(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7501621号
(P7501621)

(45)発行日　令和6年6月18日(2024.6.18)

(24)登録日　令和6年6月10日(2024.6.10)

(51)国際特許分類
Ｇ０６Ｆ　16/73　(2019.01)

ＦＩ
Ｇ０６Ｆ　16/73

請求項の数　7　（全39頁）

(21)出願番号　特願2022-522434(P2022-522434)
(86)(22)出願日　令和2年5月14日(2020.5.14)
(86)国際出願番号　PCT/JP2020/019255
(87)国際公開番号　WO2021/229750
(87)国際公開日　令和3年11月18日(2021.11.18)
審査請求日　令和4年11月4日(2022.11.4)

(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
(74)代理人　100110928
弁理士　速水 進治
(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
(72)発明者　潘 雅冬
東京都港区芝五丁目７番１号 日本電気株式会社内
(72)発明者　川合 諒
東京都港区芝五丁目７番１号 日本電気株式会社内
(72)発明者　劉 健全

最終頁に続く

(54)【発明の名称】　画像選択装置、画像選択方法、およびプログラム

(57)【特許請求の範囲】

【請求項１】

複数の第１フレーム画像を含むクエリ動画を取得するクエリ取得手段と、

前記複数の第１フレーム画像から複数のクエリフレームを選択するクエリフレーム選択手段と、

前記複数のクエリフレームを用いて動画を選択する動画選択手段と、

を備え、

前記クエリフレーム選択手段は、第１の前記クエリフレームの次の前記クエリフレームとして前記第１のクエリフレームからの変化量が第３基準以上となった前記第１フレーム画像を選択する処理を繰り返し、

前記動画選択手段は、前記動画を選択する条件として、少なくとも以下の（１）及び（２）を用いる、画像選択装置。

（１）「前記クエリフレームに対する類似度が第１基準を満たす類似フレーム画像が存在する」という条件が少なくとも２つの前記クエリフレームについて満たされる。

（２）前記少なくとも２つの類似フレーム画像のそれぞれに対応する前記クエリフレームを前記少なくとも２つの類似フレーム画像と同じ順序で並べたときに、当該並び順が、前記クエリ動画における前記少なくとも２つの前記クエリフレームの並び順に一致する。

【請求項２】

請求項１に記載の画像選択装置において、

前記クエリ動画は人物を含んでおり、

前記動画選択手段は、前記類似度として、前記人物の姿勢の類似度を用いる画像選択装置。
【請求項３】
請求項２に記載の画像選択装置において、
前記クエリフレームが満たすべき条件の一つは、当該クエリフレームに含まれる前記人物に関する情報量が第２基準を満たしていることであり、
前記複数の第１フレーム画像のそれぞれは、前記人物の姿勢を示す情報として、当該人物の関節の位置を示す関節情報を含んでおり、
前記第２基準は、前記関節情報に含まれる関節の数が基準を満たしていることである画像選択装置。
【請求項４】
請求項１～３のいずれか一項に記載の画像選択装置において、
前記クエリフレーム選択手段は、ユーザからの入力を用いて前記第３基準を設定する画像選択装置。
【請求項５】
請求項１～３のいずれか一項に記載の画像選択装置において、
前記クエリフレーム選択手段は、前記複数の第１フレーム画像を用いて当該クエリフレームに用いる前記第３基準を設定する画像選択装置。
【請求項６】
コンピュータが、
複数の第１フレーム画像を含むクエリ動画を取得する取得工程と、
前記複数の第１フレーム画像から複数のクエリフレームを選択するクエリフレーム選択工程と、
前記複数のクエリフレームを用いて動画を選択する動画選択工程と、
を行い、
前記クエリフレーム選択工程において、第１の前記クエリフレームの次の前記クエリフレームとして前記第１のクエリフレームからの変化量が第３基準以上となった前記第１フレーム画像を選択する処理を繰り返し、
前記動画選択工程において、前記動画を選択する条件として、少なくとも以下の（１）及び（２）を用いる、画像選択方法。
（１）「前記クエリフレームに対する類似度が第１基準を満たす類似フレーム画像が存在する」という条件が少なくとも２つの前記クエリフレームについて満たされる。
（２）前記少なくとも２つの類似フレーム画像のそれぞれに対応する前記クエリフレームを前記少なくとも２つの類似フレーム画像と同じ順序で並べたときに、当該並び順が、前記クエリ動画における前記少なくとも２つの前記クエリフレームの並び順に一致する。
【請求項７】
コンピュータに、
複数の第１フレーム画像を含むクエリ動画を取得する取得機能と、
前記複数の第１フレーム画像から複数のクエリフレームを選択するクエリフレーム選択機能と、
前記複数のクエリフレームを用いて動画を選択する動画選択機能と、
を持たせ、
前記クエリフレーム選択機能は、第１の前記クエリフレームの次の前記クエリフレームとして前記第１のクエリフレームからの変化量が第３基準以上となった前記第１フレーム画像を選択する処理を繰り返し、
前記動画選択機能は、前記動画を選択する条件として、少なくとも以下の（１）及び（２）を用いるプログラム。
（１）「前記クエリフレームに対する類似度が第１基準を満たす類似フレーム画像が存在する」という条件が少なくとも２つの前記クエリフレームについて満たされる。
（２）前記少なくとも２つの類似フレーム画像のそれぞれに対応する前記クエリフレー

ムを前記少なくとも２つの類似フレーム画像と同じ順序で並べたときに、当該並び順が、前記クエリ動画における前記少なくとも２つの前記クエリフレームの並び順に一致する。
【発明の詳細な説明】
【技術分野】
【０００１】
本発明は、画像選択装置、画像選択方法、およ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】
一方、近年は動画をクエリとして利用し、このクエリに類似する動画を検索することも検討されている。例えば特許文献３には、クエリとなる参照映像を入力すると、登場人物の顔の数、並びに各登場人物の顔の位置、大きさ、及び向きを用いて、類似する映像を検索することが記載されている。
【先行技術文献】
【特許文献】
【０００４】
【文献】特表２０１４－５２２０３５号公報
【文献】特開２００６－２６０４０５号公報
【文献】国際公開第２００６／０２５２７２号
【非特許文献】
【０００５】
【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
【発明の概要】
【発明が解決しようとする課題】
【０００６】
動画をクエリとして、このクエリに類似する動画を選択する場合、選択の精度を上げることは難しい。本発明の目的の一つは、動画をクエリとして、このクエリに類似する動画を選択する場合において、選択の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、複数の第１フレーム画像を含むクエリ動画を取得するクエリ取得手段と、
前記複数の第１フレーム画像から複数のクエリフレームを選択するクエリフレーム選択手段と、
前記複数のクエリフレームを用いて動画を選択する動画選択手段と、
を備え、
前記動画選択手段は、前記動画を選択する条件として、少なくとも以下の（１）及び（２）を用いる、画像選択装置が提供される。
（１）「前記クエリフレームに対する類似度が第１基準を満たす類似フレーム画像が存在する」という条件が少なくとも２つの前記クエリフレームについて満たされる。
（２）前記少なくとも２つの類似フレーム画像のそれぞれに対応する前記クエリフレー

ムを前記少なくとも２つの類似フレーム画像と同じ順序で並べたときに、当該並び順が、前記クエリ動画における前記少なくとも２つの前記クエリフレームの並び順に一致する。

【０００８】

本発明によれば、コンピュータが、

複数の第１フレーム画像を含むクエリ動画を取得する取得工程と、

前記複数の第１フレーム画像から複数のクエリフレームを選択するクエリフレーム選択工程と、

前記複数のクエリフレームを用いて動画を選択する動画選択工程と、

を行い、

前記動画選択工程において、前記動画を選択する条件として、少なくとも以下の（１）及び（２）を用いる、画像選択方法が提供される。

（１）「前記クエリフレームに対する類似度が第１基準を満たす類似フレーム画像が存在する」という条件が少なくとも２つの前記クエリフレームについて満たされる。

（２）前記少なくとも２つの類似フレーム画像のそれぞれに対応する前記クエリフレームを前記少なくとも２つの類似フレーム画像と同じ順序で並べたときに、当該並び順が、前記クエリ動画における前記少なくとも２つの前記クエリフレームの並び順に一致する。

【０００９】

本発明によれば、コンピュータに、

複数の第１フレーム画像を含むクエリ動画を取得する取得機能と、

前記複数の第１フレーム画像から複数のクエリフレームを選択するクエリフレーム選択機能と、

前記複数のクエリフレームを用いて動画を選択する動画選択機能と、

を持たせ、

前記動画選択機能は、前記動画を選択する条件として、少なくとも以下の（１）及び（２）を用いるプログラムが提供される。

（１）「前記クエリフレームに対する類似度が第１基準を満たす類似フレーム画像が存在する」という条件が少なくとも２つの前記クエリフレームについて満たされる。

（２）前記少なくとも２つの類似フレーム画像のそれぞれに対応する前記クエリフレームを前記少なくとも２つの類似フレーム画像と同じ順序で並べたときに、当該並び順が、前記クエリ動画における前記少なくとも２つの前記クエリフレームの並び順に一致する。

【発明の効果】

【００１０】

本発明によれば、動画をクエリとして、このクエリに類似する動画を選択する場合において、選択の精度は上がる。

【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３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における検索部の機能構成を示す図である。
【図４１】動画選択部が類似フレームを選択する方法の一例を説明するための図である。
【図４２】検索部が行う処理の一例を示すフローチャートである。
【図４３】図４２に示す処理を説明するための図である。
【図４４】図４２の変形例を示すフローチャートである。
【図４５】クエリフレームの選択規則の第１例を説明するための図である。
【図４６】クエリフレームの選択規則の第２例を説明するための図である。
【図４７】第３基準を設定する方法を説明するための図である。

【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。そして「キーポイント」の位置は関節情報の一例になる。

【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索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徴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徴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、分類対象の特徴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タ（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）の特徴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。分類と同様に、類似度は、骨格構造の特徴量間の距離である。検索部１０５は、骨格構造の全体の特徴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徴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）の特徴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徴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徴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、骨格構造の特徴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徴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１０

４０、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。骨格構造検出部１０２は、検出した人物の骨格構造をデータベース１１０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検出され、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。

【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。

【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１００は、検出された骨格構造の特徴量を算出する（Ｓ１０３）。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１０３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１０３は、求めた骨格構造の特徴量をデータベース１１０に格納する。なお、この骨格構造の特徴量は、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）、算出された類似度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Ｃ３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、データベース１１０は動画を記憶している。動画を構成する複数のフレーム画像のそれぞれには、上記した処理が行われている。そして画像処理装置１００はクエリとなる動画（以下、クエリ動画と記載）に類似する動画をデータベース１１０から検索する。
【００６７】
図４０は、本検索方法における検索部１０５の機能構成を示す図である。本図に示す検索部１０５は、クエリ取得部６１０、クエリフレーム選択部６２０、及び動画選択部６３０を備えている。
【００６８】
クエリ取得部６１０は、上記したクエリ動画を取得する。以下の説明において、クエリ動画に含まれるフレーム画像を第１フレーム画像と記載する。言い換えると、クエリ画像は複数の第１フレーム画像を含んでいる。
【００６９】
クエリフレーム選択部６２０は、複数の第１フレーム画像から複数のクエリフレームを選択する。クエリフレームは、検索対象となっている複数の動画からクエリ動画に類似する動画を選択する際に用いられる。クエリフレームの数は、２以上であればよいが、多いほど好ましい。クエリフレームの選択基準については後述する。
【００７０】
なお、あるクエリフレーム画像と、その次のクエリフレーム画像の間に、他のフレーム画像が存在することがある。後述する選択基準のいずれにおいても、クエリフレームの間のフレーム画像の数（又は時間）が基準以内であるのが好ましい。ここで用いられる基準は、例えば１フレーム以上１０フレーム以下、又は０．０２５秒以上１秒以下である。
【００７１】
動画選択部６３０は、複数のクエリフレームを用いて動画を選択する。この際、動画選択部６３０は、動画を選択する条件として、以下の基準（１）及び（２）を用いる。
基準（１）「クエリフレームに対する類似度が第１基準を満たす類似フレーム画像が存在する」という条件が少なくとも２つのクエリフレームについて満たされる。
基準（２）少なくとも２つの類似フレーム画像のそれぞれに対応するクエリフレームを少なくとも２つの類似フレーム画像と同じ順序で並べたときに、当該並び順が、クエリ動画における少なくとも２つのクエリフレームの並び順に一致する。
【００７２】
まず、上記（１）について説明する。第１の動画と第２の動画が互いに類似する場合、第１の動画に含まれる少なくとも２つのフレーム画像は、それぞれ、第２の動画に含まれるいずれかのフレーム画像に類似しているはずである。上記（１）は、これに対応している。
【００７３】
次に、上記（２）について説明する。複数のフレーム画像が互いに類似していたとしても、これら複数のフレーム画像の出現順が一致していない場合、第１の動画と第２の動画は互いに異なると判断すべきである。基準（２）は、複数のフレーム画像の出現順が一致していることを求めている。
【００７４】
上記した基準（１）及び（２）は、以下のようにも置き換えることができる。
基準（１）´少なくとも第１のクエリフレームに対する類似度が基準を満たす第１の類似フレーム画像と、第２の前記クエリフレームに対する類似度が基準を満たす第２の類似フレーム画像とを含んでいる。
基準（２）´第１の類似フレーム画像と第２の類似フレーム画像の順序が第１のクエリフレームと第２のクエリフレームの順序と同一である。
【００７５】
なお、第１の類似フレーム画像と第２の類似フレーム画像の間に、他のフレーム画像が

存在することがある。この場合、第１の類似フレームと第２の類似フレームの間のフレーム画像の数（又は時間）が基準以内であることを、基準（２）（又は基準（２）′）にさらに加えてもよい。ここで用いられる基準は、例えば１フレーム以上１０フレーム以下、又は０．０２５秒以上１秒以下である。

【００７６】

本検索方法が適用できる動画は、人物の姿勢を含む動画に限定されない。ただし、人物の姿勢を含む動画を検索する場合、フレーム画像間の類似度には、人物の姿勢の類似度を用いることができる。

【００７７】

フレーム画像間の類似度として人物の姿勢の類似度を用いる場合、クエリフレームとして選択されるための条件は、「クエリフレームに含まれる人物に関する情報量が第２基準を満たしていること」を含むのが好ましい。言い換えると、画像解析によってその人物に関する情報が得られるが、クエリフレームは、この情報量がある程度多いことが必要である。第２基準の一例は、クエリフレームに含まれるキーポイント（関節）の数が基準を満たしていることである。この基準は、例えばキーポイントの全数に対して３０％以上（例えば１４個中５個）の値として設定される。ここで、手足に相当するキーポイントを少なくとも一つ以上含んでいるのが好ましい。なお、キーポイントのうち首、両肩、及び頭部に相当する部分は欠損しにくい。キーポイントの数が少ない場合、人物の姿勢に関する情報量が少なくなるため、動画の検索精度が低下してしまう。

【００７８】

図４１は、動画選択部６３０が類似フレームを選択する方法の一例を説明するための図である。本図に示す例において、人物の姿勢は、上記した骨格構造の特徴量（パラメータ）によって定義されている。そして動画選択部６３０は、この特徴量によって定義される空間（以下、特徴量空間と記載）における距離を用いて類似フレームを選択する。詳細には、動画選択部６３０は、特徴量空間においてクエリフレームからの距離が第１基準以内のフレーム画像を、類似フレーム画像として選択する。なお、動画選択部６３０は、第１基準をユーザからの入力に従って設定する。ただし動画選択部６３０は、第１基準として予め定められた値を用いてもよい。

【００７９】

図４２は、本検索方法に係る検索部１０５が行う処理の一例を示すフローチャートである。図４３は、図４２に示す処理を説明するための図である。

【００８０】

まず、クエリ取得部６１０はクエリ動画を取得する（ステップＳ３００）。一例として、クエリ取得部６１０は、ユーザが指定した動画をクエリ動画として取得する。ユーザは、データベース１１０に記憶されている動画からクエリ動画を選択してもよいし、外部の装置または記憶媒体からクエリ動画を検索部１０５に取得させてもよい。

【００８１】

次いでクエリフレーム選択部６２０は、クエリ動画に含まれる複数の第１フレーム画像から、複数のクエリフレームを選択する。本図に示す例では、クエリフレーム選択部６２０は、ユーザの入力に従って複数のクエリフレームを選択する（ステップＳ３０２）。

【００８２】

詳細には、クエリフレーム選択部６２０は、図４３に示すように、クエリ動画に含まれる複数の第１フレーム画像（好ましくはすべての第１フレーム画像）を出現順に並べて表示部１０７に表示させる。するとユーザは、マウス等の入力デバイスを用いて、クエリフレームとすべき第１フレーム画像を選択する。

【００８３】

次いで動画選択部６３０は、ステップＳ３０２で選択された複数のクエリフレームを用いて、クエリ動画に類似する動画を選択する（ステップＳ３０４）。この選択基準は、図４０及び図４１を用いて説明した通りである。なお、図４３に示す図においても、選択された動画は上記した基準（１）及び（２）を満たしているとともに、基準（１）′及び（

２）′も満たしている。
【００８４】
なお、クエリ動画がデータベース１１０に記憶されていなかった場合、骨格構造検出部１０２及び特徴量算出部１０３は、クエリフレームを処理して骨格構造の特徴量を算出する。そして動画選択部６３０は、この骨格構造の特徴量を用いて、クエリ動画に類似する動画を選択する。
【００８５】
その後、動画選択部６３０は、選択結果をデータベース１１０に記憶させる。ここで動画選択部６３０は、選択した動画そのものを記憶してもよいし、データベース１１０に既に記憶されている当該動画に、その画像がクエリ動画に類似することを示すフラグを紐づけてもよい（ステップＳ３０６）。
【００８６】
図４４は、図４２の変形例を示すフローチャートである。本変形例は、クエリフレーム選択部６２０が予め定められた規則（以下、選択規則と記載）に従って複数のクエリ画像を選択する（ステップＳ３０３）点を除いて、図４２に示した例と同様である。
【００８７】
図４５は、クエリフレームの選択規則の第１例を説明するための図である。本図に示す例において、クエリフレーム選択部６２０は、所定の間隔で複数の第１フレーム画像から複数のクエリフレームを選択する。ここでクエリフレーム選択部６２０は、最初のクエリフレームとして、ｎ番目の第１フレーム画像を選択する。ここでｎは、予め設定されている整数であってもよいし、ユーザ入力によって設定された整数であってもよい。また、所定の間隔は、例えば２フレーム以上１０フレーム以下、又は０．０５秒以上１秒以下である。
【００８８】
図４６は、クエリフレームの選択規則の第２例を説明するための図である。本図に示す例において、クエリフレーム選択部６２０は、第１のクエリフレームが特定された後、当該第１のクエリフレームからの変化量が第３基準以上となった第１フレーム画像を、第１のクエリフレームの次のクエリフレームとして選択する。そしてクエリフレーム選択部６２０は、この処理を繰り返す。これにより、隣り合うクエリフレームに含まれる人物の姿勢は、互いにある程度異なる。従って、クエリフレームが増加することを抑制しつつ、検索部１０５による動画の検索精度を高くすることができる。
【００８９】
ここで、第３基準はユーザからの入力に従って設定されてもよいし、複数の第１クエリフレームを用いて設定されてもよい。
【００９０】
図４７は、クエリフレーム選択部６２０が複数の第１クエリフレームを用いて第３基準を設定する方法を説明するための図である。本図に示す例において、第３基準は、特徴量空間における距離である。そして具体例としては、以下の２つがある。
【００９１】
第１の例において、クエリフレーム選択部６２０は、特徴量空間における２つのクエリフレームの最大距離を特定し、この最大距離を用いて第３基準を設定する。一例として、クエリフレーム選択部６２０は、最大距離を変数とした関数を用いて、第３基準を最大距離未満の値に設定する。一例として、クエリフレーム選択部６２０は、最大距離に１未満の係数を乗じたり、最大距離から所定の値を引くことにより、第３基準を設定する。
【００９２】
第２の例において、クエリフレーム選択部６２０は、時間的に隣り合う２つの第１フレームの距離を統計処理した結果を用いて、第３基準を設定する。例えばクエリフレーム選択部６２０は、これらの距離の中央値、平均値、又は最頻値を第３基準としてもよいし、中央値、平均値、及び最頻値の少なくとも一つを変数とした関数を用いて第３基準を設定してもよい。

【００９３】
なお、図４６及び図４７に示す例においても、クエリフレーム選択部６２０は、図４５に示した例と同様に、最初のクエリフレームとしてｎ番目の第１フレーム画像を選択する。
【００９４】
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。
【００９５】
また、検索方法６によれば、クエリ動画に類似する動画を精度良く検索することができる。また、クエリ動画と検索対象となる動画との間で、フレームレートが異なったり再生時間が異なっていた場合でも、動画の検索を行うことができる。
【００９６】
（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。
【００９７】
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。
【００９８】
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。
【００９９】
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。
【０１００】
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。

【０１０１】

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。

【０１０２】

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。

【０１０３】

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（ｙｉ）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。

【０１０４】

＜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）。

【０１０５】

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。

【０１０６】

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検出されている。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。

【０１０７】

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。

【０１０８】

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。

【０１０９】

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。

【０１１０】

＜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２次元画像空間上の人物の全身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。

【０１１１】

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２次元画像空間上の全身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全身の長さとの関係（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。

【０１１２】

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。

【０１１３】

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１０８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×０．１５であるため、ボーンＢ４１の長さ／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×０．２５であるため、ボーンＢ７１の長さ／０．２５によりボーンＢ７１に基づいた身長画素数を求める。

【０１１４】

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。

【０１１５】

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。

【０１１６】

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、ボーンを正面から撮像できていない場合、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。

【０１１７】

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。

【０１１８】

＜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。

【０１１９】

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。

【０１２０】

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。

【０１２１】

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。

【０１２２】

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。

【０１２３】

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。

【０１２４】

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。

【０１２５】

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。

【０１２６】

＜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（ｙｉ）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。

【０１２７】

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。

【０１２８】

続いて、特徴量算出部１０３は、キーポイント高さ（ｙｉ）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（ｙｃ）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、（ｙｉ）及び（ｙｃ）を鉛直投影軸に沿った方向の値に変換する。

【数１】

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

【０１２９】

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（ｘ０、ｙ０）、（ｘ１、ｙ１）、・・・（ｘ１７、ｙ１７）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】

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

【０１３０】

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。

【０１３１】

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。

【０１３２】

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。

【０１３３】

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。

【０１３４】

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。

【０１３５】

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．複数の第１フレーム画像を含むクエリ動画を取得するクエリ取得手段と、

前記複数の第１フレーム画像から複数のクエリフレームを選択するクエリフレーム選択手段と、

前記複数のクエリフレームを用いて動画を選択する動画選択手段と、

を備え、

前記動画選択手段は、前記動画を選択する条件として、少なくとも以下の（１）及び（２）を用いる、画像選択装置。

（１）「前記クエリフレームに対する類似度が第１基準を満たす類似フレーム画像が存在する」という条件が少なくとも２つの前記クエリフレームについて満たされる。

（２）前記少なくとも２つの類似フレーム画像のそれぞれに対応する前記クエリフレームを前記少なくとも２つの類似フレーム画像と同じ順序で並べたときに、当該並び順が、前記クエリ動画における前記少なくとも２つの前記クエリフレームの並び順に一致する。

２．上記１に記載の画像選択装置において、

前記クエリ動画は人物を含んでおり、

前記動画選択手段は、前記類似度として、前記人物の姿勢の類似度を用いる画像選択装置。
３．上記２に記載の画像選択装置において、
前記クエリフレームが満たすべき条件の一つは、当該クエリフレームに含まれる前記人物に関する情報量が第２基準を満たしていることである、画像選択装置。
４．上記３に記載の画像選択装置において、
前記複数の第１フレーム画像のそれぞれは、前記人物の姿勢を示す情報として、当該人物の関節の位置を示す関節情報を含んでおり、
前記第２基準は、前記関節情報に含まれる関節の数が基準を満たしていることである画像選択装置。
５．上記１～４のいずれか一項に記載の画像選択装置において、
前記クエリフレーム選択手段は、ユーザからの入力に従って少なくとも一つの前記クエリフレームを選択する画像選択装置。
６．上記１～４のいずれか一項に記載の画像選択装置において、
前記クエリフレーム選択手段は、第１の前記クエリフレームの次の前記クエリフレームとして前記第１のクエリフレームからの変化量が第３基準以上となった前記第１フレーム画像を選択する処理を、繰り返す画像選択装置。
７．上記６に記載の画像選択装置において、
前記クエリフレーム選択手段は、ユーザからの入力を用いて前記第３基準を設定する画像選択装置。
８．上記６に記載の画像選択装置において、
前記クエリフレーム選択手段は、前記複数の第１フレーム画像を用いて当該クエリフレームに用いる前記第３基準を設定する画像選択装置。
９．上記１～４のいずれか一項に記載の画像選択装置において、
前記クエリフレーム選択手段は、所定の間隔で前記複数の第１フレーム画像から複数のクエリフレームを選択する画像選択装置。
１０．上記６～９のいずれか一項に記載の画像選択装置において、
前記クエリフレーム選択手段は、最初の前記クエリフレームとして、ｎ番目の前記第１フレーム画像を選択する画像選択装置。
ここで、ｎは予め又はユーザ入力により設定された整数である。
１１．コンピュータが、
複数の第１フレーム画像を含むクエリ動画を取得する取得処理と、
前記複数の第１フレーム画像から複数のクエリフレームを選択するクエリフレーム選択処理と、
前記複数のクエリフレームを用いて動画を選択する動画選択処理と、
を行い、
前記動画選択処理において、前記動画を選択する条件として、少なくとも以下の（１）及び（２）を用いる、画像選択方法。
（１）「前記クエリフレームに対する類似度が第１基準を満たす類似フレーム画像が存在する」という条件が少なくとも２つの前記クエリフレームについて満たされる。
（２）前記少なくとも２つの類似フレーム画像のそれぞれに対応する前記クエリフレームを前記少なくとも２つの類似フレーム画像と同じ順序で並べたときに、当該並び順が、前記クエリ動画における前記少なくとも２つの前記クエリフレームの並び順に一致する。
１２．上記１１に記載の画像選択方法において、
前記クエリ動画は人物を含んでおり、
前記動画選択処理において、前記コンピュータは、前記類似度として、前記人物の姿勢の類似度を用いる画像選択方法。
１３．上記１２に記載の画像選択方法において、
前記クエリフレームが満たすべき条件の一つは、当該クエリフレームに含まれる前記人物に関する情報量が第２基準を満たしていることである、画像選択方法。

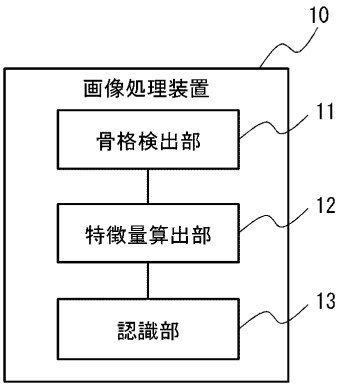
１４．上記１３に記載の画像選択方法において、
　前記複数の第１フレーム画像のそれぞれは、前記人物の姿勢を示す情報として、当該人物の関節の位置を示す関節情報を含んでおり、
　前記第２基準は、前記関節情報に含まれる関節の数が基準を満たしていることである画像選択方法。
１５．上記１１～１４のいずれか一項に記載の画像選択方法において、
　前記クエリフレーム選択処理において、前記コンピュータは、ユーザからの入力に従って少なくとも一つの前記クエリフレームを選択する画像選択方法。
１６．上記１１～１４のいずれか一項に記載の画像選択方法において、
　前記クエリフレーム選択処理において、前記コンピュータは、第１の前記クエリフレームの次の前記クエリフレームとして前記第１のクエリフレームからの変化量が第３基準以上となった前記第１フレーム画像を選択する処理を、繰り返す画像選択方法。
１７．上記１６に記載の画像選択方法において、
　前記クエリフレーム選択処理において、前記コンピュータは、ユーザからの入力を用いて前記第３基準を設定する画像選択方法。
１８．上記１６に記載の画像選択方法において、
　前記クエリフレーム選択処理において、前記コンピュータは、前記複数の第１フレーム画像を用いて当該クエリフレームに用いる前記第３基準を設定する画像選択方法。
１９．上記１１～１４のいずれか一項に記載の画像選択方法において、
　前記クエリフレーム選択処理において、前記コンピュータは、所定の間隔で前記複数の第１フレーム画像から複数のクエリフレームを選択する画像選択方法。
２０．上記１６～１９のいずれか一項に記載の画像選択方法において、
　前記クエリフレーム選択処理において、前記コンピュータは、最初の前記クエリフレームとして、ｎ番目の前記第１フレーム画像を選択する画像選択方法。
　ここで、ｎは予め又はユーザ入力により設定された整数である。

２１．コンピュータに、
　　複数の第１フレーム画像を含むクエリ動画を取得する取得機能と、
　　前記複数の第１フレーム画像から複数のクエリフレームを選択するクエリフレーム選択機能と、
　前記複数のクエリフレームを用いて動画を選択する動画選択機能と、
を持たせ、
　前記動画選択機能は、前記動画を選択する条件として、少なくとも以下の（１）及び（２）を用いるプログラム。
　（１）「前記クエリフレームに対する類似度が第１基準を満たす類似フレーム画像が存在する」という条件が少なくとも２つの前記クエリフレームについて満たされる。
　（２）前記少なくとも２つの類似フレーム画像のそれぞれに対応する前記クエリフレームを前記少なくとも２つの類似フレーム画像と同じ順序で並べたときに、当該並び順が、前記クエリ動画における前記少なくとも２つの前記クエリフレームの並び順に一致する。
２２．上記２１に記載のプログラムにおいて、
　前記クエリ動画は人物を含んでおり、
　前記動画選択機能は、前記類似度として、前記人物の姿勢の類似度を用いるプログラム。
２３．上記２２に記載のプログラムにおいて、
　前記クエリフレームが満たすべき条件の一つは、当該クエリフレームに含まれる前記人物に関する情報量が第２基準を満たしていることである、プログラム。
２４．上記２３に記載のプログラムにおいて、
　前記複数の第１フレーム画像のそれぞれは、前記人物の姿勢を示す情報として、当該人物の関節の位置を示す関節情報を含んでおり、
　前記第２基準は、前記関節情報に含まれる関節の数が基準を満たしていることであるプログラム。

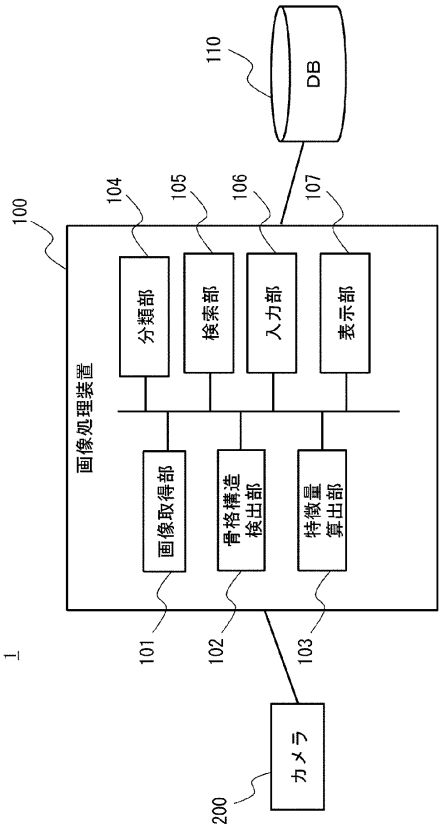
２５．上記２１～２４のいずれか一項に記載のプログラムにおいて、
　前記クエリフレーム選択機能は、ユーザからの入力に従って少なくとも一つの前記クエリフレームを選択するプログラム。
２６．上記２１～２４のいずれか一項に記載のプログラムにおいて、
　前記クエリフレーム選択機能は、第１の前記クエリフレームの次の前記クエリフレームとして前記第１のクエリフレームからの変化量が第３基準以上となった前記第１フレーム画像を選択する処理を、繰り返すプログラム。
２７．上記２６に記載のプログラムにおいて、
　前記クエリフレーム選択機能は、ユーザからの入力を用いて前記第３基準を設定するプログラム。
２８．上記２６に記載のプログラムにおいて、
　前記クエリフレーム選択機能は、前記複数の第１フレーム画像を用いて当該クエリフレームに用いる前記第３基準を設定するプログラム。
２９．上記２１～２４のいずれか一項に記載のプログラムにおいて、
　前記クエリフレーム選択機能は、所定の間隔で前記複数の第１フレーム画像から複数のクエリフレームを選択するプログラム。
３０．上記２６～２９のいずれか一項に記載のプログラムにおいて、
　前記クエリフレーム選択機能は、最初の前記クエリフレームとして、ｎ番目の前記第１フレーム画像を選択するプログラム。
　ここで、ｎは予め又はユーザ入力により設定された整数である。
【符号の説明】
【０１３６】
１　　画像処理システム
１０　　画像処理装置（画像選択装置）
１１　　骨格検出部
１２　　特徴量算出部
１３　　認識部
１００　　画像処理装置（画像選択装置）
１０１　　画像取得部
１０２　　骨格構造検出部
１０３　　特徴量算出部
１０４　　分類部
１０５　　検索部
１０６　　入力部
１０７　　表示部
１０８　　身長算出部
１１０　　データベース
２００　　カメラ
３００、３０１　　人体モデル
４０１　　２次元骨格構造
６１０　　クエリ取得部
６２０　　クエリフレーム選択部
６３０　　動画選択部

【図面】

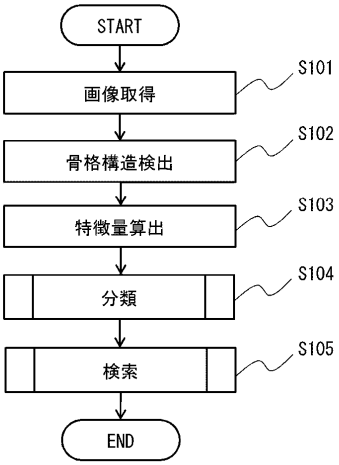
【図１】


10
画像処理装置
11
骨格検出部
12
特徴量算出部
13
認識部

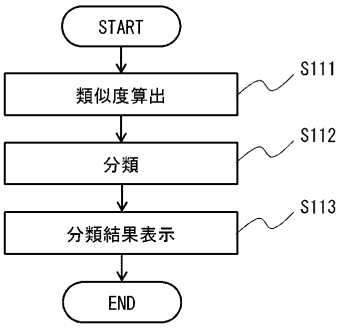
【図２】


110
DB
104
105
106
107
100
分類部
検索部
入力部
表示部
画像処理装置
画像取得部
骨格構造検出部
特徴量算出部
1
101
102
103
カメラ
200

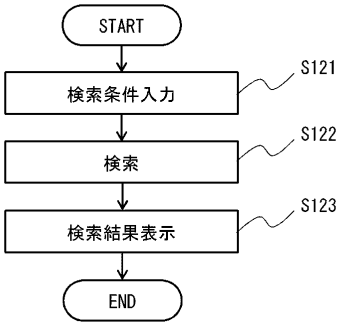
【図３】


START
画像取得
S101
骨格構造検出
S102
特徴量算出
S103
分類
S104
検索
S105
END

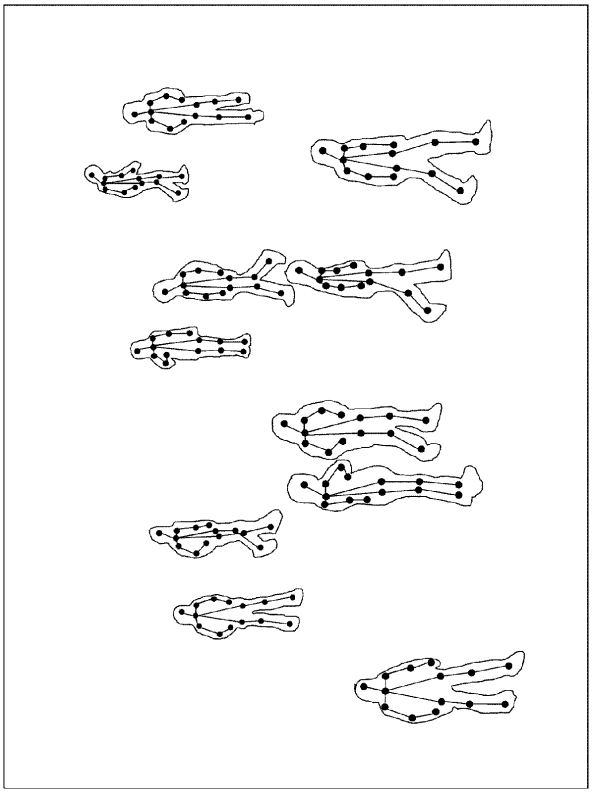
【図４】


START
類似度算出
S111
分類
S112
分類結果表示
S113
END

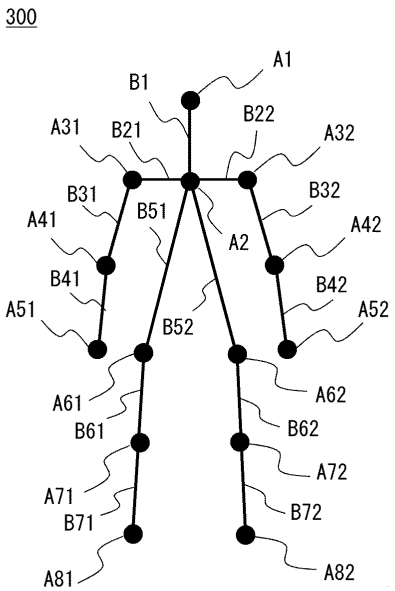
【図５】


START
検索条件入力
S121
検索
S122
検索結果表示
S123
END

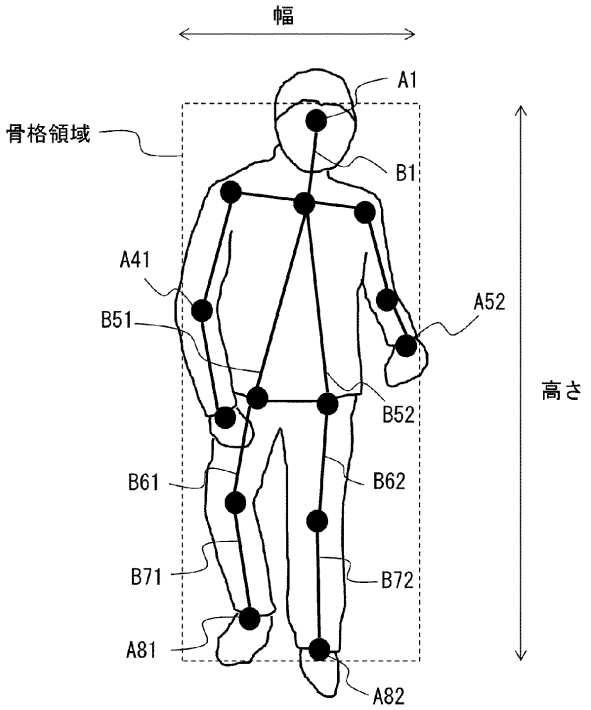
【図６】



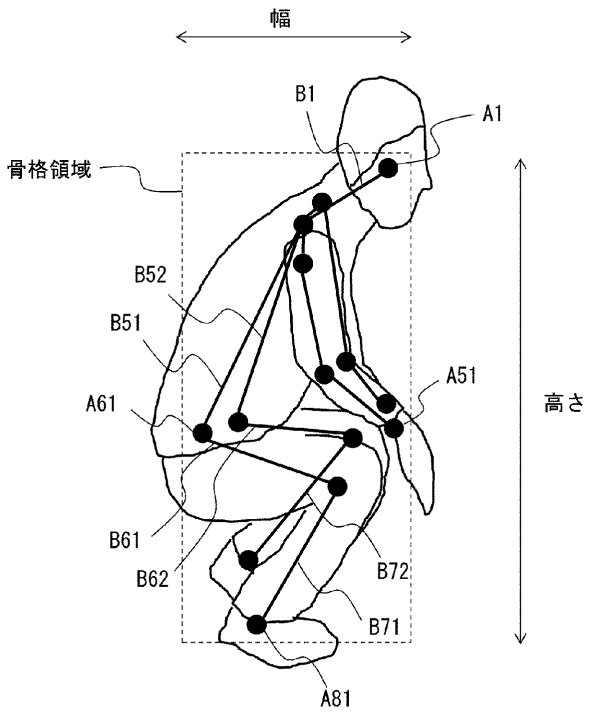
【図７】


300
A1
B1
A31
B21
B22
A32
B31
B51
B32
A41
A2
A42
B41
B42
A51
B52
A52
A61
A62
B61
B62
A71
A72
B71
B72
A81
A82

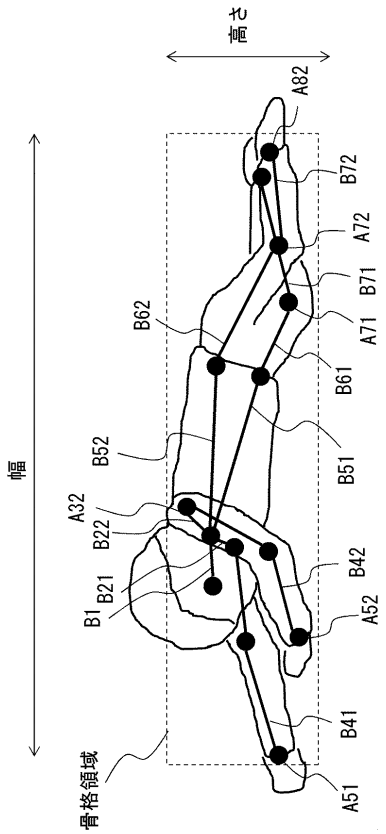
【図８】


幅
骨格領域
A1
B1
A41
A52
B51
B52
高さ
B61
B62
B71
B72
A81
A82

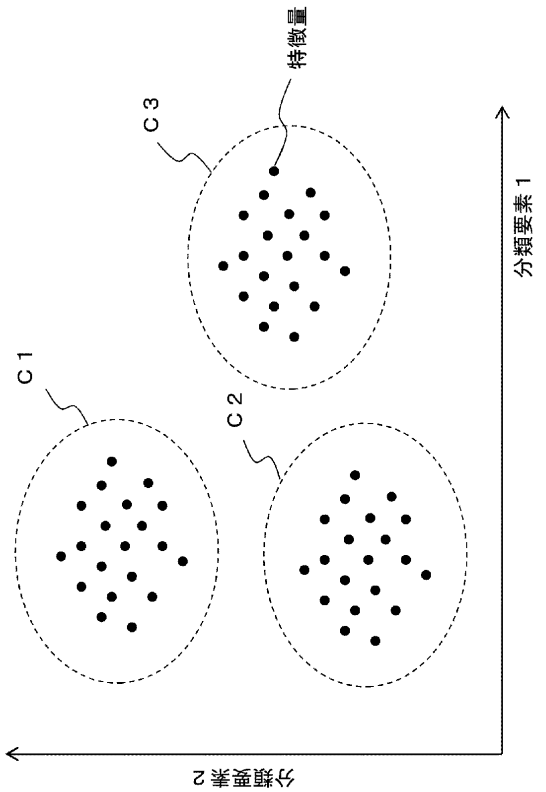
【図９】



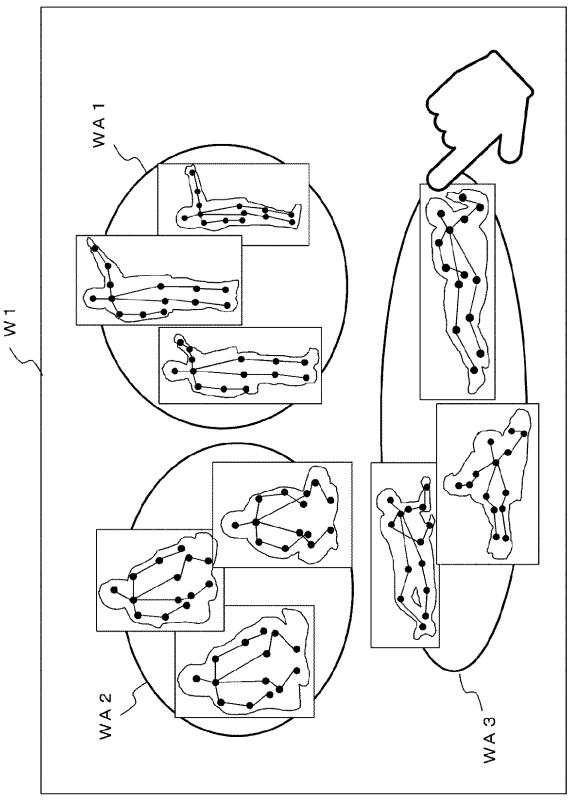
【図１０】



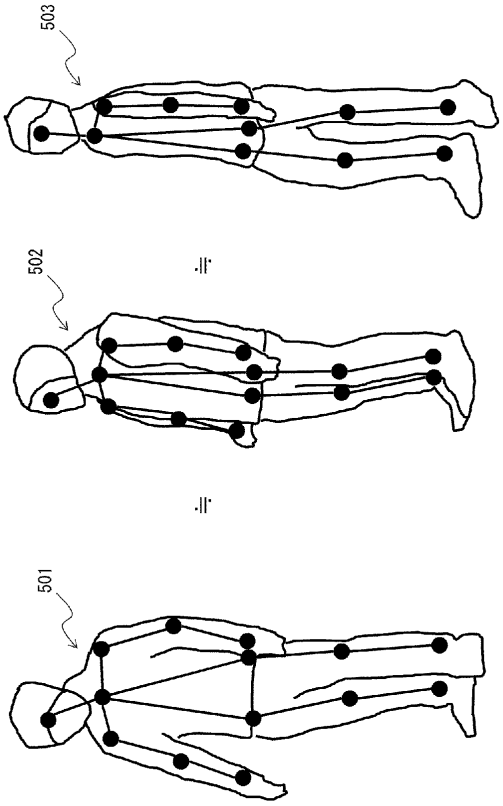
【図１１】



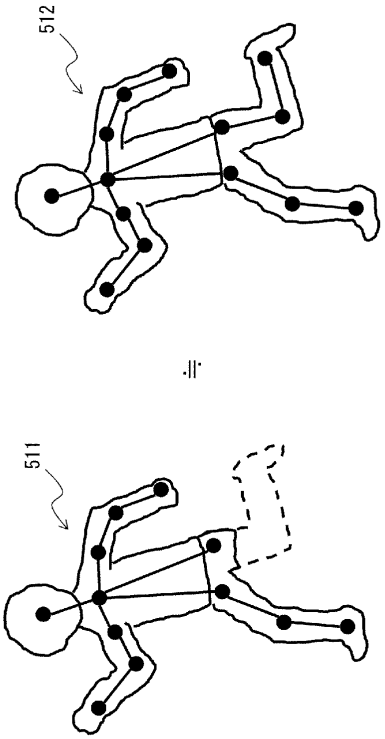
【図１２】



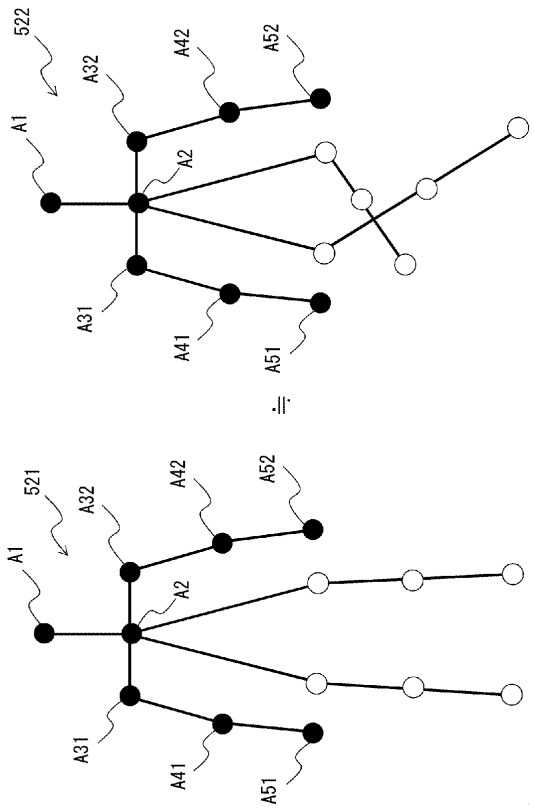
【図１３】



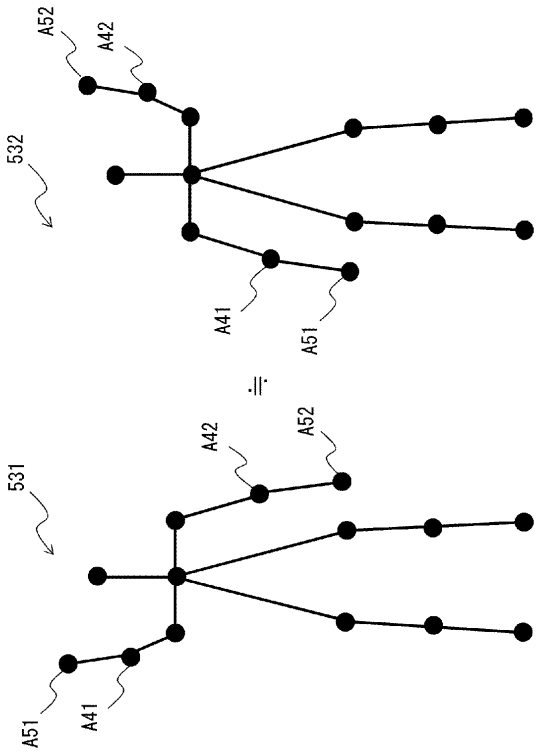
【図１４】



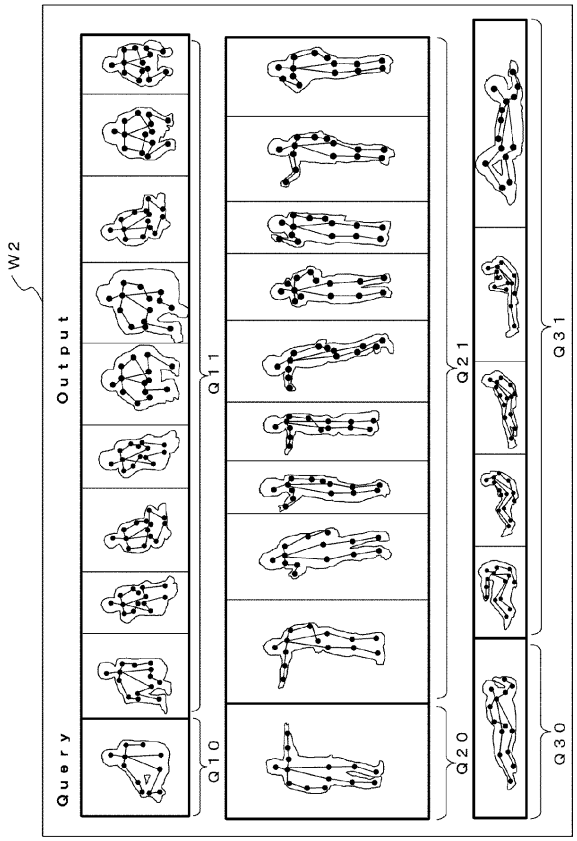
【図１５】



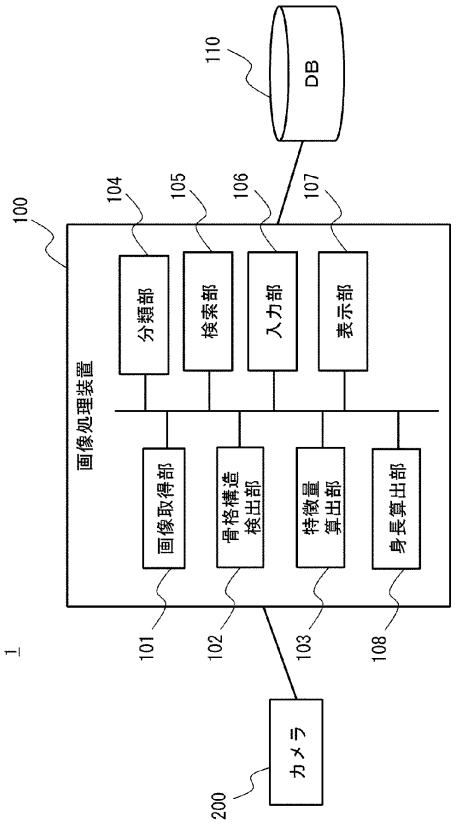
【図１６】



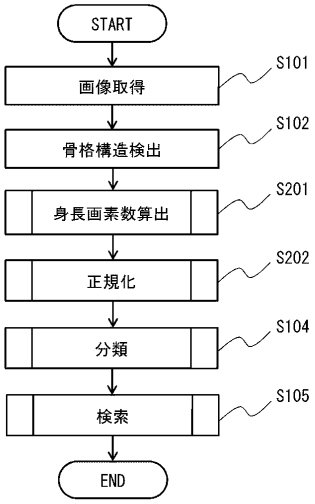
【図１７】



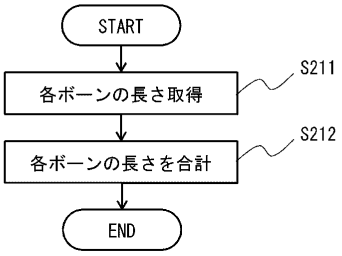
【図１８】



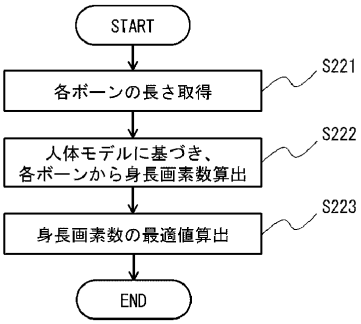
【図１９】



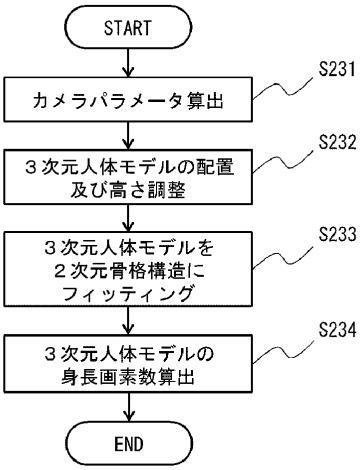
【図２０】



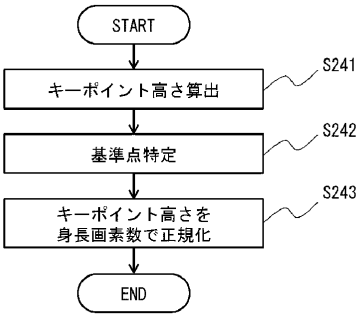
【図２１】



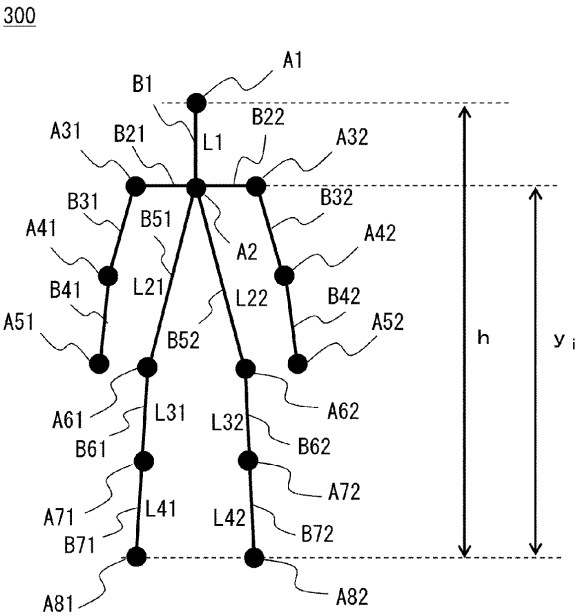
【図２２】



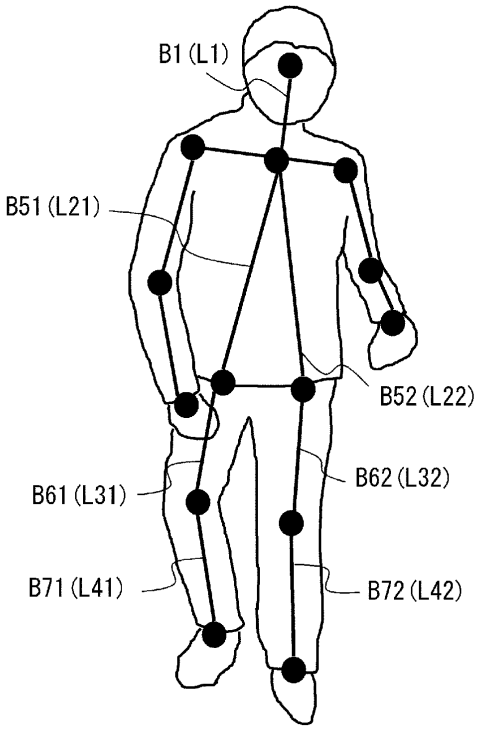
【図２３】



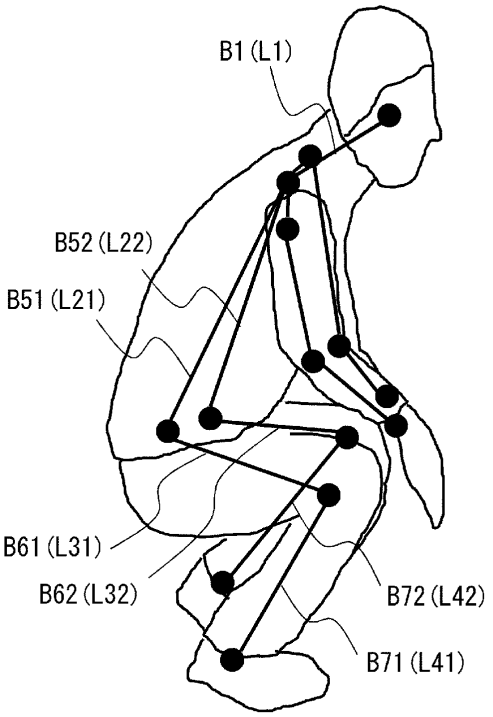
【図２４】



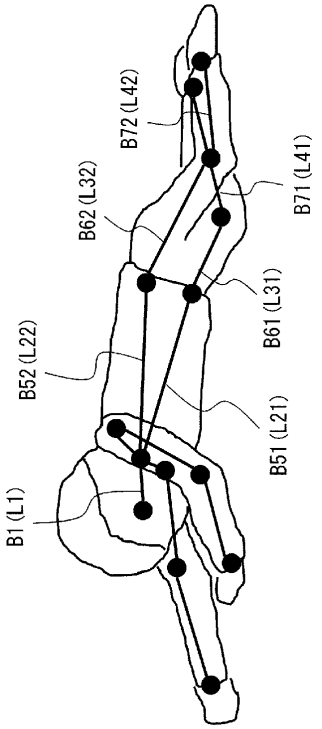
【図２５】



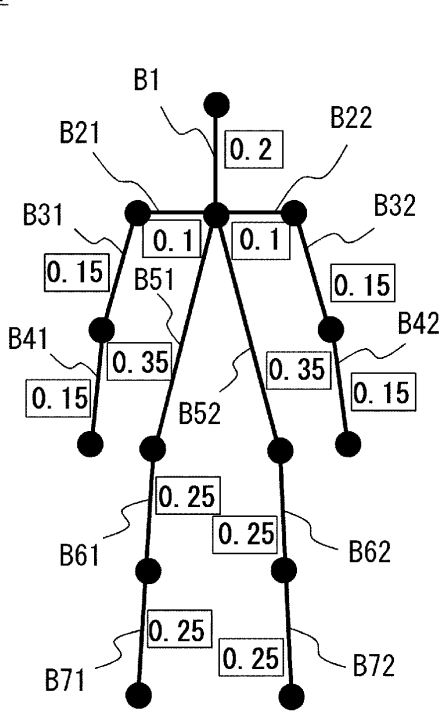
【図２６】



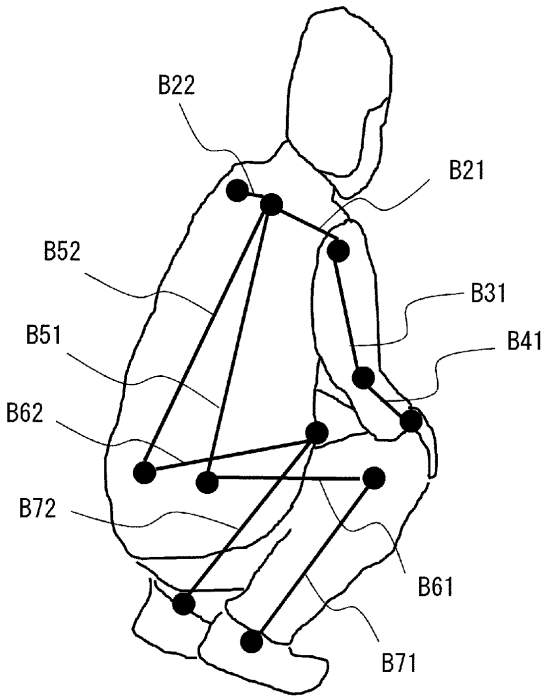
【図２７】



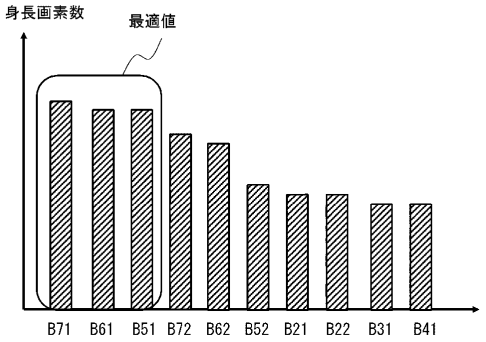
【図２８】



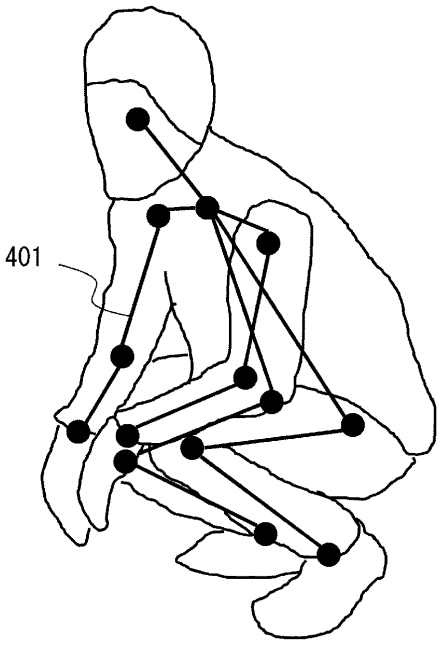
【図２９】



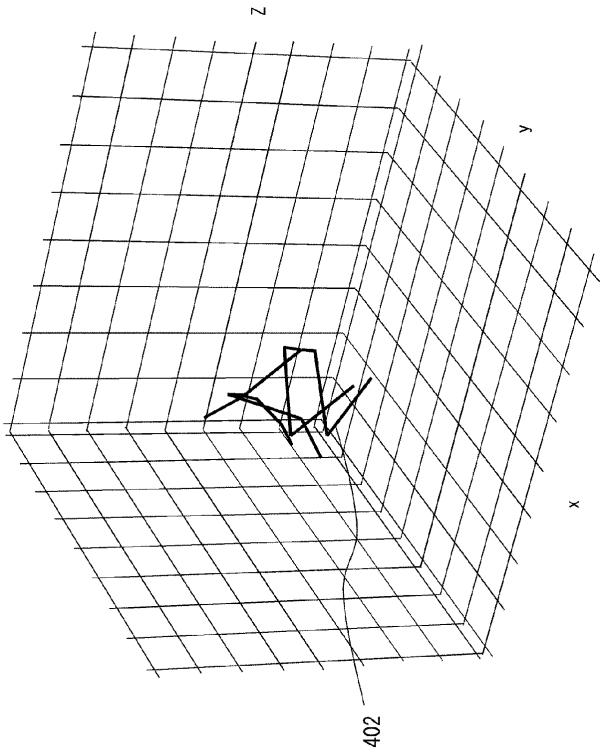
【図３０】



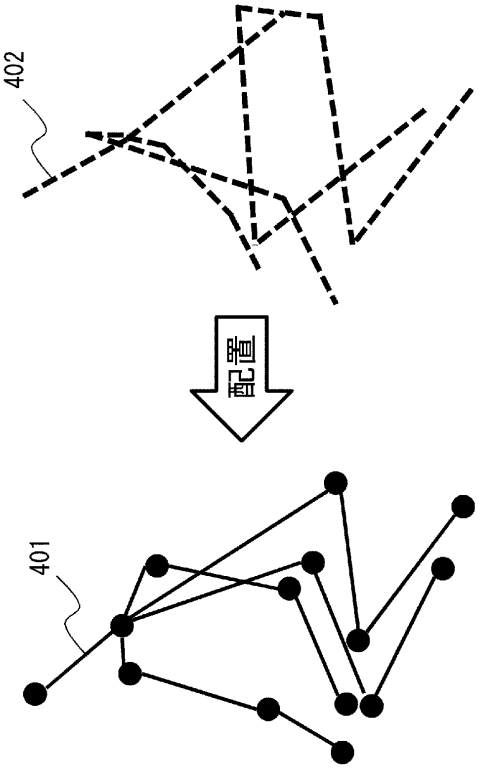
【図３１】



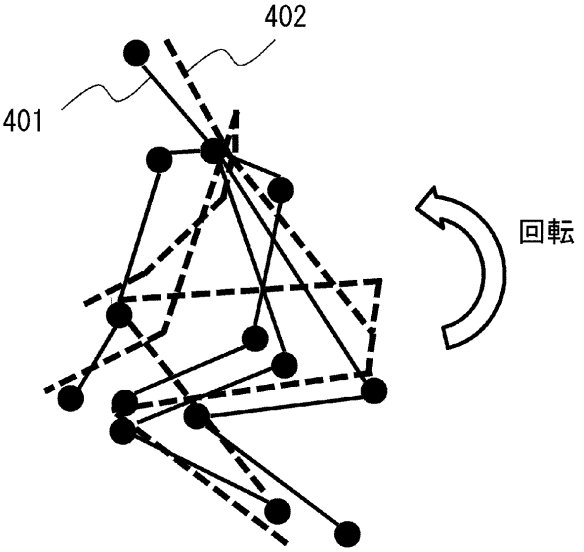
【図３２】



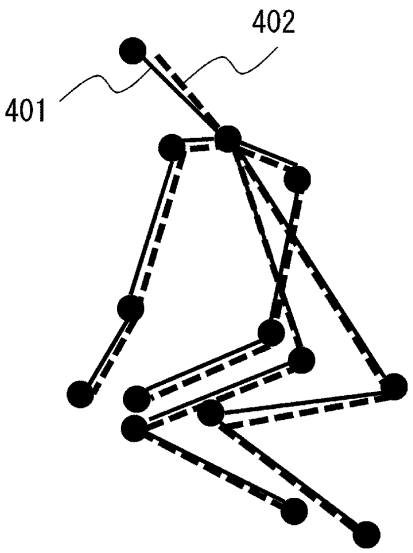
【図３３】



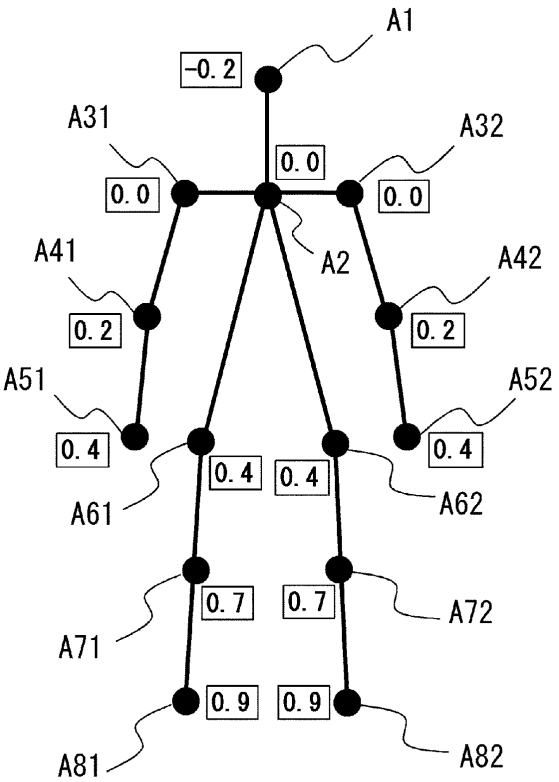
【図３４】



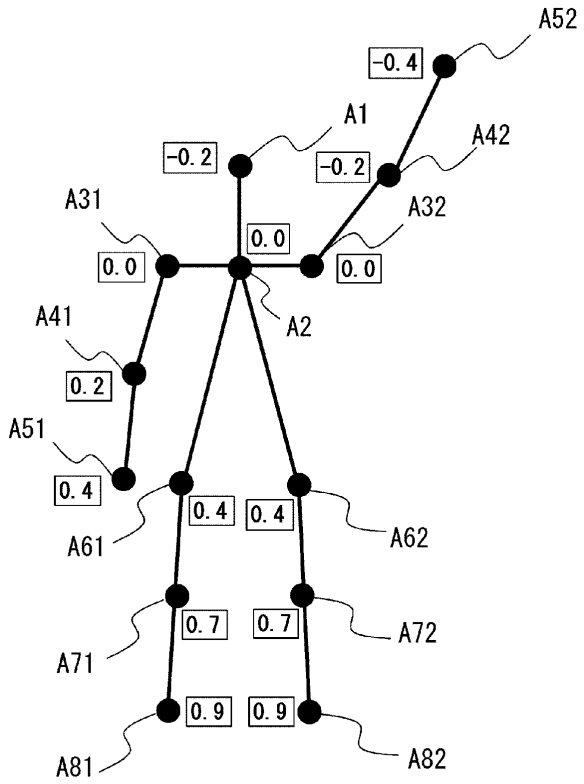
【図３５】



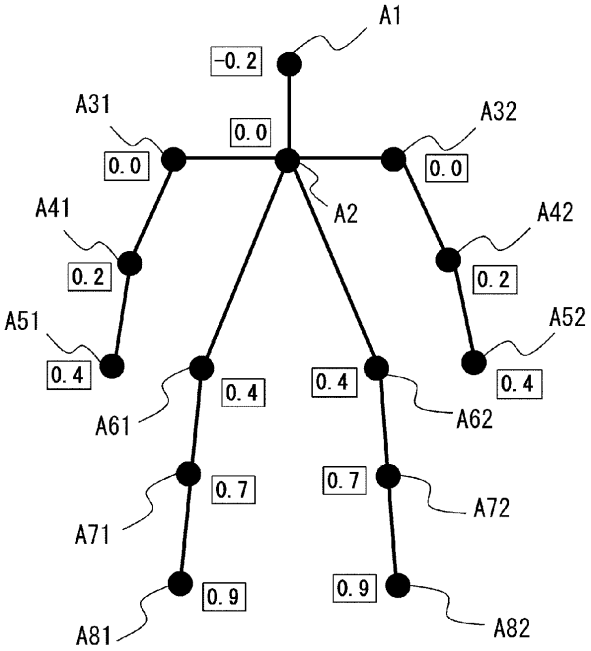
【図３６】



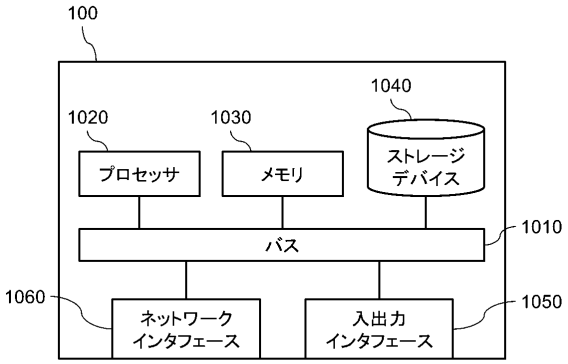
【図３７】



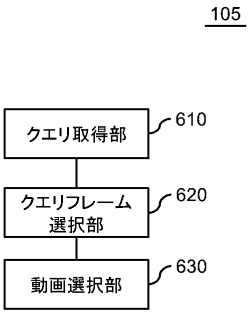
【図３８】



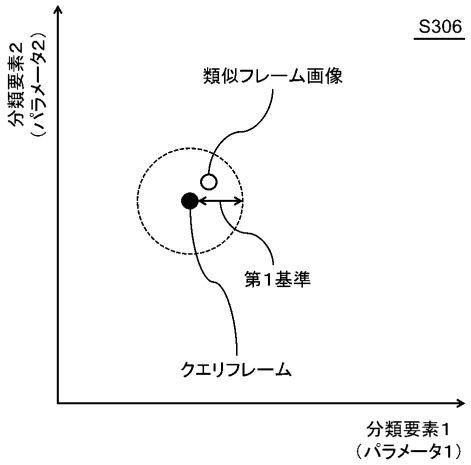
【図３９】



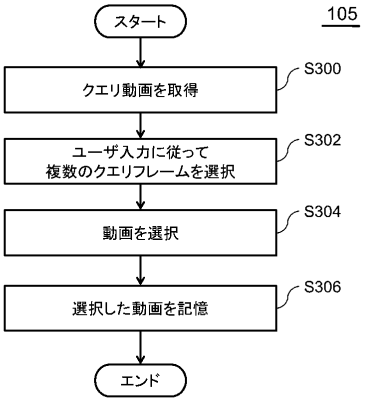
【図４０】



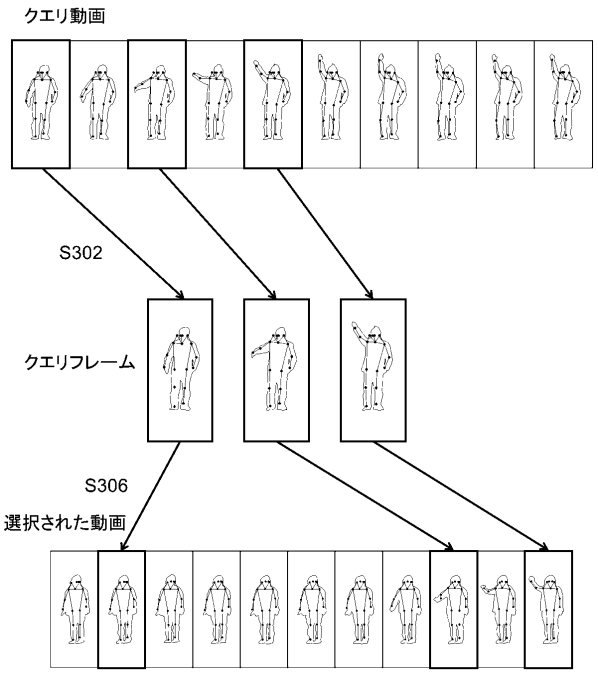
【図４１】



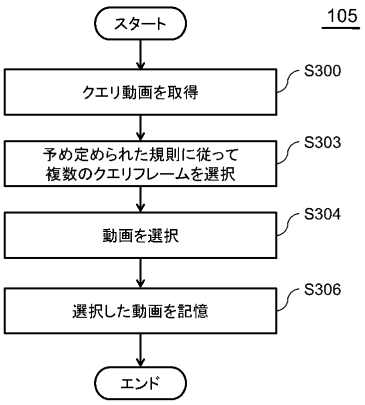
【図４２】



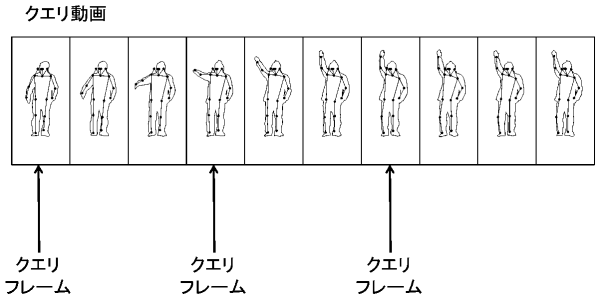
【図４３】



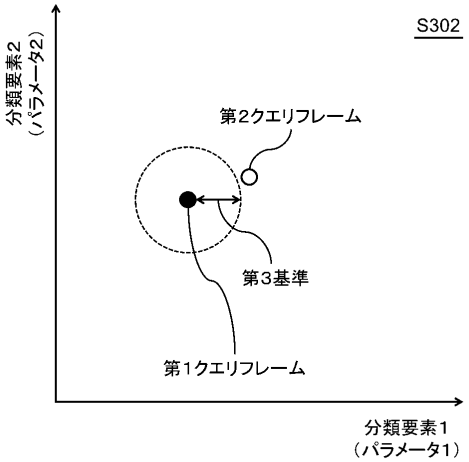
【図４４】



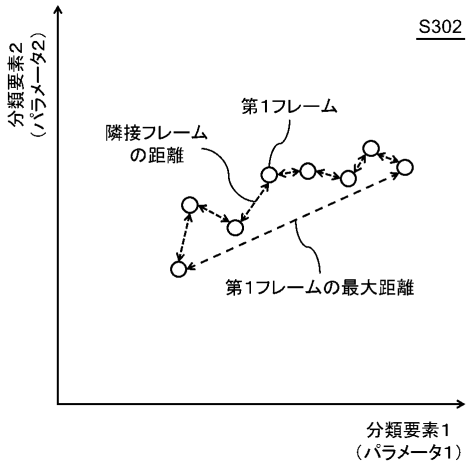
【図４５】


クエリ動画
クエリ
フレーム
クエリ
フレーム
クエリ
フレーム

【図４６】


S302
分類要素2
（パラメータ2）
第2クエリフレーム
第3基準
第1クエリフレーム
分類要素1
（パラメータ1）

【図４７】


S302
分類要素2
（パラメータ2）
第1フレーム
隣接フレーム
の距離
第1フレームの最大距離
分類要素1
（パラメータ1）

フロントページの続き

東京都港区芝五丁目７番１号 日本電気株式会社内
(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
審査官 松尾 真人
(56)参考文献 特開２０１７－１１７３０２（ＪＰ，Ａ）
特開２０１９－０９１１３８（ＪＰ，Ａ）
特開２０１１－１１８７９０（ＪＰ，Ａ）
中国特許出願公開第１１０７２８２０９（ＣＮ，Ａ）
米国特許出願公開第２０１６／０１１０４５３（ＵＳ，Ａ１）
特開２００１－１３４５８９（ＪＰ，Ａ）
国際公開第２００６／０２５２７２（ＷＯ，Ａ１）
(58)調査した分野 (Int.Cl.，ＤＢ名)
Ｇ０６Ｆ １６／００－１６／９５８
Ｇ０６Ｔ １／００－１／４０
Ｇ０６Ｔ ３／００－５／５０
Ｇ０６Ｔ １１／６０－１３／８０
Ｇ０６Ｔ １９／００－１９／２０
Ｇ０６Ｑ １０／００－９９／００