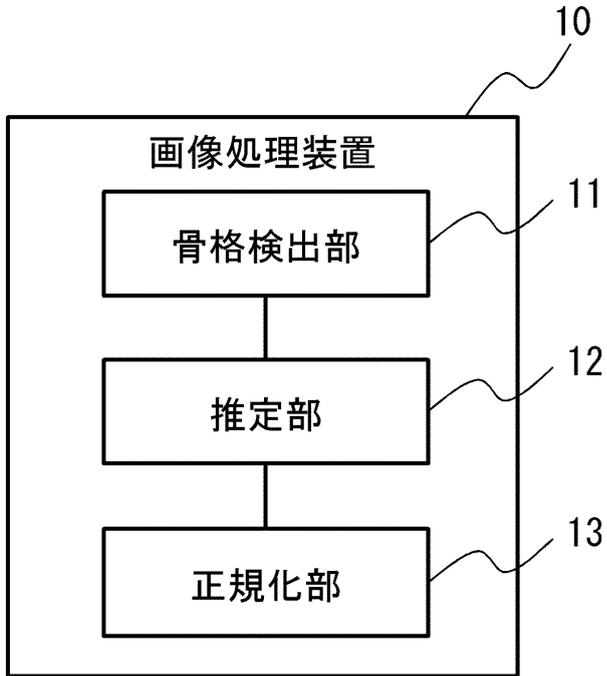
40

50

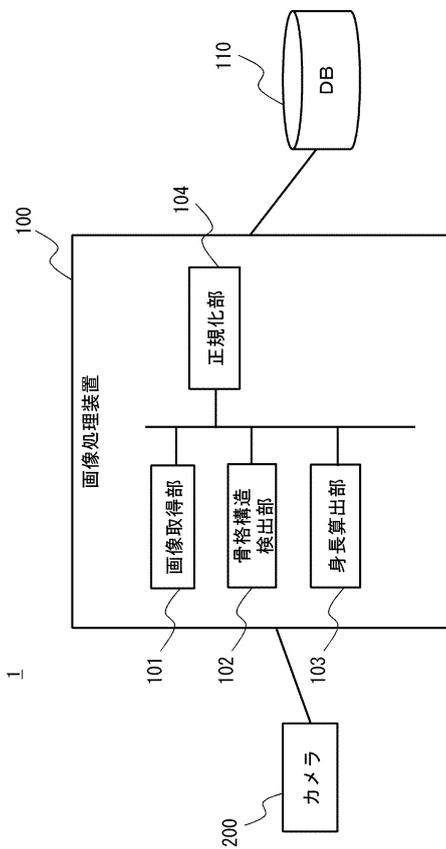
- 200 カメラ
- 300、301 人体モデル
- 401 2次元骨格構造
- 402 3次元人体モデル

【図面】

【図1】



【図2】



10

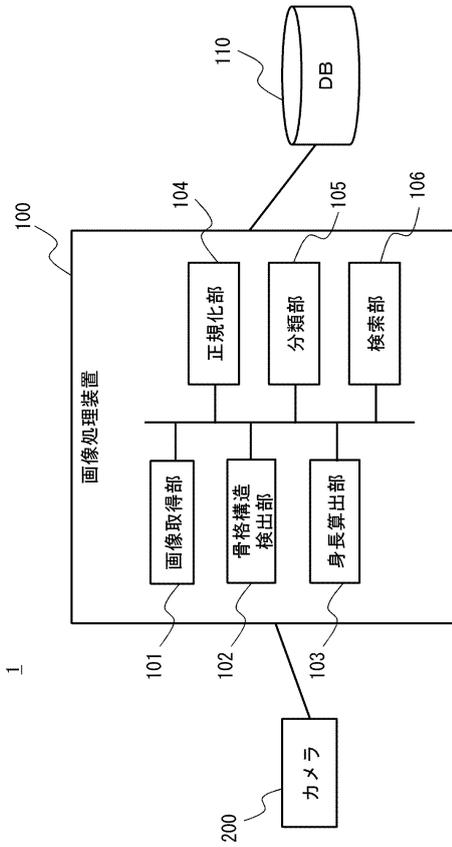
20

30

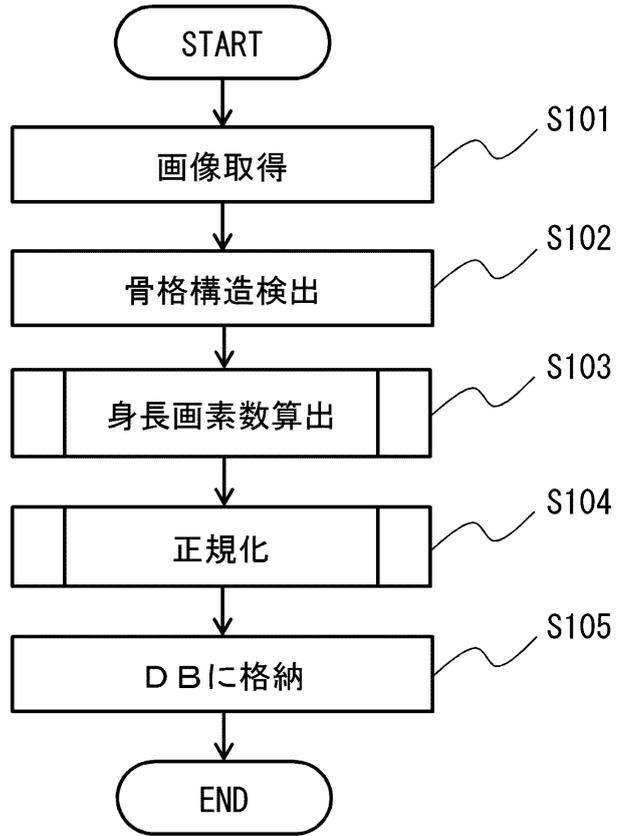
40

50

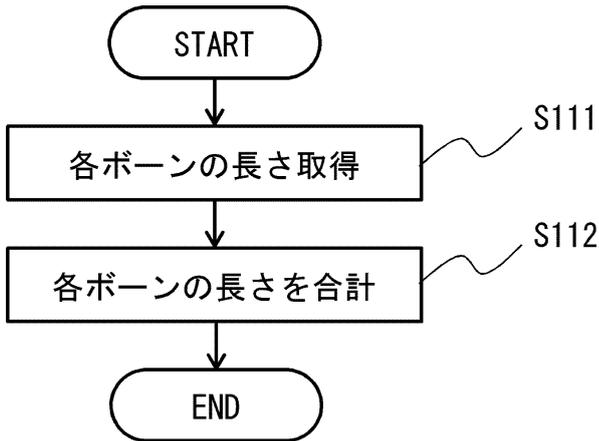
【図3】



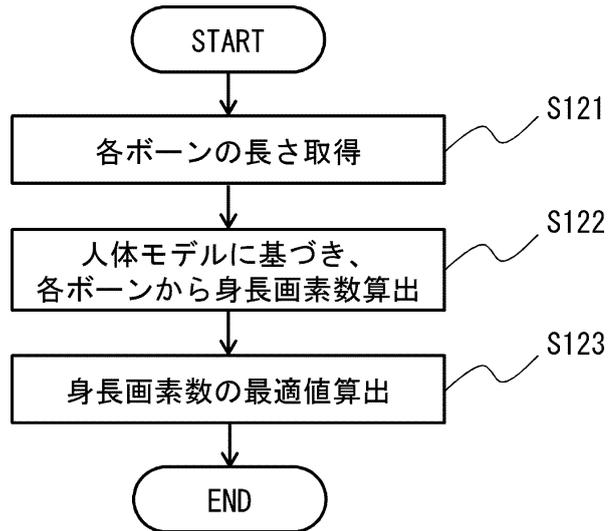
【図4】



【図5】



【図6】



10

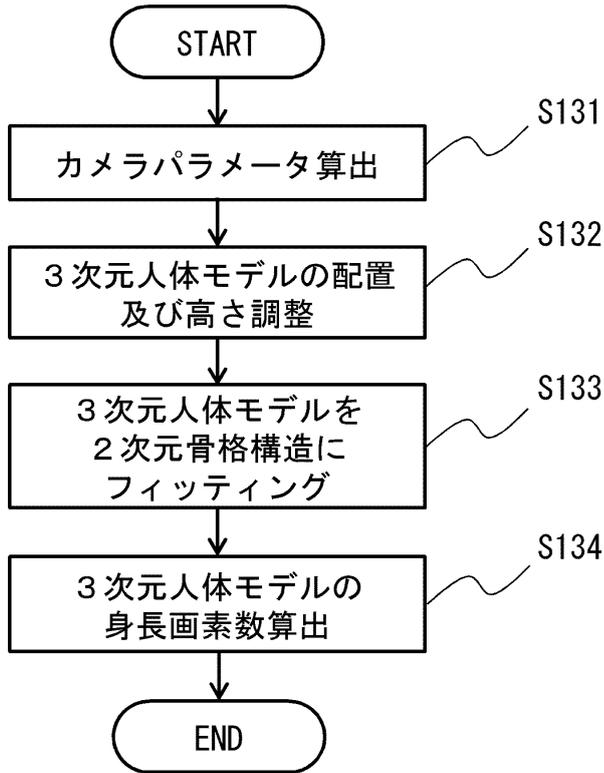
20

30

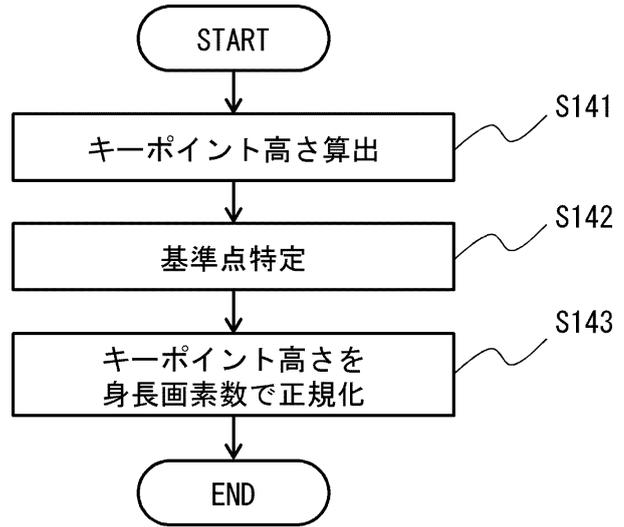
40

50

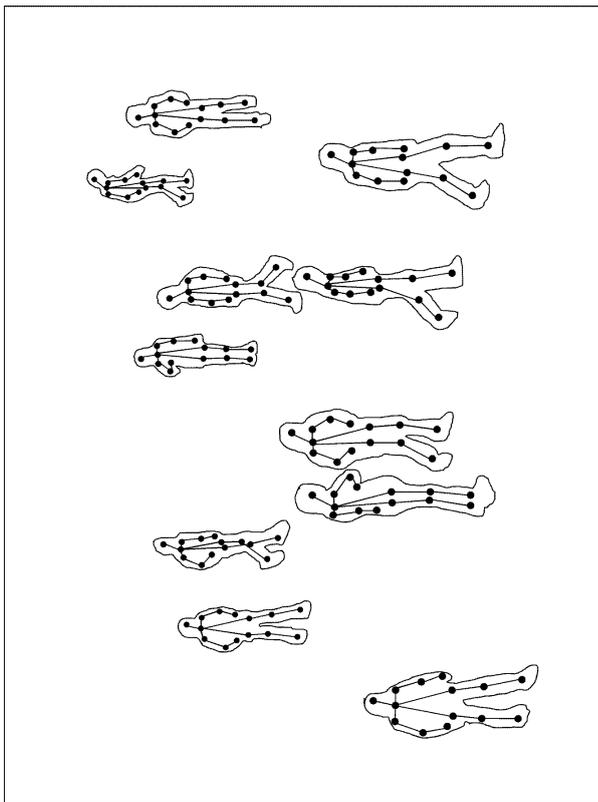
【図7】



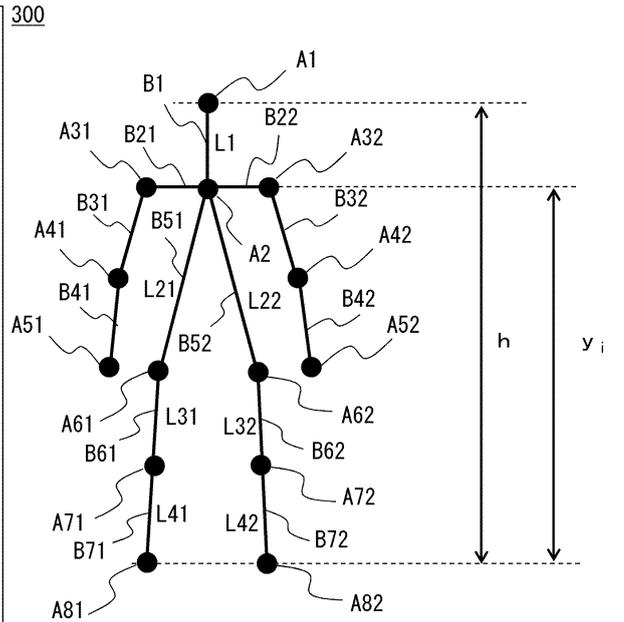
【図8】



【図9】



【図10】



10

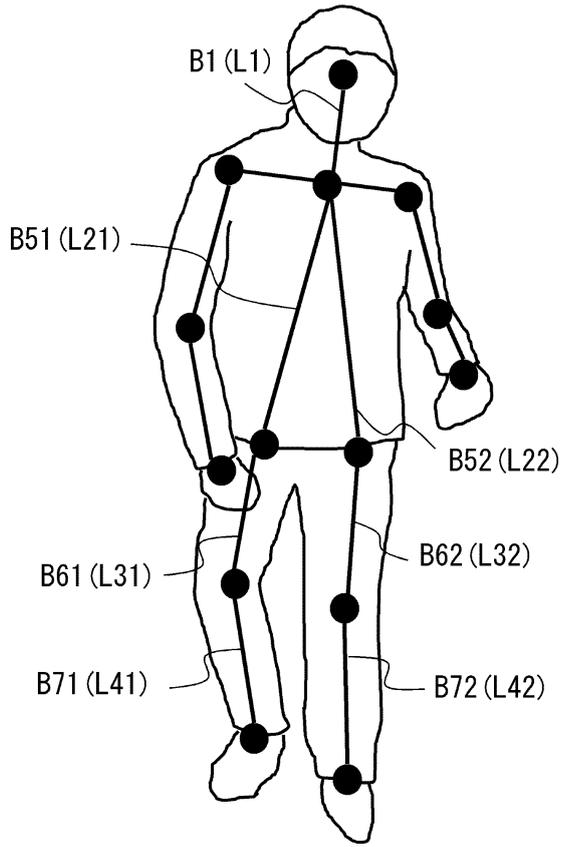
20

30

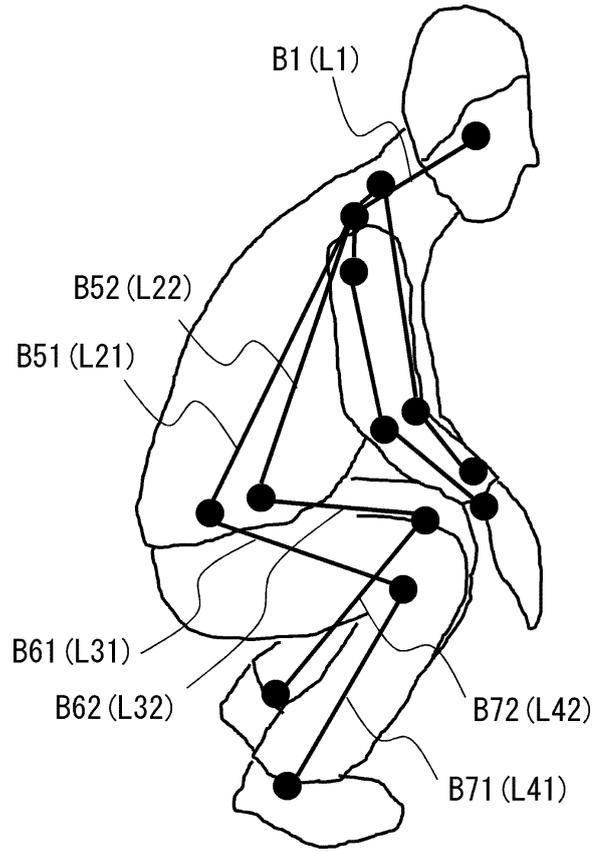
40

50

【図 1 1】



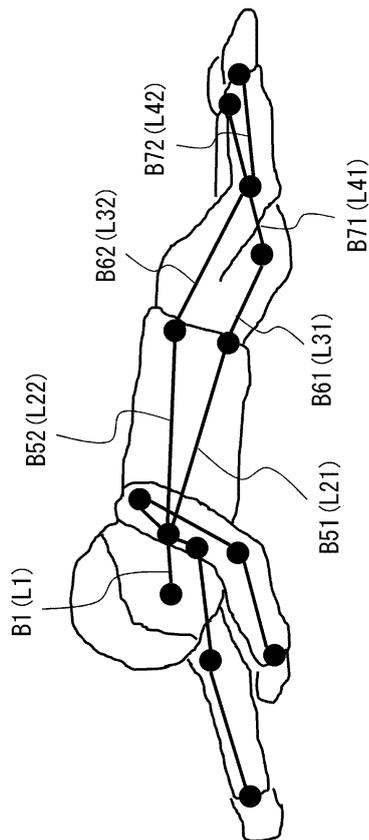
【図 1 2】



10

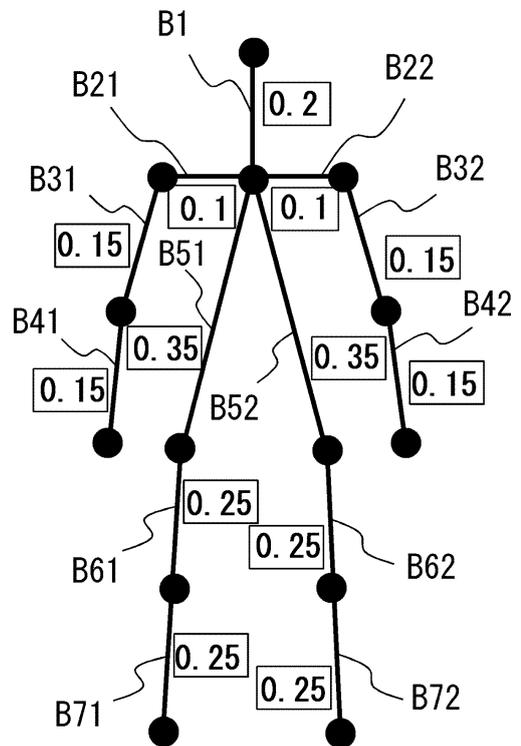
20

【図 1 3】



【図 1 4】

301

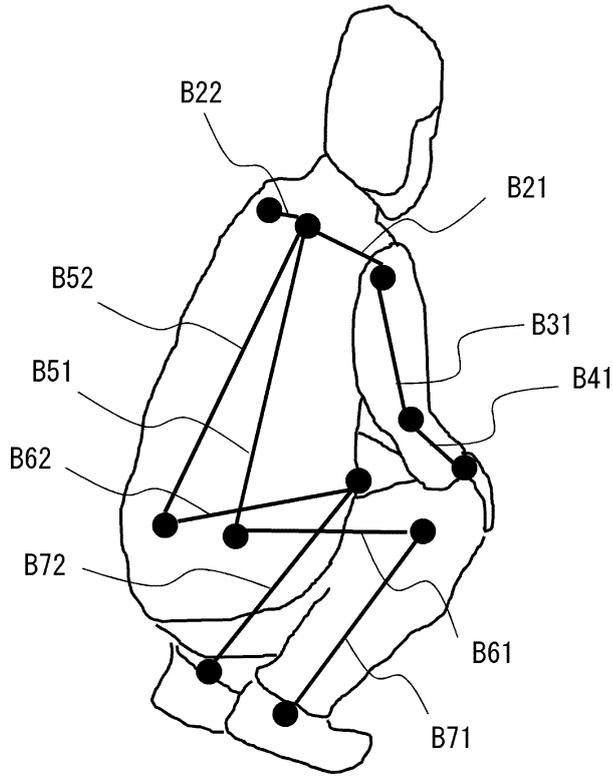


30

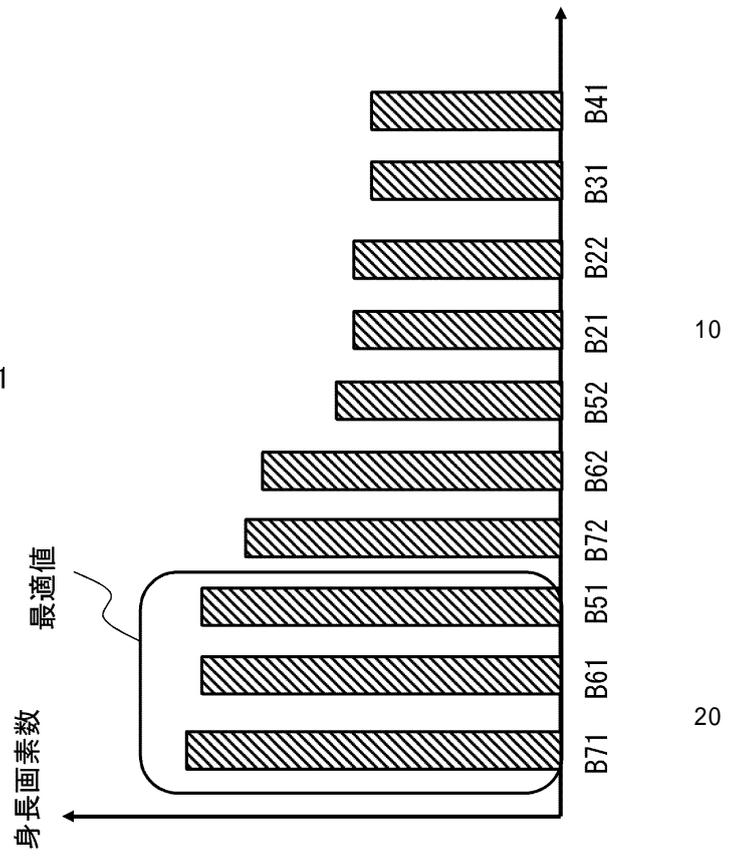
40

50

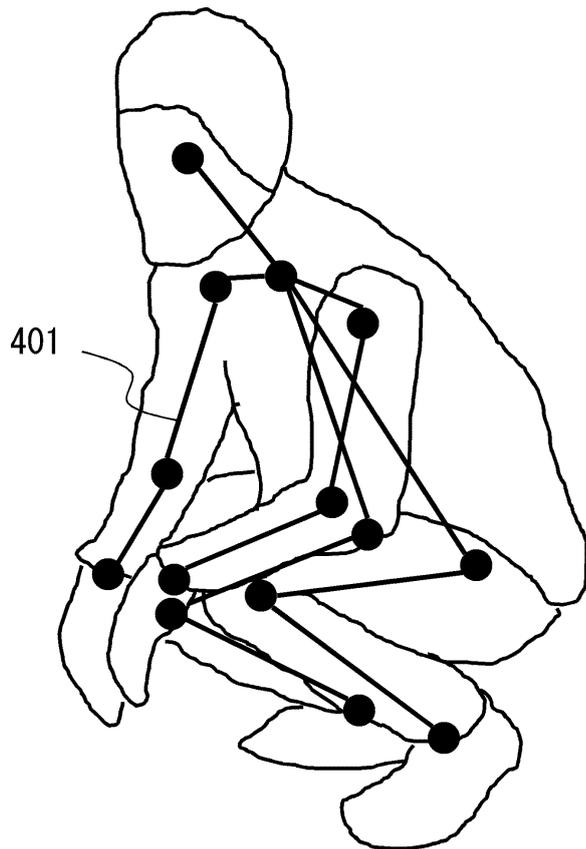
【 図 1 5 】



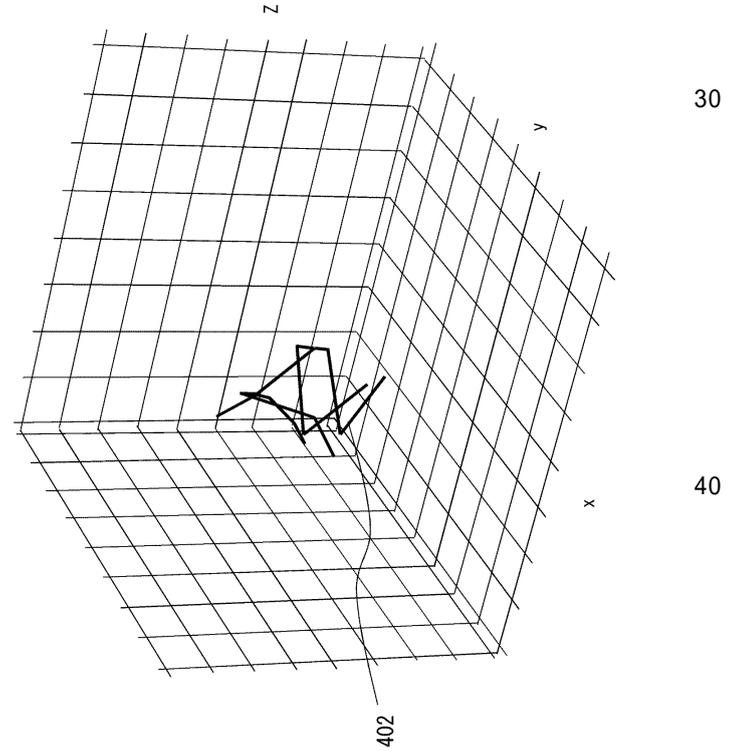
【 図 1 6 】



【 図 1 7 】



【 図 1 8 】



10

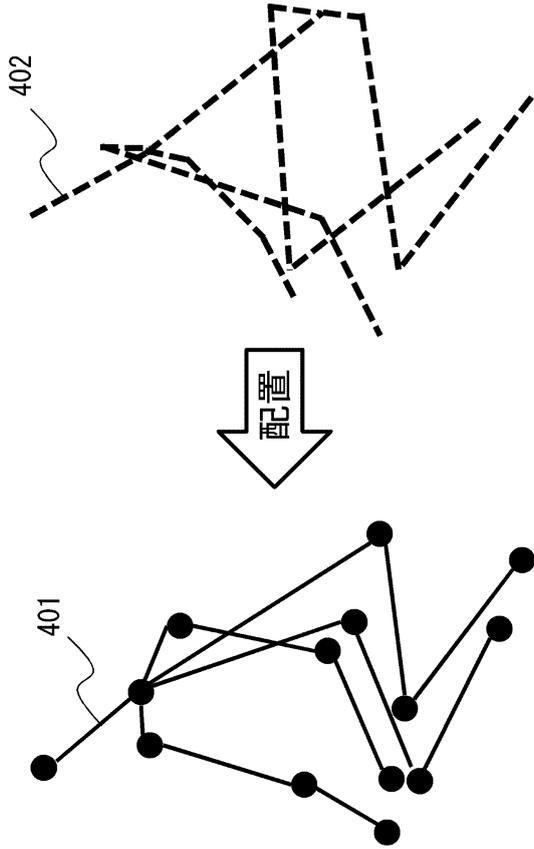
20

30

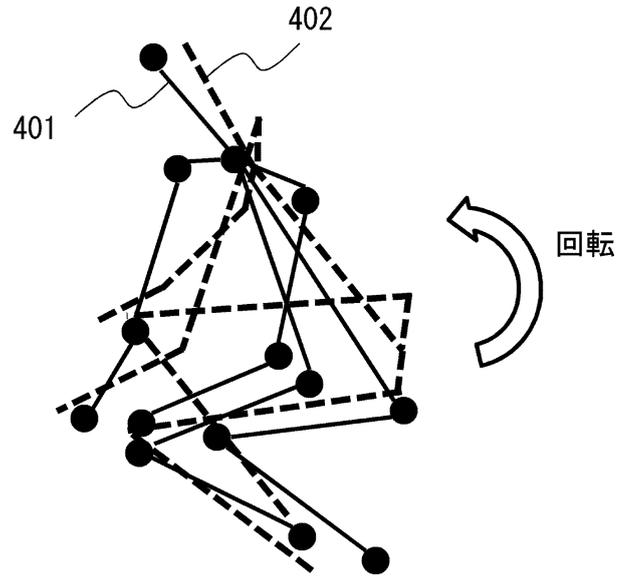
40

50

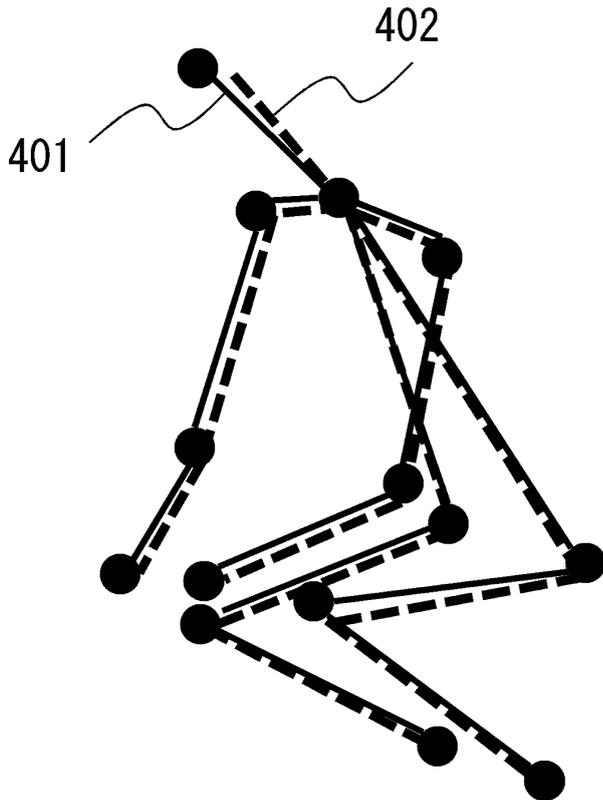
【図 19】



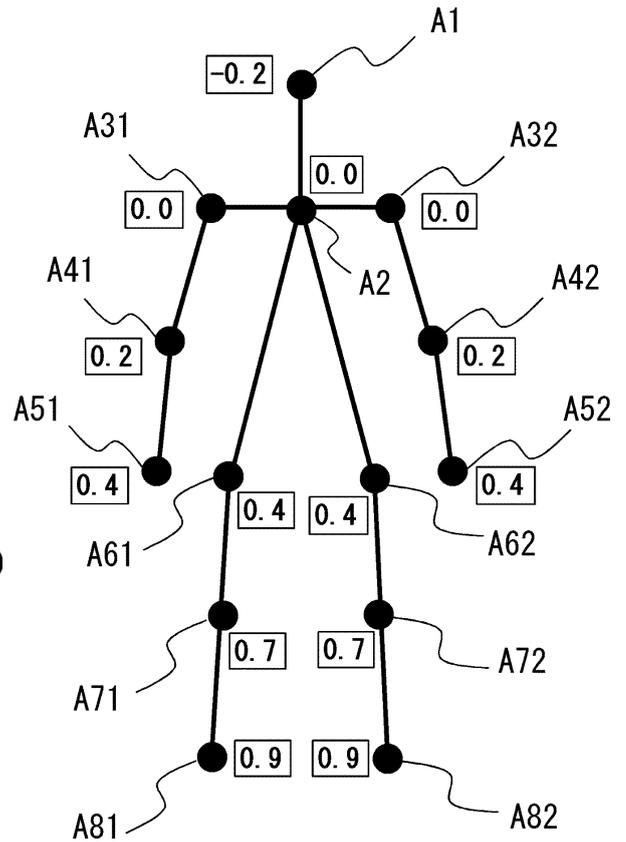
【図 20】



【図 21】



【図 22】



10

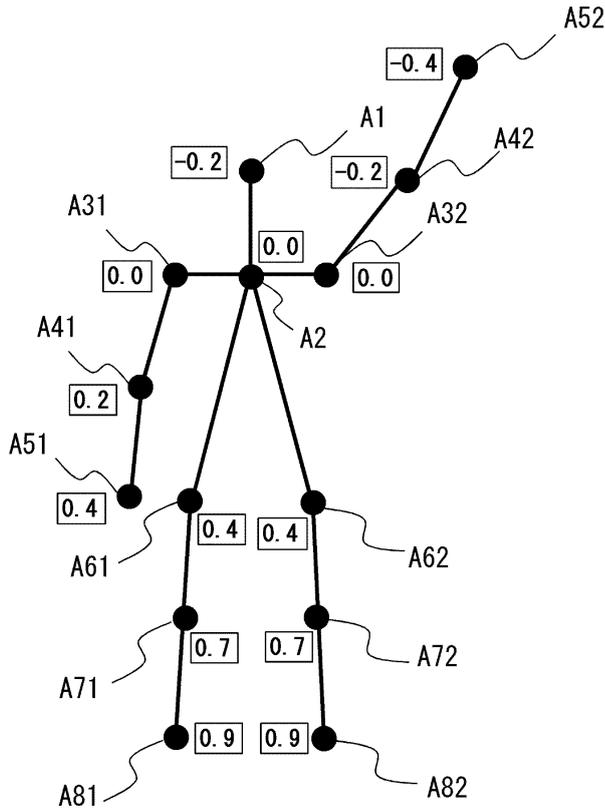
20

30

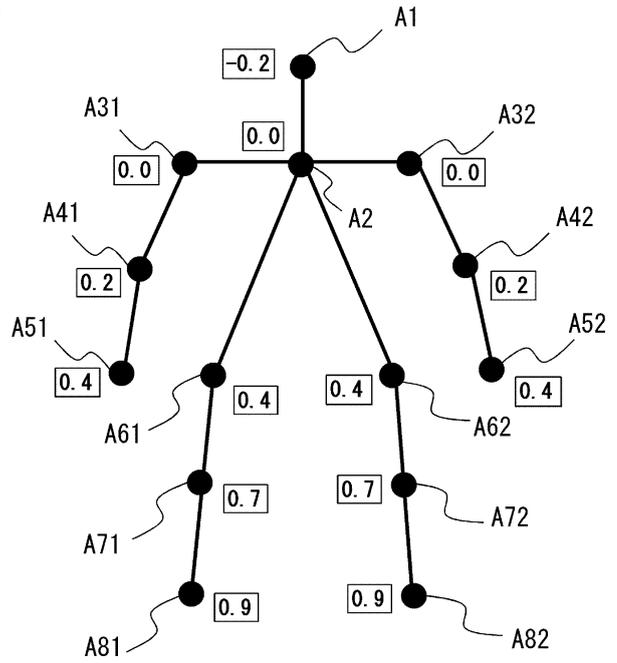
40

50

【 2 3 】



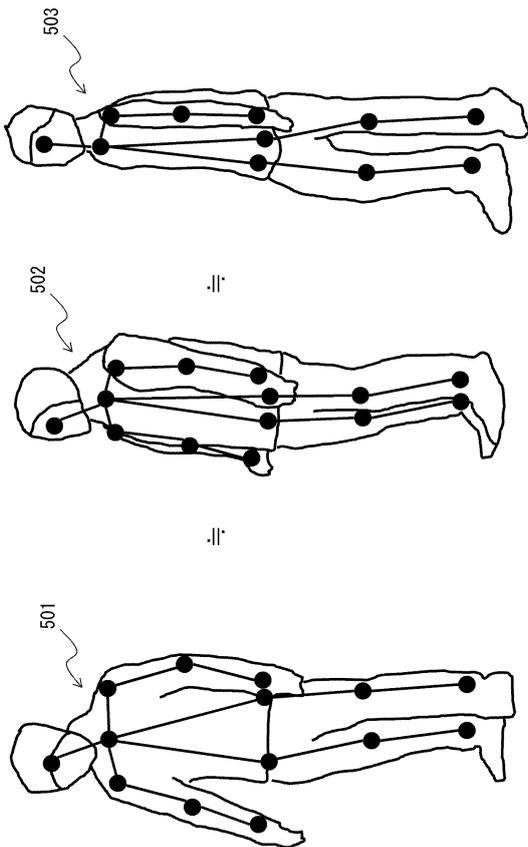
【 2 4 】



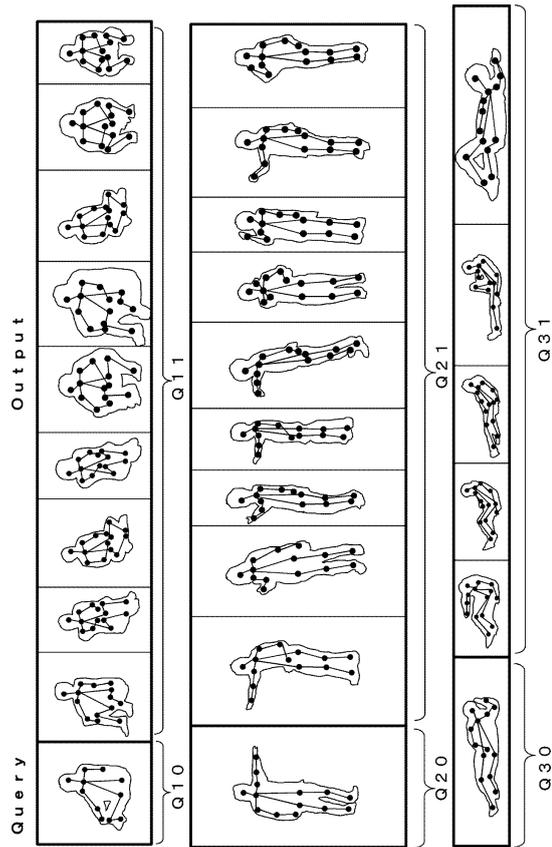
10

20

【 2 5 】



【 2 6 】

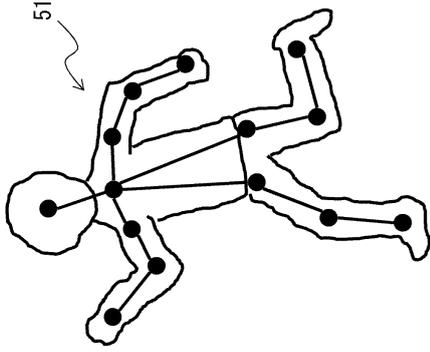


30

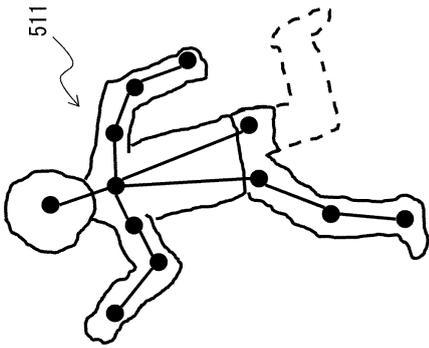
40

50

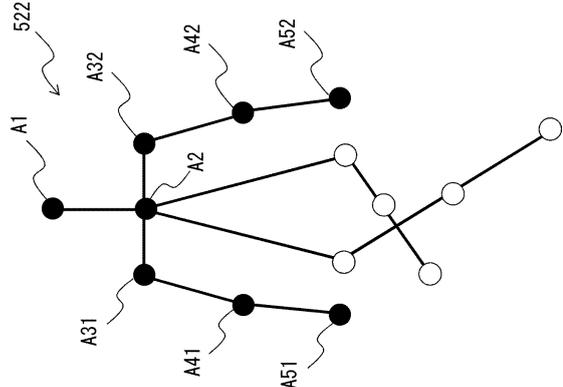
【図 27】



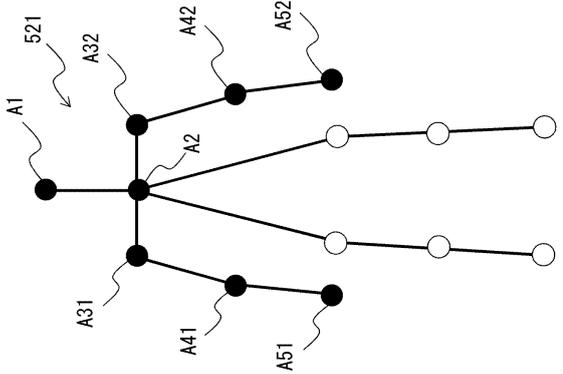
||



【図 28】

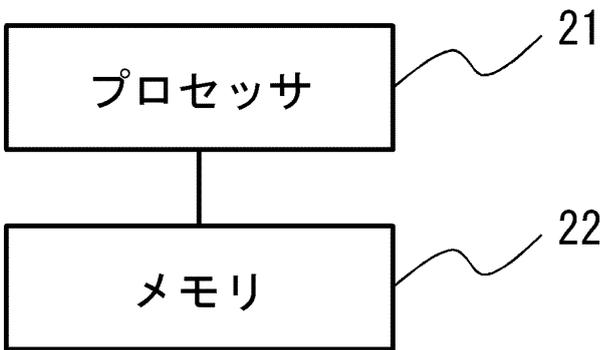


||



【図 29】

20



10

20

30

40

50

フロントページの続き

- (56)参考文献 米国特許出願公開第2019/0251340 (US, A1)
中国特許出願公開第109919141 (CN, A)
特開2006-253932 (JP, A)
特開2019-091138 (JP, A)
米国特許出願公開第2009/0232353 (US, A1)
特許第6534499 (JP, B1)
特開2015-097639 (JP, A)
崎翔太, 外2名, 歩行動作に基づく人物識別に関する研究, 情報処理学会研究報告, 2018年03月17日, Vol.2018-HCI-177 No.31
十代淳貴, 外3名, 動作予測のためのRGB-Dカメラを用いた実時間3次元姿勢推定, ロボティクスメカトロニクス講演会2019講演会論文集, 2019年06月05日
Zhe Cao et al., Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields, [online], 2017年04月14日, <https://arxiv.org/abs/1611.08050>
- (58)調査した分野 (Int.Cl., DB名)
G06T 7/00